Using binocular rivalry to tag foreground sounds: Towards an objective visual measure for auditory multistability

Wolfgang Einhäuser Chemnitz University of Technology, Institute of Physics, Physics of Cognition Group, Chemnitz, Germany

Sabine Thomassen Chemnitz University of Technology, Institute of Physics, Cognitive Systems Lab, Chemnitz, Germany

Alexandra Bendixen Chemnitz University of Technology, Institute of Physics, Cognitive Systems Lab, Chemnitz, Germany

In binocular rivalry, paradigms have been proposed for unobtrusive moment-by-moment readout of observers' perceptual experience ("no-report paradigms"). Here, we take a first step to extend this concept to auditory multistability. Observers continuously reported which of two concurrent tone sequences they perceived in the foreground: high-pitch (1008 Hz) or low-pitch (400 Hz) tones. Interstimulus intervals were either fixed per sequence (Experiments 1 and 2) or random with tones alternating (Experiment 3). A horizontally drifting grating was presented to each eye; to induce binocular rivalry, gratings had distinct colors and motion directions. To associate each grating with one tone sequence, a pattern on the grating jumped vertically whenever the respective tone occurred. We found that the direction of the optokinetic nystagmus (OKN)—induced by the visually dominant grating—could be used to decode the tone (high/low) that was perceived in the foreground well above chance. This OKN-based readout improved after observers had gained experience with the auditory task (Experiments 1 and 2) and for simpler auditory tasks (Experiment 3). We found no evidence that the visual stimulus affected auditory multistability. Although decoding performance is still far from perfect, our paradigm may eventually provide a continuous estimate of the currently dominant percept in auditory multistability.

Introduction

Ambiguity about the distal sources of a proximal sensory signal is a universal property that any perception system has to deal with (Helmholtz, 1867). A common approach to study the resolution of such ambiguity are multistable phenomena, where, for a given stimulus, perceptual interpretations switch back and forth between two or more alternatives. In vision, multistability is often equated with *rivalry* and comes in many flavors, such as perspective reversals (Necker, 1832), figure-ground reversals (Rubin, 1921), content reversals (Boring, 1930), integration and segregation of transparently overlaid components (Breese, 1899; Wallach, 1935), ambiguous apparent motion (Wertheimer, 1923), or object emergence from moving parts (Lorenceau & Shiffrar, 1992). A widely studied form of rivalry is so-called binocular rivalry (Wheatstone, 1838): When two distinct stimuli are presented to the eyes, these alternate in awareness, and the switching dynamics share many characteristics with other forms of rivalry (Brascamp, Klink, & Levelt, 2015; Klink, van Ee, & van Wezel, 2008; O'Shea, Parker, La Rooy, & Alais, 2009).

In audition, two main forms of multistability are known. Verbal transformations (Warren & Gregory, 1958) occur when a repetitively presented word becomes subject to perceptual reorganization such that different arrangements of the input are heard (e.g., "fly" for the repetitive presentation of "life"). Auditory streaming (van Noorden, 1975), which will be used in the present study, occurs when two repeating tone sequences are presented in an interleaved manner (e.g., "ABABAB…"). They either form a coherent ("integrated") percept of a single tone sequence ("ABABAB…") or are perceived as separate ("segregated") sequences ("A-A-A-…" and "-B-B-B…"), of which either one can be perceived in the foreground. As such, there are three distinct main percepts that alternate over time (one tone in the foreground, the other tone in the foreground, the integrated percept),

Citation: Einhäuser, W., Thomassen, S., & Bendixen, A. (2017). Using binocular rivalry to tag foreground sounds: Towards an objective visual measure for auditory multistability. *Journal of Vision*, *17*(1):34, 1–19, doi:10.1167/17.1.34.

doi: 10.1167/17.1.34 Received September 21, 2016; published January 27, 2017 ISSN 1534-7362

although other perceptual interpretations can emerge after prolonged listening (Denham et al., 2014). The likelihood of perceiving one of the interpretations can be controlled by pitting temporal separation of the individual tones against the (dis)similarity of physical parameters (e.g., frequency) across the two sequences (see the review by Moore & Gockel, 2012): The segregated interpretations become more likely with a higher rate of physical feature change. Further manipulations (e.g., an intensity difference between the sequences) can render one of the segregated interpretations dominant over the other.

The arguably closest visual analogue to auditory streaming is given by grating/plaid rivalry (Wallach, 1935), in which two overlaid gratings drift in opposite directions, with the same three main perceptual options as in auditory streaming (one grating in the foreground, the other grating in the foreground, and the integrated [plaid] percept) and a similar parametric control (by the angle between the gratings, their contrast, etc.). Indeed, there have been attempts to exploit this similarity to link visual and auditory multistability, although with limited to mixed success (Kondo et al., 2012; Pressnitzer & Hupé, 2006).

Most paradigms for studying multistable perception rely on the observers' subjective report: By pressing or holding a button or lever, participants indicate at any point in time which percept they are currently experiencing. Besides the necessity to trust the observers about the veridicality of their report, this poses additional challenges. For example, if a rivalry paradigm shall be combined with another rivalry paradigm or another task, it is hard to disentangle interactions on the perceptual side from interactions on the level of monitoring one's own perception and on the level of the response. Similarly, in situations in which multistability is combined with attentional or valuation paradigms, veridical report can become strategically suboptimal. Several strategies have therefore been proposed to circumvent the reliance on report.

First, catch trials can be introduced that replace the perceptual (endogenous) by a physical (exogenous) transition. One challenge for such catch trials is to mimic the perceptual transition as closely as possible to be indistinguishable from actual rivalry for the observer. In vision, this is virtually impossible for some classes of perceptual-rivalry stimuli, such as perspective reversals (e.g., the Necker cube), where either perceptual interpretation includes all stimulus components. Moreover, seemingly subtle stimulus changes that disambiguate such stimuli may have profound physiological consequences (Kornmeier & Bach, 2009). For binocular rivalry, specifically, transitions have complex spatiotemporal dynamics that correspond to traveling waves (Wilson, Blake, & Lee, 2001). Moreover, perceptual transitions from one state to another can occur without the switching itself being noticeable (Brascamp, Blake, & Knapen, 2015). This renders it difficult to match the perceptual experience of a transition in binocular rivalry by exogenous stimulation (see also Blake, Brascamp, & Heeger, 2014). Failure to do so, however, may be a substantial confound when comparing exogenous to endogenous switches (Knapen, Brascamp, Pearson, van Ee, & Blake, 2011). In audition, the catch-trial approach is likewise complicated by the fact that complete disambiguation of the stimulus is very difficult to achieve by means of unobtrusive physical parameter changes: For instance, increasing the intensity of one sequence over the other in an "ABA. . ." streaming paradigm will increase the likelihood of that stream to be perceived in the foreground, yet dominance of the other stream remains a valid percept as long as the intensity difference is not so extreme as to make the manipulation overly salient (and hence no longer comparable with the ambiguous case). For this reason, catch-trial performance is often far from perfect (e.g., Farkas, Denham, Bendixen, & Winkler, 2016). In any case, even if the physical and perceptual switches are indistinguishable, it remains open whether both would have the same behavioral consequences, for example, similar reaction times from the transition to the overt response (Kornmeier & Bach, 2012). Another problem with the catch-trial approach is that it can be difficult to convey to participants the notion that in some parts of the experiment, there is "no correct answer," whereas in other parts, the experimenter checks the correctness of their responses and reinstructs them if necessary.

Second, observers can be required to perform a task on the stimulus, which they cannot accomplish unless they perceive a certain percept. In auditory streaming, a deviant detection task can be designed that can be accomplished only when a certain percept is held (Micheyl & Oxenham, 2010); by combining such tasks with electroencephalography (EEG), even passive versions are possible, in which no response by the observer is required (Sussman, Ritter, & Vaughan, 1999; Winkler et al., 2003). A similar idea can be applied in binocular rivalry: Targets are presented on the dominant and suppressed stimulus, and successful detection is used as indirect measure of dominance. This approach has, for example, proved useful to assess statistical properties of rivalry transitions (Alais, Keetels, & Freeman, 2014) or to validate the veridicality of report in the context of a reinforcement paradigm (Wilbertz, van Slooten, & Sterzer, 2014). However, to achieve a moment-by-moment readout, sampling of the target-detection task has to be dense, such that the task of reporting the dominant percept is replaced by the task of detecting the target, leaving little resources for combinations with other tasks and

prohibiting passive-viewing conditions. Moreover, the target-detection approach might be limited by the fact that in binocular rivalry, suppression of the "invisible" stimulus is rarely complete. Instead, suppression manifests itself in increased detection thresholds (e.g., Wales & Fox, 1970), and even contrast decrements in the suppressed stimulus (i.e., changes that decrease visibility further) can be detected (Ling, Hubert-Wallander, & Blake, 2010). Similar to the auditory modality, EEG can be used to determine changes in rivalry perception. However, physical changes that do not yield a perceptual change, specifically swapping the stimuli between the two eyes, can elicit responses that are similar to perceptual or perceived physical changes (van Rhijn, Roeber, & O'Shea, 2013). Such results render the use of task performance or related electrophysiological signals difficult for binocular rivalry. Likewise, in audition, the inference from task performance to percept is far from perfect: Observers can find strategies to solve the task without holding the required percept (Dowling, Lung, & Herrbold, 1987); vice versa, observers can fail to solve the task although the supportive percept is being held (e.g., because they fail to detect the deviant per se). Likewise, results of EEG-based testing show some dissociations between EEG data and behavioral task performance or perceptual reports (Bendixen, Schröger, Ritter, & Winkler, 2012; Spielmann, Schröger, Kotz, & Bendixen, 2014; Szalárdy, Winkler, Schröger, Widmann, & Bendixen, 2013). Hence, despite the elegance of the task-performance approach, it remains an indirect measure whose conformance with perception needs to be documented specifically for each paradigm.

Third, the content of the percept itself can be decoded by means of electrophysiological or imaging techniques. In vision, this has first been demonstrated for binocular rivalry between faces and houses, whose perceptual dominance differentially activates the fusiform face area and the parahippocampal place area, respectively (Tong, Nakayama, Vaughan, & Kanwisher, 1998). Similarly, each of the two rivaling stimuli can be "tagged" by a specific frequency, and the frequency of the dominant one is reflected more strongly in the steady-state visual evoked potential, both in EEG (Brown & Norcia, 1997) and magnetoencephalography (MEG; Tononi, Srinivasan, Russell, & Edelman, 1998). In audition, frequency tagging can be applied when setting up tone sequences with nontrivial rhythmic relations between the different perceptual alternatives (e.g., Pannese, Herrmann, & Sussman, 2015). In all of these cases, it is the content, rather than the transition that is decoded; this limits this procedure to specific stimuli or requires their tagging, which in itself may influence rivalry.

An important distinction has to be drawn between methods that allow near-perfect decoding on a moment-by-moment basis from those that merely show above-chance decoding on average across longer time periods. The latter also have their use, as they allow verifying—for example, in situations in which rewards are at stake—veridical report, at least on average per condition. The more reliable a cue is on a moment-by-moment basis, however, the closer it will reach the ultimate objective of continuously monitoring the subjective visual experience with objective measures. All of the aforementioned methods *for principled reasons* fall short of such a continuous-monitoring demand. This is either because the method is designed to probe the system (catch trials, task performance) only occasionally, requires averaging over time to achieve a sufficient signal-to-noise ratio (EEG, MEG), or is based on a comparably slow signal (functional magnetic resonance imaging [fMRI]). In binocular rivalry, this issue has been solved by using the direction of the optokinetic nystagmus (Enoksson, 1963; Fox, Todd, & Bettinger, 1975; Frässle, Sommer, Jansen, Naber, & Einhäuser, 2014; Marx & Einhäuser, 2015; Naber, Frässle, & Einhäuser, 2011): When two gratings are presented that drift in opposite directions, the optokinetic nystagmus (OKN) slow phase reliably follows the perceptually dominant grating. Similarly, when two stimuli of different luminance are used, the perceptual dominance can be reliably inferred from pupil size (Fahle, Stemmler, & Spang, 2011; Naber et al., 2011; see also Lorber, Zuber, & Stark, 1965). To date, no similar "no-report" paradigm has been available for auditory multistability.

In the present study, we aim at a first step toward a no-report paradigm for auditory streaming. Rather than searching a direct (peripheral) physiological correlate of auditory dominance, we propose a binocular-rivalry stimulus whose dominant percept is controlled by the currently dominant auditory percept. We test to what extent the currently dominant auditory percept can be read off from the OKN induced by the corresponding visual stimulus. Specifically, we presented two tone sequences and asked observers to continuously indicate which of them they perceived in the foreground. At the same time, we presented two horizontally drifting gratings, one to each eye, each of which "jumped" vertically in synchrony with the tones of one of the two sequences. We hypothesized that the OKN follows the grating corresponding to the tone that was perceived in the foreground. To test for effects of familiarity with the auditory stimulus and task, in Experiment 2, we added an additional 15 min of auditory task performance without visual stimulus prior to the audiovisual combination. In Experiments 1 and 2, tone sequences were isochronous; that is, intervals between the tones in each sequence were fixed. To test whether this rhythmicity of the auditory stimulus was critical to the hypothesized effect, in

Experiment 3, intervals between subsequent tones were varied randomly.

Methods

Participants

Twenty-four volunteers (age 19–29 years, $M \pm SD$: 22.79 ± 2.30; 14 female, 10 male) participated in the study, eight in each experiment. All were naïve to the purpose of the experiment, had normal color vision, and reported normal hearing. One participant reported a mild form of tinnitus that did not interfere with normal hearing and did not occur during the experiment. All participants gave written informed consent to participation. All procedures conformed to the principles laid out in the Declaration of Helsinki and were determined by the applicable body (*Ethikkommission der Fakultät für Human- und Sozialwissenschaften, TU Chemnitz*) to not require in-depth ethics evaluation.

Setup

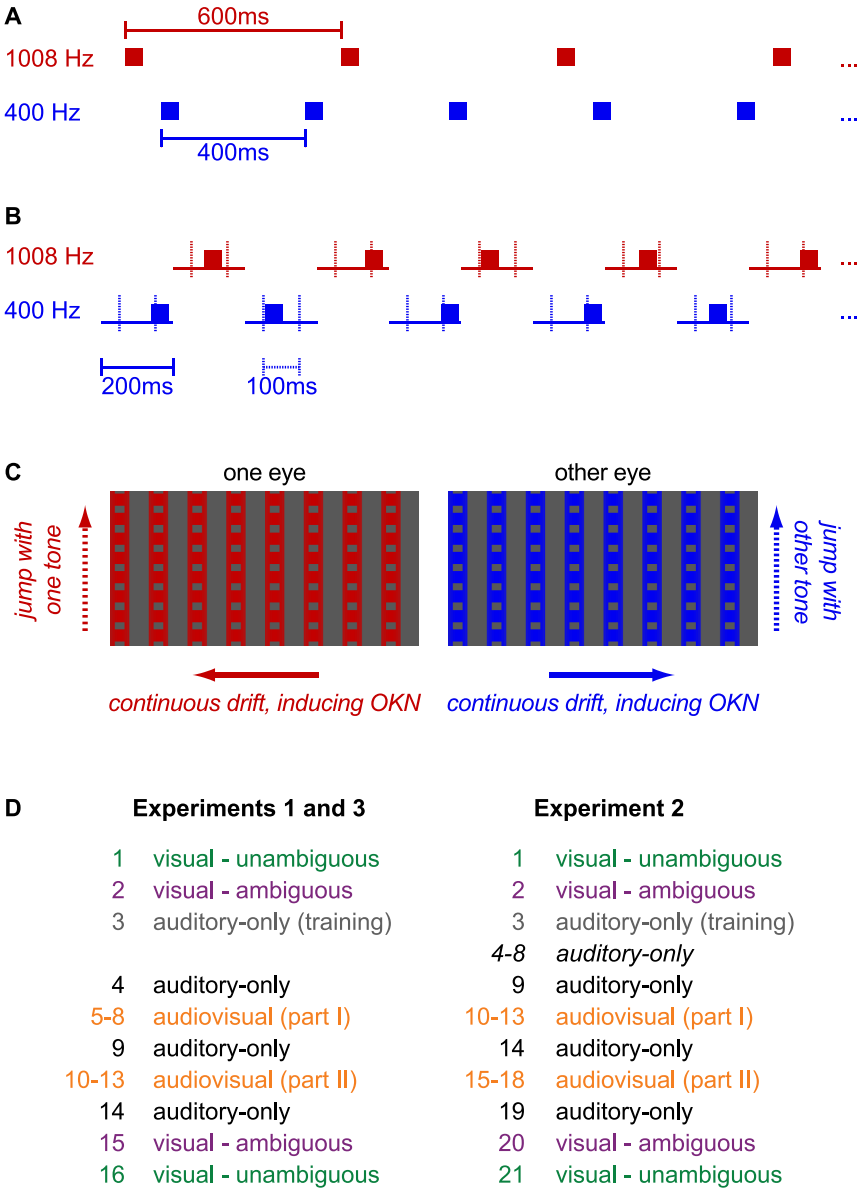
Visual stimuli were presented dichoptically at a viewing distance of 30 cm by means of a stereoscope, whose mirrors were transparent to infrared light ("cold mirrors") to allow eye tracking to be performed through the mirrors. Stimuli were displayed on two 21-in. CRT screens (Samsung, Seoul, South Korea) at a resolution of 1,024 × 768 pixels and a frame rate of 85 Hz. Auditory stimuli were presented diotically at 75 dB(A) delivered through calibrated Sennheiser HD 25-1 (70 Ω) headphones. The timing of visual and auditory stimulation was adjusted to account for the measured (and stable) latency of the sound presentation to be below one visual frame. Throughout the experiment, the participants' eye position was recorded binocularly at 500 Hz by an infrared eye-tracking device (Eyelink-1000, SR Research, Ottawa, ON, Canada). Because both eyes carry highly redundant information in this setting (Naber et al., 2011), data of only one eye were analyzed further. Participants entered their responses using the back two buttons (left/right) of a digital USB game pad and were instructed to press and hold those with their left and right index fingers, respectively. The state of the game pad was recorded together with the eye-position data. Experiments were conducted in a sound-attenuated room with no source of light or sound other than screens and the headphones. Stimulus presentation used Matlab (Mathworks, Natick, MA) with its Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002).

Auditory stimulus and task

Auditory stimuli were variants of the auditory-streaming stimulus (van Noorden, 1975): Two sequences of sine tones spaced 16 semitones apart (400 Hz and 1008 Hz) were presented. Tones were 50 ms in duration with 5-ms raised-cosine onset and offset amplitude ramps. In Experiments 1 and 2, the tones within each of the two sequences were presented at onset asynchronies of 400 ms and 600 ms, respectively, such that the onset asynchrony of distinct tones across the sequences was 100 ms, −100 ms, 300 ms, or −300 ms (Figure 1A). The association of pitch (400 Hz/1008 Hz) with interval (400 ms/600 ms) was counterbalanced across observers but fixed within each individual. In Experiment 3, tone onsets were placed in the 100-ms center of a 200-ms interval, with 200-ms intervals alternating between tones (Figure 1B). Placement within the 100 ms was random, with a uniform distribution. This led to an alternating sequence of the two tones with a minimum spacing of 50 ms (interstimulus interval, i.e., offset to the next onset) and a maximum spacing of 250 ms between two subsequent tones. In all experiments, participants were asked to report the tone they perceived in the foreground by pressing and holding the respective button. The association between button and tone was fixed in each individual but counterbalanced across participants. They were further instructed to press and hold both buttons whenever they experienced both tones in the foreground (including an integrated percept), as well as no button whenever the percept was ambiguous or distinct from the available options.

Visual stimulus

The visual stimulus consisted of two gratings, each presented centrally to one eye. Both gratings were isoluminant (10 cd/m^2) vertical square-wave gratings: one consisted of a red and a gray phase and the other of a blue and a gray phase. Gray and chromatic phases were equally wide; that is, the gratings had a 50% duty cycle. Color coordinates were (x = 0.623; y = 0.344) and (x = 0.151; y = 0.065), respectively. Gratings were 256 pixels (19.8°) high and 512 pixels (39.6°) wide with eight cycles on the full width (0.20 cyc/°). To minimize possible effects of temporal aliasing during grating drift, the transition between the chromatic and gray phase of the square wave was slightly smoothed: The luminance of each chromatic phase was gradually reduced from its central peak to the color/gray transitions on each side by 30% following an inverse square function. Each grating drifted horizontally at 240 pixels/s (18.4°/s), with different directions (left/right) between the two eyes; that is, either both gratings



D

Experiments 1 and 3		Experiment 2	
1	visual - unambiguous	1	visual - unambiguous
2	visual - ambiguous	2	visual - ambiguous
3	auditory-only (training)	3	auditory-only (training)
		4-8	*auditory-only*
4	auditory-only	9	auditory-only
5-8	audiovisual (part I)	10-13	audiovisual (part I)
9	auditory-only	14	auditory-only
10-13	audiovisual (part II)	15-18	audiovisual (part II)
14	auditory-only	19	auditory-only
15	visual - ambiguous	20	visual - ambiguous
16	visual - unambiguous	21	visual - unambiguous

Figure 1. Stimuli and procedure. (A) Illustration of the auditory stimulus used in Experiments 1 and 2. Tones of 50-ms duration were spaced 600 ms apart in one sequence and 400 ms in the other, with ±100-ms shift between the sequences. The association of interval (400/600 ms) to frequency (400 Hz/1008 Hz) was counterbalanced across observers. (B) Illustration of the auditory stimulus used in Experiment 3. Tone onsets were randomly placed alternating between low and high tone in the central 100 ms of 200-ms intervals. (C) Visual stimulus. Gratings drifted continuously in the horizontal direction and jumped vertically, each in synchrony with one of the tones. A movie of the stimulus including the sounds and the checkerboard frame is available as supplemental material. (D) Order of conditions. Additional auditory-only blocks in Experiment 2 result in the same experience with the auditory task in audiovisual Part I of Experiment 2 as in audiovisual Part II of Experiment 1.

drifted inward or both outward (for half of the participants, drift was inward; for the other half, it was outward). On the chromatic phase of the grating, there were 16 × 12 pixel wide rectangles of the same gray as the grating's achromatic phase. Rectangles were horizontally centered on the chromatic phase and spaced 32 pixels apart vertically (Figure 1C). Each grating "jumped" by 16 pixels vertically in synchrony with one of the tones (Figure 1C and Supplementary Video S1). The association between tone and grating was balanced across four subsequent blocks in each observer (see the Procedure section for details). To provide a fixed depth plain, gratings were surrounded by a background of random white (85 cd/m^2) or black (<0.1 cd/m^2) checks of width 16 pixels (1.3°), which was identical in both eyes.

Procedure

Experiments 1 and 3 consisted of 16 blocks and Experiment 2 of 21 blocks (Figure 1D). At the start of each block, the eye tracker was calibrated with a standard nine-point procedure, and calibration was validated. Participants were encouraged to take short breaks between blocks. In all experiments, there were eight "*audiovisual*" *blocks* of 180 s each that used the combined stimulus and the task as described above. In Experiments 1 and 3, these were Blocks 5–8 and 10–13; in Experiment 2, these were Blocks 10–13 and 15–18. Within each set of four blocks, all four combinations of eye and color with tone were used in random order. For the analysis of audiovisual blocks, the first set of four blocks will be referred to as first part and the second set of four blocks as the second part. In addition to the audiovisual blocks, each experiment started and ended with a block in which one of the gratings was presented to each eye and the other eye was presented a blank gray surface in lieu of the grating. In these "*visual-unambiguous*" *blocks*, sounds were silenced (vertical jumps were still present, as if the tones would sound), and the association of eye, color, and 400 ms/600 ms jumps (in Experiments 1 and 2) was changed every 30 s and covered all eight possible combinations in random order (240 s in total per block). The second and second-to-last block in each experiment was a "*visual-ambiguous*" *block* that matched the audiovisual blocks except that the tones were silenced and the association of color and 400 ms/600 ms to eye changed every 30 s. In visual-ambiguous and visual-unambiguous blocks, the observers had no task but to look at the visual stimulus. The remaining blocks (3, 4, 9, and 14 in Experiments 1 and 3; 3–9, 14, and 19 in Experiment 2) were "*auditory-only*" *blocks*, in which the gratings were replaced by a gray surface. The task in auditory-only blocks was identical to the audiovisual blocks as described above. Block 3 in all experiments was considered a pure training/familiarization block that was not used for further analysis; the other auditory-only blocks were used for comparison between reported switching patterns in auditory and audiovisual blocks. In one participant of Experiment 1, Block 3 was repeated as the responses during the block and immediate question by the experimenter revealed a misconception of the report/percept association.

Analysis

OKN slow phase

All eye-movement analyses were based on the slow phases of the OKN induced by the perceptually dominant drifting grating. Because OKN fast phases have similar velocity profiles to saccades, fast phases of the OKN were determined from the raw eye-position data (Figure 2A) using the system's saccade detection algorithm with thresholds of 35°/s for eye velocity and 9,500°/s^2 for eye acceleration. These periods, as well as periods of eye blinks, were treated as missing data for analysis (Figure 2B). In each continuous period of the remaining horizontal eye-position data, a linear function was fit (Figure 2C). The slope of this fit corresponds to the eye velocity for this particular period. The gain of the OKN was then defined as the thus determined velocity divided by the horizontal speed of the grating. For visual-unambiguous blocks, the sign of the gain was defined to be positive in the direction of the presented grating. For the audiovisual blocks, the gain was defined as positive when it matched the direction of the grating corresponding to the low-pitch (400 Hz) tone (Figure 2D).

Sensitivity (d′) and signed gain

For periods in audiovisual blocks in which one stream was reported uniquely in the foreground, we analyzed how well the OKN represented the auditory percept. These analyses were based on two complementary measures of readout success.

The first measure was based on signal detection theory (SDT) and quantifies how well the direction of the OKN can discriminate between perceiving the low-pitch or the high-pitch tone in the foreground. We used participants' report as ground truth and define hits, false alarms, correct rejections, and misses for each time point as in standard SDT (Green & Swets, 1966; Macmillan & Creelman, 2005):

- Hit: Low-pitch tone is reported in the foreground and the gain is positive.
- Miss: Low-pitch tone is reported in the foreground and the gain is negative.
- False alarm: High-pitch tone is reported in the foreground and the gain is positive.
- Correct rejection: High-pitch tone is reported in the foreground and the gain is negative.

Note that the assignment of hits/correct rejections is arbitrary and could be swapped, provided misses and false alarms are also swapped. As usual, the hit rate was then defined as

$$\text{hit rate} = \text{hits}/(\text{hits} + \text{misses})$$

$$\text{false alarm rate} = \text{false alarms}/(\text{false alarms} + \text{correct rejections})$$

and the sensitivity as

$$d' = z(\text{hit rate}) - z(\text{false alarm rate})$$

A d′ significantly larger than 0 would imply above-chance decoding of the foreground tone from the OKN data.

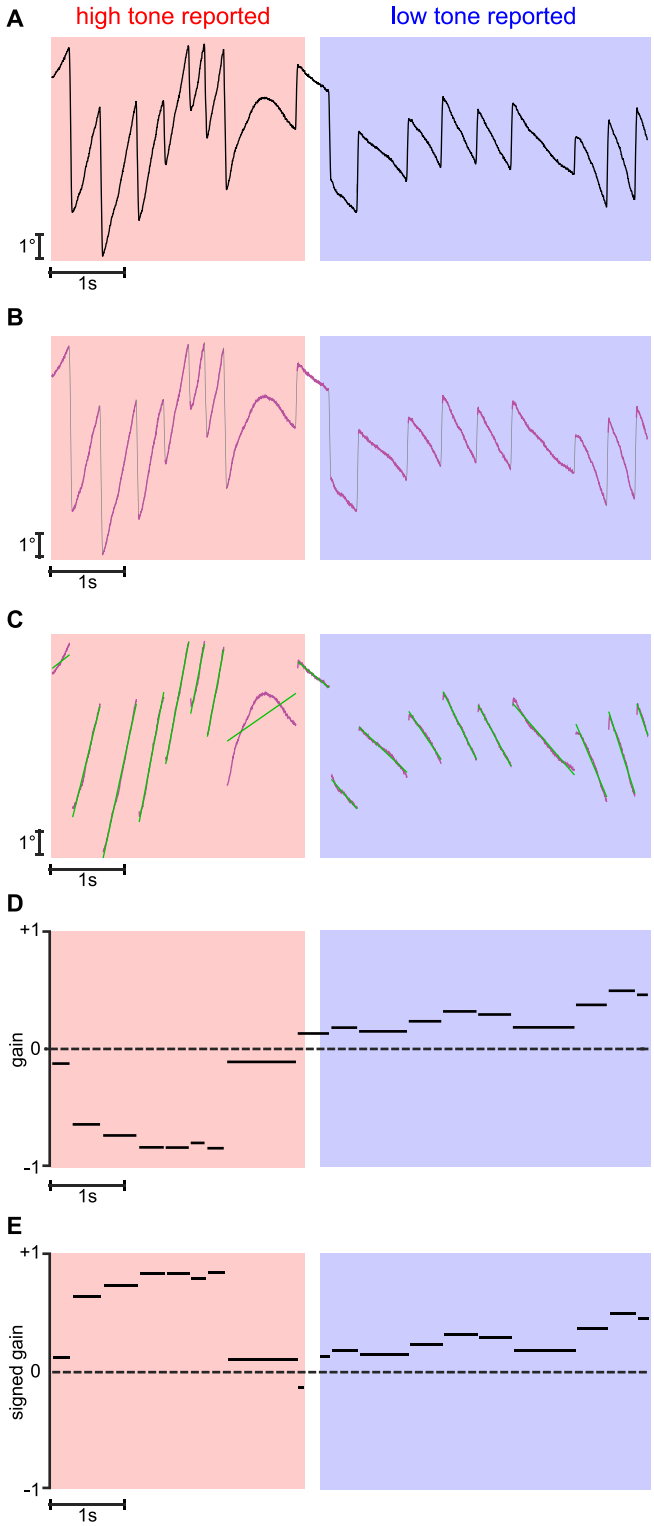


Figure 2. Determining optokinetic nystagmus (OKN) gain and signed gain from raw data. (A) Excerpt of horizontal eye-position trace of an audiovisual block. The red box indicates a report of high tone, blue box report of low tone, white interval report of an undefined percept (no button pressed). (B) The signal of panel A separated into slow phases of the OKN (magenta) and fast phases (gray). (C) Fast phases were removed from the signal and a linear function (green) fitted to each slow phase. (D) Slopes of linear fits divided by stimulus speed define the gain of the OKN. Sign was defined as positive for the direction of the grating that was associated with the low-pitch tone (irrespective of report). (E) Signed gain: For periods in which the percept of the high tone in the foreground was reported, the sign of the gain was inverted; periods with no clear foreground percept of one of the tone sequences were removed from analysis. Signed gain is thus positive if report and OKN match, negative otherwise.

The second measure, which hereafter will be referred to as "signed gain," assigns a positive sign to the gain whenever the OKN slow phase has the same direction as the grating matching the currently reported auditory percept and a negative sign otherwise (Figure 2E).

For example, if the blue grating drifts to the right and its jumps are in synchrony with the low-pitch tone, the signed gain would be

- positive, if the slow phase is directed to the right and the observer reports perceiving the low-pitch tone in the foreground;
- positive, if the slow phase is directed to the left and the observer reports perceiving the high-pitch tone in the foreground;
- negative, if the slow phase is directed to the left and the observer reports perceiving the low-pitch tone in the foreground; and
- negative, if the slow phase is directed to the right and the observer reports perceiving the high-pitch tone in the foreground.

If the OKN slow phase matches the auditory report as hypothesized, the average signed gain will be significantly larger than 0. The theoretical upper limit would be an average signed gain of +1, implying that OKN perfectly follows the auditory percept. The practical upper limit, which takes the limits of the oculomotor system into account, would be to match the gain of the visual-unambiguous block. Unlike d′, the signed gain does not only consider the direction of the eye velocity but also its absolute speed.

Absolute gain

The subjective perceptual experience of a dominant stimulus in binocular rivalry may differ from the experience when viewing the same stimulus in isolation. Moreover, in particular with large stimuli, patches of the nondominant stimulus can locally become visible or dominant, leading to mixed percepts ("piecemealing"). As a possible measure of the degree of dominance (i.e., the vividness of the perceptual impression or its exclusiveness), the absolute value of the OKN gain (i.e., speed irrespective of direction) was compared between the different conditions. Because, for the audiovisual blocks, we expected less vivid visual percepts for periods in which both tones were perceived in the foreground as compared with periods in which a single tone was perceived in the foreground, we analyzed these two phases separately.

Dominance durations

For auditory-only and audiovisual blocks, mean dominance durations of the three percepts (high tone in the foreground, low tone in the foreground, both tones in the foreground) were determined for each individual based on the button-press data. Dominance durations smaller than 100 ms were excluded from this analysis (8.7% of dominance periods) as they likely resulted from motor limitations in changing between the buttons. In addition, the relative dominance of each of the three percepts was defined as the aggregate time for which the percept was reported divided by the total viewing time of the respective condition.

To compare the durations of the visual percepts (for which no button-press data are available) in visual-ambiguous blocks to audiovisual blocks, the changes in the direction (sign) of the OKN slow phase were determined. The time between two subsequent sign changes was then analyzed in lieu of reported perceptual transitions. For consistency with the analysis of report-based auditory dominance durations, periods smaller than 100 ms (2.4% of OKN-defined periods) were excluded from this analysis.

Time course of OKN gain relative to perceptual transitions

To analyze the time course of the OKN gain relative to the report of perceptual transitions by button press, we determined all time points at which the high-tone stream started to be perceived in the foreground and the percept reported immediately prior had been the low-tone stream in the foreground (hereafter "transition low tone to high tone"), as well as all time points with the converse transition ("transition high tone to low tone"). To allow for some glitch in the motor response when switching from holding one button exclusively to holding the other button exclusively, reports of both streams in the foreground (both buttons pressed) or no report (no button pressed) between the high-tone and low-tone report were ignored if their duration was less than 100 ms. With these criteria, 21 of 24 observers showed such direct high-tone-to-low-tone or low-tone-to-high-tone transitions and were included in this analysis. Time Point 0 was defined as the onset of the new percept, that is, the point in time from which exactly one button was pressed. The gain traces were aligned at this time point. Aligned traces were first averaged separately in each individual and then the resulting mean traces were averaged across individuals, such that each individual contributed with equal weight. Before averaging, gain traces were cropped at the preceding and subsequent perceptual transition and the remaining time points treated as missing data for the particular trace. Hence, except for the transition time between $t = -100$ ms and $t = 0$, only data were averaged for which either the high or the low tone was reported in the foreground. Mean traces were compared statistically at each time point with a paired t

test; to correct for multiple comparisons, the alpha level was adjusted to an expected false discovery rate of 5% following the procedure by Benjamini and Hochberg (1995). An analogous procedure was applied to compare the transitions from either high-tone or low-tone percept to the percept with both tones in the foreground ("low to both", "high to both") and of the reverse transitions ("both to low", "both to high"). All observers except one, who had only a single "both tones in the foreground" percept in the audiovisual condition, contributed data to this analysis.

Results

OKN gain in the visual-unambiguous blocks

Across all experiments, the gain in the visual-unambiguous blocks ranged between 0.49 and 0.92 ($M \pm SD$: 0.71 ± 0.14), with a trend to decrease from the beginning to the end of the experiment, $t(23) = 1.99$, $p = 0.06$. This shows that—albeit not always close to the perfect value of 1—the gratings used in the present set of experiments induced a reliable OKN.

Absolute OKN gain

Although, by definition, the mean absolute gain must be larger than or equal to the mean gain, the numerical difference during unambiguous blocks was negligible (0.72 ± 0.14 compared with the 0.71 ± 0.14 given above). This verifies that the OKN slow phase is nearly always in the direction of the unambiguously presented grating. For the audiovisual blocks, the absolute gain dropped to 0.51 ± 0.22 for periods in which a single tone was perceived in the foreground and was thus smaller than for the unambiguous visual blocks, $t(23) = 8.72$, $p < 0.001$. This may indicate that the dominant visual percept in rivalry is not as vivid as an isolated stimulus. However, the correlation between the absolute gains in the two conditions across observers was large, $r(22) = 0.86$, $p < 0.001$, suggesting that the gain was also limited by idiosyncratic oculomotor factors. The absolute gain dropped further to 0.40 ± 0.21 for the periods in which both tones are perceived in the foreground [difference to single tone in the foreground, $t(23) = 4.11$, $p < 0.001$], although the correlation across observers with the absolute gain in the unambiguous visual condition remained high, $r(22) = 0.73$, $p = 0.001$. When comparing the absolute gains of the audiovisual blocks to the absolute gains in the visually ambiguous blocks (0.42 ± 0.17), the absolute gains when reporting a single tone in the foreground were significantly larger, $t(23) = 4.02$, $p < 0.001$, whereas the absolute gains when reporting both tones in the foreground were statistically indistinguishable from the visually ambiguous blocks without auditory task, $t(23) = 0.51$, $p = 0.62$.

Relative dominance and dominance durations across conditions

In the auditory-only and the audiovisual conditions, auditory percepts were indicated by button presses. We found no evidence that relative dominance or dominance durations differed between these conditions (Table 1, top and middle), suggesting that the presence of the visual stimulus had no evident impact on the statistics of auditory multistability. Conversely, when comparing the visual-ambiguous to the audiovisual conditions, the presence of the auditory stimulus significantly prolonged the time between sign changes in the OKN slow phase (Table 1, bottom). This suggests that the auditory stimulus or task exerted an influence on visual dominance as reflected in the OKN direction.

Readout of auditory percept from OKN gain in audiovisual blocks

Experiment 1

In the audiovisual blocks of Experiment 1, the sensitivity was significantly different from 0, $t(7) = 2.54$, $p = 0.04$ (Figure 3A), as was the signed gain, $t(7) = 2.41$, $p = 0.046$ (Figure 3B), which indicates successful readout of the auditory percept from the OKN data. To test for effects of training and/or familiarity with the stimulus, we split the analysis into the first and second parts of four audiovisual blocks each. We found significant deviations from zero for the second audiovisual part in sensitivity, $t(7) = 3.01$, $p = 0.02$, and signed gain, $t(7) = 2.58$, $p = 0.04$, but not for the first part, $t(7) = 1.11$, $p = 0.30$, and $t(7) = -0.12$, $p = 0.91$. Indeed, there was a significant difference between the first and the second parts for sensitivity, $t(7) = 2.39$, $p = 0.048$ (Figure 3C). Signed gain shows a similar trend for an increase from the first to the second part, $t(7) = 2.33$, $p = 0.052$ (Figure 3D). This is remarkable in view of the overall trend for the gain to decline over the course of the experiment (cf. analysis of the visual-unambiguous blocks). In sum, Experiment 1 showed above-chance relations between the auditory percept and the OKN as hypothesized. Those relations tended to increase over the course of the experiment. This raises the question as to whether it is an increase in the audiovisual coupling or whether the increase results from an increasing familiarity with the auditory stimulus and the corresponding task.

	Condition	Experiment 1	Experiment 2[a]	Experiment 3	All experiments
Relative dominance (button press)					
Low tone	Auditory-only	30.1% ± 15.4%	17.1% ± 15.0%	39.8% ± 11.5%	29.0% ± 16.4%
	Audiovisual	27.9% ± 15.7%	22.4% ± 15.6%	43.9% ± 11.8%	31.4% ± 16.7%
	Difference	$t(7) = 0.78$	$t(7) = 1.56$	$t(7) = 1.36$	$t(23) = 1.32$
		$p = 0.46$	$p = 0.16$	$p = 0.22$	$p = 0.20$
High tone	Auditory-only	24.0% ± 13.6%	22.4% ± 10.7%	37.0% ± 12.1%	27.8% ± 13.5%
	Audiovisual	29.9% ± 14.9%	20.4% ± 10.2%	32.8% ± 12.0%	27.5% ± 13.1%
	Difference	$t(7) = 1.99$	$t(7) = 0.75$	$t(7) = 1.43$	$t(23) = 0.19$
		$p = 0.09$	$p = 0.48$	$p = 0.20$	$p = 0.85$
Both tones	Auditory-only	37.3% ± 15.9%	59.3% ± 25.6%	20.9% ± 19.2%	39.2% ± 25.4%
	Audiovisual	35.5% ± 15.9%	55.5% ± 25.0%	20.5% ± 18.7%	37.1% ± 24.2%
	Difference	$t(7) = 0.56$	$t(7) = 1.02$	$t(7) = 0.22$	$t(23) = 1.17$
		$p = 0.59$	$p = 0.34$	$p = 0.83$	$p = 0.25$
Mean dominance duration (s)					
Low tone	Auditory-only	10.3 ± 11.6	8.0 ± 4.4	15.2 ± 19.3	11.2 ± 13.0
	Audiovisual	7.6 ± 5.7	8.9 ± 7.2	21.0 ± 7.2	12.5 ± 20.3
	Difference	$t(7) = 1.16$	$t(7) = 0.62$	$t(7) = 1.07$	$t(23) = 0.65$
		$p = 0.28$	$p = 0.56$	$p = 0.32$	$p = 0.52$
High tone	Auditory-only	5.9 ± 3.2	11.5 ± 8.4	15.8 ± 22.2	11.1 ± 13.9
	Audiovisual	7.2 ± 4.4	8.7 ± 5.2	12.7 ± 14.8	9.5 ± 9.3
	Difference	$t(7) = 1.62$	$t(7) = 1.42$	$t(7) = 1.13$	$t(23) = 1.31$
		$p = 0.15$	$p = 0.20$	$p = 0.30$	$p = 0.20$
Both tones	Auditory-only	12.2 ± 10.5	23.2 ± 18.6	3.0 ± 1.3	13.2 ± 14.7
	Audiovisual	10.7 ± 7.9	22.8 ± 17.4	3.0 ± 2.0	12.5 ± 13.6
	Difference	$t(7) = 0.73$	$t(7) = 0.07$	$t(6) = 0.02$[b]	$t(22) = 0.29$[b]
		$p = 0.49$	$p = 0.95$	$p = 0.99$	$p = 0.77$
Mean period between OKN sign changes (s)	Visual-ambiguous	1.8 ± 0.5	1.9 ± 0.5	2.5 ± 0.6	2.1 ± 0.6
	Audiovisual	2.5 ± 0.9	2.5 ± 0.8	3.6 ± 1.0	2.9 ± 1.0
	Difference	$t(7) = 3.08$	$t(7) = 3.69$	$t(7) = 3.47$	$t(23) = 5.65$
		$p = 0.02$	$p = 0.008$	$p = 0.01$	$p < 0.001$

Table 1. Relative dominance and dominance durations across conditions. *Notes*: Top: Comparison of relative dominance of each auditory percept (low tone in the foreground, high tone in the foreground, both tones in the foreground) between auditory-only and audiovisual conditions based on button-press data. Middle: Comparison of mean dominance durations for each auditory percept between auditory-only and audiovisual conditions based on button-press data. Bottom: Comparison between visual-ambiguous and audiovisual conditions of mean durations between sign changes of optokinetic nystagmus (OKN) as a possible proxy of a perceptual transition. Values denote means and standard deviations across observers for the respective experiment; statistics reported refer to paired *t* tests. Right column: Aggregated data over all experiments. [a]For analysis of Experiment 2, auditory-only Blocks 4–9, 14, and 19 were used. Restricting analysis to the same number of auditory-only blocks as in Experiments 1 and 3 (i.e., using 9, 14, and 19) did not alter the results qualitatively. [b]One participant of Experiment 3 did not report "both tones in the foreground" in the auditory-only block and had one single instance of a "both-tones in the foreground" percept (of 104 ms) in the audiovisual blocks. This observer was thus excluded from this analysis of the "both-tones in the foreground" percept. Hence, the degrees of freedom for the *t* test were 22 (or 6) for this comparison and 23 (or 7) for the other two comparisons.

Experiment 2

To address the question whether familiarity with auditory stimulus and task suffices for the improvement of the OKN readout observed in Experiment 1, in Experiment 2 we added five more auditory-only blocks prior to the first audiovisual block (Figure 1D). Aggregated over all audiovisual blocks, sensitivity was significantly different from 0, $t(7) = 3.76$, $p = 0.007$ (Figure 4A), as was the signed gain, $t(7) = 3.63$, $p = 0.008$ (Figure 4B). Both measures were significantly different from 0 already for the first audiovisual part, Blocks 10–13, d′: $t(7) = 2.70$, $p = 0.03$ (Figure 4C); signed gain: $t(7) = 3.39$, $p = 0.01$ (Figure 4D), and remained so for the second audiovisual part, Blocks 15–19; d′: $t(7) = 3.61$, $p = 0.009$; signed gain: $t(7) = 3.39$, $p = 0.01$. Between the first and second audiovisual parts, we found little evidence for changes in the relevant measures: d′: $t(7) = 2.11$, $p = 0.07$; signed gain: $t(7) = 1.50$, $p = 0.18$.

When comparing the first audiovisual part of Experiment 2 to the second audiovisual part in Experiment 1 (i.e., when comparing participants who

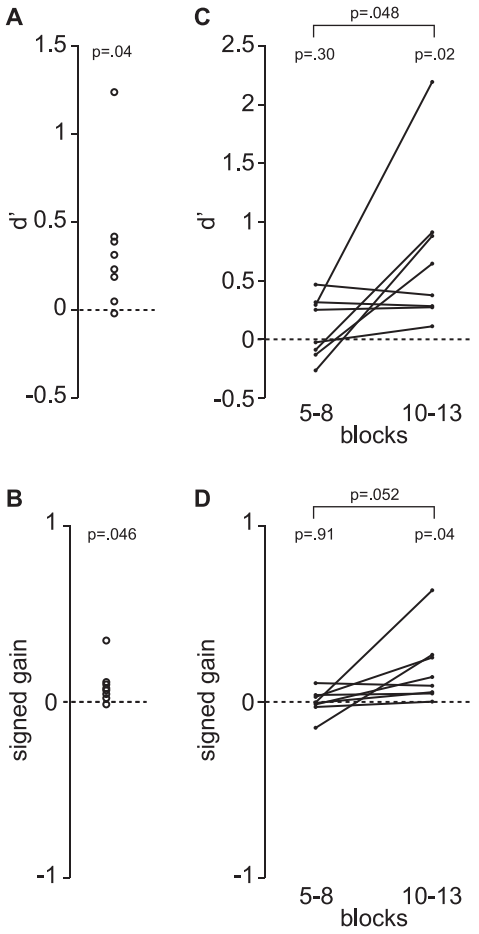


Figure 3. Results of Experiment 1. Sensitivity (panels A and C) and signed gain (panels B and D) for determining the auditory percept from the OKN data. Data in A and B are aggregated over all audiovisual blocks; data in C and D separate first and second parts of the respective experiment (four blocks each). Each data point (A and B) or line (C and D) corresponds to one individual. The *p* values on top of each column refer to *t* tests on the null hypothesis that the data are not different from 0; *p* values on top of panels C and D to paired *t* tests on the difference between left and right columns.

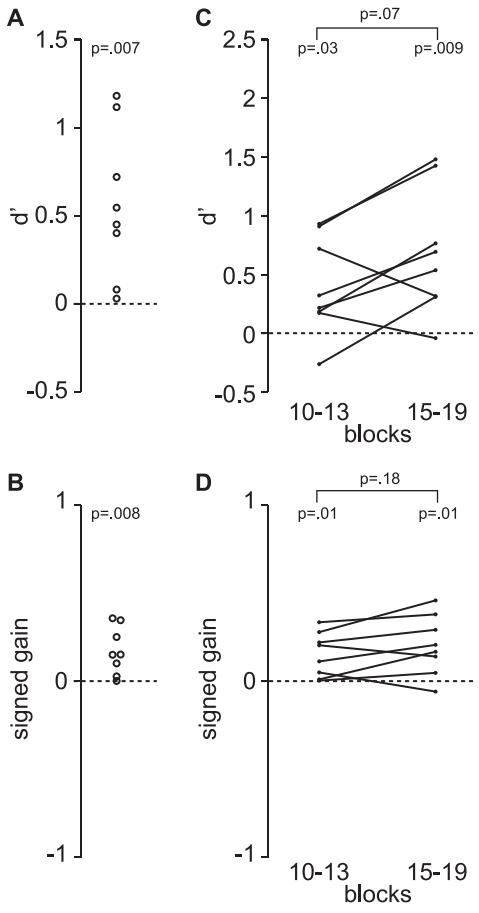


Figure 4. Results of Experiment 2. Notation as in Figure 3; in panel B some markers were shifted horizontally for readability.

have performed an equal amount of blocks on the auditory task), we found no difference between these parts, d′: $t(14) = 1.13$, $p = 0.28$; signed gain: $t(14) = 0.42$, $p = 0.68$. This is unlikely a consequence of lack in statistical power, as between the first audiovisual part of Experiment 1 and the first audiovisual part of Experiment 2, there was a difference in signed gain, $t(14) = 3.00$, $p = 0.01$. Moreover, for both measures, there was a difference between the first audiovisual part of Experiment 1 and the second audiovisual part of Experiment 2, d′: $t(14) = 2.76$, $p = 0.02$; signed gain: $t(14) = 3.16$, $p = 0.007$. This between-experiment comparison suggests that the effect increase in Experiment 1 is mostly a consequence of the naïve observers getting more acquainted with the auditory stimulus and task rather than of the increased experience with the audiovisual stimulus.

Experiment 3

Experiment 3 tested whether the effects observed in Experiments 1 and 2 require the rhythmicity of the tone sequences. As an intended side effect, the lacking rhythmicity was expected to increase the amount of time the two tones were perceived as segregated (see the Discussion section) and thus to increase the amount of useable data. Indeed, the manipulation was effective in this respect: The fraction of time both tones were reported in the foreground decreased from 45.5% ± 22.7% of the total audiovisual presentation time in Experiments 1 and 2 to 20.5% ± 18.7% in Experiment 3, $t(22) = 2.68$, $p = 0.01$. Conversely, the fraction of time a single tone was reported to be perceived in the foreground, and thus the amount of useable data, increased from 50.0% ± 22.9% in Experiments 1 and 2 to 76.7% ± 17.7% in Experiment 3, $t(22) = 2.88$, $p = 0.009$.

Over all audiovisual blocks of Experiment 3, OKN was indicative of the dominant auditory percept according to the signed-gain measure and showed a

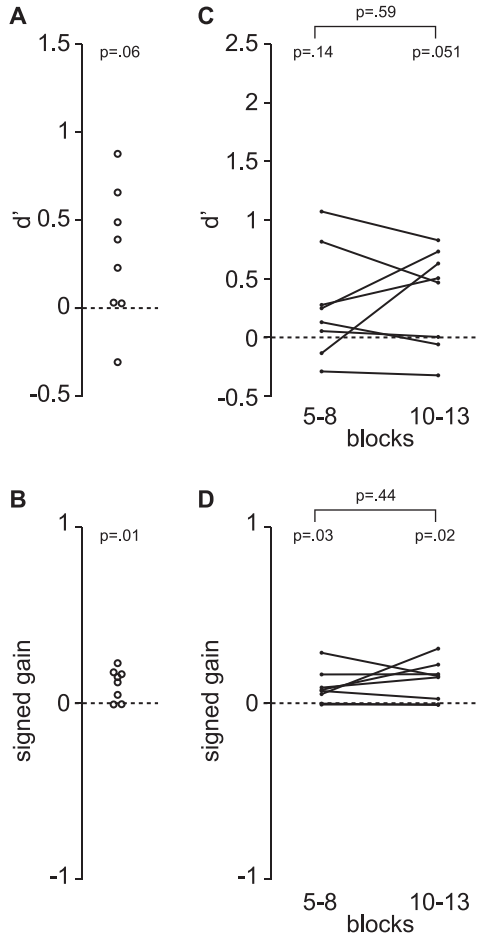


Figure 5. Results of Experiment 3. Notation as in Figures 3 and 4.

similar trend for sensitivity, d′: $t(7) = 2.21$, $p = 0.06$ (Figure 5A); signed gain: $t(7) = 3.51$, $p = 0.01$ (Figure 5B). In neither measure was any evidence of effect changes across the course of the experiment observed, d': $t(7) = 0.56$, $p = 0.59$ (Figure 5C); signed gain: $t(7) = 0.82$, $p = 0.44$ (Figure 5D). Hence, successful OKN readout does not require the rhythmicity of the stimulus.

As a final analysis to assess the potential usefulness of the OKN as an objective marker of auditory perception, we aggregated data over all audiovisual blocks of all three experiments. With the 24 observers in total, both measures are clearly different from 0, d′: $t(23) = 4.96$, $p < 0.001$; signed gain: $t(23) = 5.42$, $p < 0.001$, which underlines the robustness of the method across slight variations in the paradigm.

Time course of OKN gain relative to the report of transitions

All analyses so far related the OKN slow phase to the *current* report of the auditory percept. This ignores the time an observer requires to transform a change in their percept to a manual response. To analyze the relative time course of OKN to perceptual reports, the time course of the gain (defined positive for the grating corresponding to the low-pitch tone) was aligned to transitions between the two distinct "segregated" percepts. At reported transitions from "high tone in the foreground" to "low tone in the foreground," the gain increased (Figure 6A, blue), whereas the gain decreased for the reverse transition (Figure 6A, red). The difference between the transitions of different polarity (low-to-high vs. high-to-low) was minimal at 1.14 s before reporting the transitions, with significant differences reemerging 0.52 s prior to the button press (Figure 6B). These data are consistent with a reaction time of about 1 s from experiencing the perceptual transition to pressing the respective button. The shallow change is likely an effect of averaging (first over transitions, then over participants) and consistent with substantial jitter in the reaction time to a perceptual transition both within and across individuals. Besides transitions between the two percepts with a single tone in the foreground, there are transitions from a single foreground tone to both tones in the foreground and vice versa. When shifting to a percept with both tones in the foreground, the former background tone briefly biases the average OKN gain in its direction until the OKN on average remains unbiased by the tone (Figure 6C, D). This may indicate that extra perceptual evidence in favor of the suppressed tone needs to be recruited before it is included in the overt report. In turn, the transition from both tones in the foreground to a single tone in the foreground builds up slowly toward the time of reported transition (Figure 6E, F). Although readout of the current percept considered only the two percepts with a single tone in the foreground, the time-course analysis demonstrates that our OKN-based approach also allows assessing temporal properties of the transitions between all three percepts.

Discussion

In this study, we introduced and validated a paradigm that can be used to tag the dominant percept in auditory multistability by binocular rivalry. The dominant auditory percept can be read out from the OKN *in principle* on a moment-by-moment basis. Although the readout is still far from perfect, we demonstrated that readout success improves when observers become more acquainted with the auditory stimulus and task (Experiments 1 and 2) or without prior experience when the auditory report is easier (Experiment 3). The improvement in readout success in

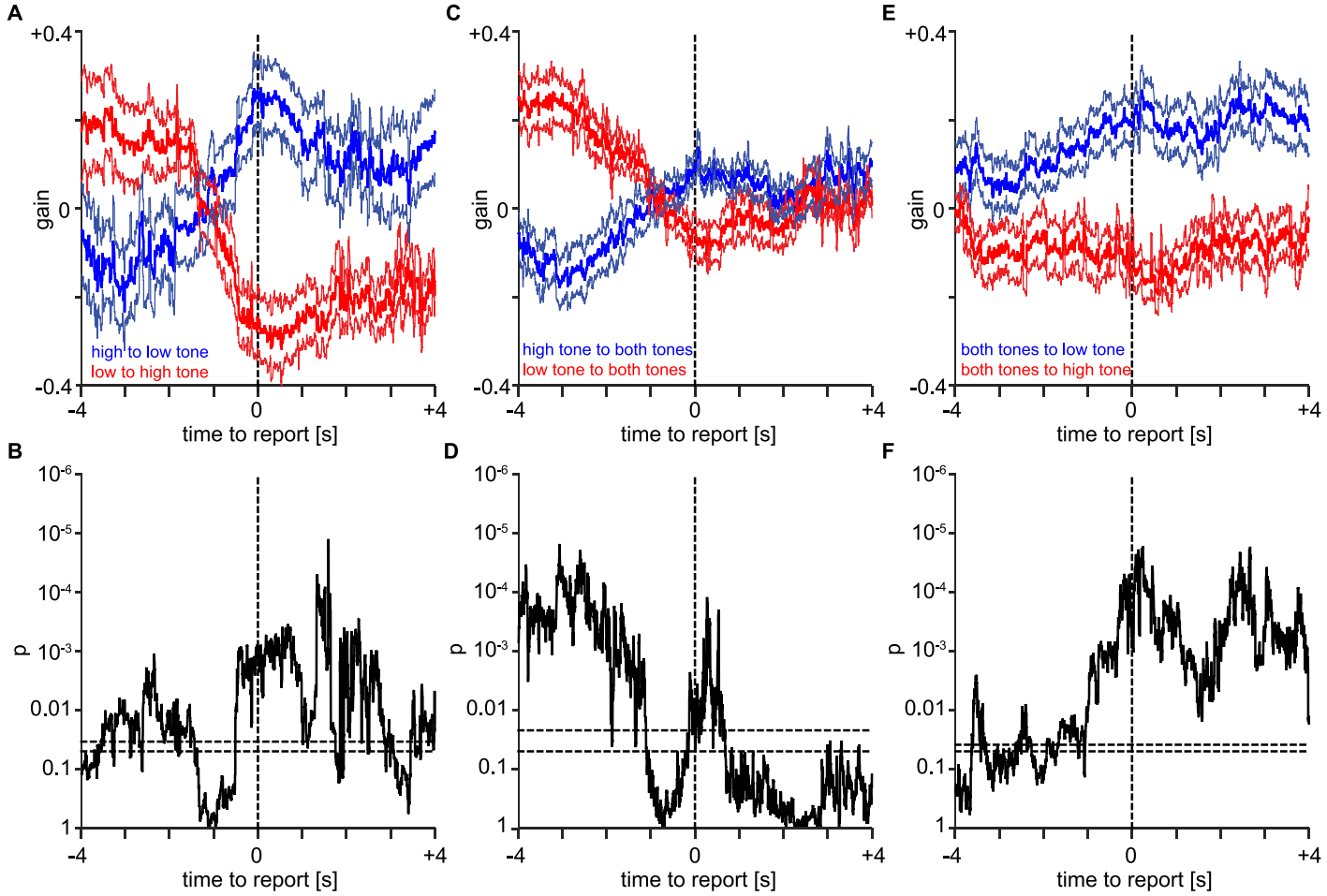


Figure 6. Time course of the gain of the optokinetic nystagmus (OKN) relative to time of report of perceptual change. (A, C, E) Time course of OKN gain; time 0 denotes the report of a transition. Thick lines: mean across observers; thin lines: *SEM* across observers. (B, D, F) Time-point by time-point comparison between red and blue traces of panels A, C, and E, respectively, by means of individual *t* tests; *p* values plotted on a logarithmic scale (upward implies lower *p* values). Lower dashed lines denote uncorrected 5% level, and the upper dashed line denotes the false discovery rate–corrected alpha level (i.e., the level corresponding to an expected false discovery rate of 5% according to the procedure by Benjamini and Hochberg, 1995). (A, B) Transitions from low to high tone in the foreground (red) or high to low tone in the foreground (blue). (C, D) Transitions from low tone in the foreground (red) or high tone in the foreground (blue) to both tones in the foreground. (E, F) Transitions from both tones in the foreground to either low tone in the foreground (blue) or high tone in the foreground (red).

turn suggests that subjective report might not provide a completely error-free ground truth. Hence, OKN readout success might be underestimated, especially in the early parts of each experiment. Importantly, we found no evidence that the presence of the visual stimulus influences auditory perception; that is, the readout mechanism is unlikely to have an effect on the quantity to be measured. In turn, the comparison to visual-ambiguous blocks (i.e., binocular-rivalry blocks without auditory stimulation) showed that the visual stimulus is under some control of the auditory percept, as intended. Finally, the readout success did not depend on rhythmicity, opening the paradigm to a potentially large class of auditory multistable phenomena; this distinguishes the present method, for instance, from tagging approaches that require specifically arranged rhythmic relations of the auditory stimulus streams (e.g., Pannese et al., 2015).

Although there is no *principled* reason for the suggested OKN-based approach not to yield a perfect readout of moment-by-moment dominance, it is self-evident that the present proof-of-principle study is only a first step toward such a no-report paradigm for auditory multistability. First, the readout success is robustly above chance but clearly below perfection. Second, the number of OKN sign changes exceeds the number of reported perceptual switches (Table 1), which indicates that determining the exact time of perceptual transitions from the OKN alone would be difficult at the present stage. Provided the time-course data (Figure 6), however, it is also conceivable that the transitions in auditory multistability indeed are not

abrupt but take some time to develop. Here, the OKN-based measure, possibly in combination with other techniques, can provide valuable insight into the dynamics of transitions. The present study clearly demonstrates that the three percepts (either tone uniquely in the foreground or both tones in the foreground) can be differentiated *on average* based on the OKN data.

One advantage of using an eye-position signal, which can be measured unobtrusively, as a measure of perceptual dominance in audition is the possibility of combining it with other measures that have been suggested for the validation of perception in auditory multistability. In particular, a combination with measures that are based on task performance is straightforward, as the eye movements as such are unlikely to interfere with an auditory task. Similarly, the catch-trial approach lends itself readily to combinations with OKN-based measurements; catch-trial segments might even be valuable in providing an upper limit of achievable readout success on a single-participant basis. In principle, OKN can also be combined with EEG-based measures, alone or in addition to task-based and catch-trial approaches. Such a design will have to take the challenges of correcting for eye-movement artifacts in EEG into account. Independent component analysis allows the efficient removal of artifacts caused by saccades (Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011), whose dynamics are similar to OKN fast phases. To our knowledge, this has not been attempted for slow smooth movements such as the OKN slow phase, but there is no principled reason that should preclude such an application. Similarly, for visual rivalry, OKN readout has successfully been combined with fMRI (Frässle et al., 2014), and it has recently been demonstrated that the combination of fMRI and OKN provides a more reliable readout than either measure alone (Ketkar, Wilbertz, & Sterzer, 2016). The combination with fMRI may thus also be useful for the present paradigm. In sum, although clearly beyond the scope of the present proof-of-principle study, the possibility of combining the present paradigm with other techniques to validate reports of perceptual experience are ample.

The use of eye movements as a marker of perceptual awareness has been subject to some debate (see Spering & Carrasco, 2015, for a review). One critical issue for the OKN measure specifically is that it typically precedes the button press signaling a perceptual switch (on the order of about half a second to a second in binocular rivalry, cf. Naber et al., 2011); hence, the OKN is sometimes considered too early to follow perception (cf. Spering & Carrasco, 2015). Yet, in fact, it is likely that the OKN gives a more accurate account of the timing of perceptual switching than the manual response. In visual multistability, this view is supported by observations that not only the OKN (Naber et al., 2011) but also other measures of transitions substantially precede manual report. These include the pupil dilation associated with a switch (Einhäuser, Stout, Koch, & Carter, 2008; Hupé, Lamirel, & Lorenceau, 2009; Kietzmann, Geuter, & König, 2011), pupil size changes when changing from a bright to a dark percept or vice versa (Fahle et al., 2011; Naber et al., 2011), event-related potential (ERP) components related to the perceptual transition (Kornmeier & Bach, 2012), and ERP components preceding a later percept in intermittently presented binocular-rivalry stimuli (O'Shea, Kornmeier, & Roeber, 2013). In auditory multistability, if anything, the time from the actual start of the perceptual transition to its report can be assumed to be longer than in vision, as by the discrete nature of the tones evidence has to be accumulated over a period of time. This is reflected in a typical approach in auditory multistability that discards responses in the initial phase (usually 1 s) of catch trials (Szalárdy et al., 2013) to allow for decision and response time. Some studies have reported average latencies of up to 2.2 s between the initiation of a catch-trial segment and the corresponding button press (Denham et al., 2014). Note that the response would be expected to be faster for catch trials than for actual rivalry trials; thus, the delay between multistable perceptual change and button press must be regarded as substantial. Hence, the time course of the OKN gain relative to the switch we observe in the present study can plausibly be interpreted by the OKN following the actual perceptual transition and the overt manual response being delayed by a further second on average.

Under the plausible assumption that the gain of the OKN is related to the vividness or exclusivity of the currently dominant percept, the analysis of the absolute gain suggests that when a single tone is perceived in the foreground, the dominant visual stimulus is perceived more exclusively or vividly than during the periods when both tones are perceived in the foreground. The exclusiveness does not reach the level of perceiving an unambiguous visual stimulus, however, and provided the size of the stimuli, it is likely that the gain is indeed reduced by local intrusions of the other, currently nondominant, grating (piecemeal percept). This is in line with the view that the percept of a single tone in the foreground strengthens the dominance of the corresponding percept as compared with the percept of both tones in the foreground. Other than this distinction in average gain, our current version of OKN-based readout of auditory multistable perception is limited in that it captures segregated percepts (distinguishing which of two streams is perceived in the foreground), although it cannot yield continuous OKN-based evidence for an integrated percept or for both tones

perceived simultaneously in the foreground (other than in an aggregate measure such as mean absolute gain). To exploit the potential of our OKN approach in a situation in which the distinction between the two segregated alternatives is more relevant than the distinction between segregation and integration, we introduced a substantial distance between the two tone sequences in frequency space (thereby increasing the probability of segregation, cf. Moore & Gockel, 2012). In Experiment 3, we further increased the time a single tone was perceived in the foreground by introducing random variations in timing separately in each stream, in contrast to the isochronous, fully predictable arrangement in Experiments 1 and 2. How such random variation affects auditory stream segregation and integration is a topic of current interest in auditory multistability (see Bendixen, 2014, for a review). It has been suggested that temporal predictability works as a cue toward stream integration (Rajendran, Harper, Willmore, Hartmann, & Schnupp, 2013), and despite unresolved issues on theoretical grounds (Bendixen, 2014), introducing unpredictable feature changes in both streams indeed tends to produce slight increases in segregation relative to a constant (thereby predictable) arrangement (Bendixen, Denham, & Winkler, 2014). In the current Experiment 3, this effect was boosted by decreasing the average temporal distance between tones from the two streams, which likewise acts in favor of stream segregation (Moore & Gockel, 2012). Together, those effects explain the higher proportion of unique segregated percepts in Experiment 3.

Unlike in Experiment 1, in Experiment 3, we observed a high readout success already in the first audiovisual part, although the experience with the auditory stimulus at this point was identical in both experiments. Part of this difference may be attributable to the reduced prevalence of the integrated percept. This does not only reduce a possible source of confusion for report in inexperienced observers, but also increases the experience with the percepts with either one tone in the foreground, because after the same period of time, the observers have experienced more segregated percepts in Experiment 3 than in Experiments 1 and 2. However, it is also conceivable that the irregularity of the sequences improved audiovisual coupling as such. Two *irregular* sequences of visual and auditory stimuli are optimally integrated into a multisensory percept if and only if both are correlated (Parise, Spence, & Ernst, 2012). This is understandable, given that the probability that *all* events of two *independent* irregular sequences co-occur is exceedingly low and approaches zero when the sequence length approaches infinity. Hence, strong correlations between irregular sequences of visual and auditory events (such as in Experiment 3) render their dependence likely and thus foster their cross-modal integration. In contrast, for two independent sequences that share the same rhythm, the co-occurrence of all their events is as probable as the co-occurrence of one pair of events (if one pair is synchronous, so are all others). Hence, the repetitive co-occurrence of events (e.g., tone and grating jump) is far stronger evidence against independence for irregular sequences (Experiment 3) than it is for rhythmic sequences (Experiments 1 and 2). Therefore, it is likely that synchronous irregular sequences in different modalities are bound more easily than synchronous rhythmic sequences. This implies a stronger audiovisual coupling for the irregular sequences of Experiment 3, which is consistent with our observations.

The fact that a visually ambiguous stimulus can be influenced by the auditory percept renders cross-modal effects in multistable perceptions conceivable. However, one must keep in mind that the present visual stimulus was explicitly designed to be controlled by auditory dominance and to maximize audiovisual coupling. Establishing OKN readout as possible no-report paradigm for auditory multistability is therefore a distinct question from studying auditory and visual multistability jointly. Studies with that aim typically focus on the question whether the same observer shows similar patterns in both modalities and so far have found little evidence for such relations, even for multistable phenomena that are similar across modalities (such as auditory streaming and plaid/grating rivalry; Pressnitzer & Hupé, 2006). Even studies that found evidence for such relations argue in favor of separate factors governing multistability in either modality (Kondo et al., 2012). Although the present study does not aim at addressing this issue, no-report paradigms as developed here might help overcome a central issue in multimodal multistability: the ability to track perception in two multistability paradigms in parallel without response interference between the two.

It is tempting to speculate as to how the coupling between visual and auditory stimulus is realized in the nervous system. Provided that the dominance durations are substantially longer than the intertone intervals, it seems unlikely that the tones themselves cause rivalry transitions, but once a transition has taken place, the tone perceived in the foreground might stabilize the current percept.[1] Temporal coincidence is a strong cue that signals from two modalities stem from the same source. In binocular rivalry, matched temporal frequency between visual signals and auditory or tactile signals biases perception, and this effect has been attributed to a supramodal binding mechanism (Lunghi, Morrone, & Alais, 2014). As directing attention toward features of a stimulus in binocular rivalry increases its dominance (Marx & Einhäuser, 2015; Ooi & He, 1999), it is also conceivable that the foreground tone directs attention to the corresponding stimulus and thereby yields the observed effect. A possible neural mechanism to bind different modalities is long-

range synchronization (cf. Engel, Fries, König, Brecht, & Singer, 1999). Indeed, it has recently been demonstrated that simultaneously directing attention to audition and vision increases alpha-band synchrony in the human EEG (van Driel, Knapen, van Es, & Cohen, 2014) as compared with directing attention to one modality. The extent to which audiovisual coupling, such as the one described here, is contingent on attention will be an issue for further research and might be an ideal application case for no-report paradigms.

In vision, no-report paradigms for multistability have sparked a substantial debate about their usefulness in assessing the physiological substrates underlying awareness or "consciousness" (Koch, Massimini, Boly, & Tononi, 2016; Overgaard & Fazekas, 2016; Tsuchiya, Frässle, Wilke, & Lamme, 2015, 2016). Irrespective of such considerations, no-report paradigms indisputably offer the practical advantage of making new experimental designs possible. As outlined above, these include the combination of multistable perception with other paradigms (e.g., attention or reinforcement-learning tasks), the simultaneous measurement of multistable phenomena in different modalities without response interference, as well as the combination with other approaches to objectify and validate perceptual experience. In conjunction with the complementary knowledge in the fields of auditory and visual multistability, the proposed no-report approach for auditory multistability may therefore provide an important step toward a truly multimodal understanding of multistable perception.

Supplemental material

Supplemental Movie. Audiovisual stimulus used in Experiments 1 and 2. The blue grating on the left jumps in synchrony with the low (400 Hz) tone at the 400-ms interstimulus interval (ISI), and the red grating on the right is in synchrony with the high (1008 Hz) tone at 600-ms ISI. Each grating was presented to a different eye through a stereoscope. Note that the movie is for illustration only; speed and background are not to scale, and audiovisual synchrony may depend on screen and sound card settings.

Keywords: multistability, *multimodal*, *binocular rivalry*, *no-report paradigm*

Acknowledgments

The authors thank Monique Michl for help with data collection.

Commercial relationships: none.
Corresponding author: Wolfgang Einhäuser.
Email: wolfgang.einhaeuser-treyer@physik.tu-chemnitz.de.
Address: Chemnitz University of Technology, Institute of Physics, Physics of Cognition Group, Chemnitz, Germany.

Footnote

[1]Given that the OKN gain direction change precedes the report of the switch (Figure 6), it is likely that an auditory perceptual transition quickly triggers the corresponding visual transition; this does not contradict the notion that *a single tone* itself does not trigger a transition in the present paradigm.

References

Alais, D., Keetels, M., & Freeman, A. W. (2014). Measuring perception without introspection. *Journal of Vision*, *14*(11):1, 1–8, doi:10.1167/14.11.1. [PubMed] [Article]

Bendixen, A. (2014). Predictability effects in auditory scene analysis: A review. *Frontiers in Neuroscience*, *8*, 60.

Bendixen, A., Denham, S. L., & Winkler, I. (2014). Feature predictability flexibly supports auditory stream segregation or integration. *Acta Acustica United With Acustica*, *100*, 888–899.

Bendixen, A., Schröger, E., Ritter, W., & Winkler, I. (2012). Regularity extraction from non-adjacent sounds. *Frontiers in Psychology*, *3*, 143.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. *Series B (Methodological)*, *57*, 289–300.

Blake, R., Brascamp, J., & Heeger, D. J. (2014). Can binocular rivalry reveal neural correlates of consciousness? Philosophical Transactions of the Royal Society of London. *Series B, Biological Sciences*, *369*, 20130211.

Boring, E. G. (1930). A new ambiguous figure. *American Journal of Psychology*, *42*, 444–445.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Brascamp, J., Blake, R., & Knapen T. (2015) Negligible fronto-parietal BOLD activity accompanying un-

reportable switches in bistable perception. *Nature Neuroscience*, *18*, 1672–1678.

Brascamp, J. W., Klink, P. C., & Levelt, W. J. (2015). The 'laws' of binocular rivalry: 50 years of Levelt's propositions. *Vision Research*, *109*, 20–37.

Breese, B. B. (1899). On inhibition. *The Psychological Review: Monograph Supplements*, *3*(1), i–65.

Brown, R. J., & Norcia, A. M. (1997). A method for investigating binocular rivalry in real-time with the steady-state VEP. *Vision Research*, *37*, 2401–2408.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*, 613–617.

Denham, S. L., Bőhm, T. M., Bendixen, A., Szalárdy, O., Kocsis, Z., Mill, R., & Winkler, I. (2014). Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. *Frontiers in Neuroscience*, *8*, 25.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, *140*, 552–572.

Dowling, W. J., Lung, K. M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, *41*, 642–656.

Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences, USA*, *105*, 1704–1709.

Engel, A. K., Fries, P., König, P., Brecht, M., & Singer, W. (1999). Temporal binding, binocular rivalry, and consciousness. *Conscious and Cognition*, *8*, 128–151.

Enoksson, P. (1963). Binocular rivalry and monocular dominance studied with optokinetic nystagmus. *Acta Ophthalmologica (Copenhagen)*, *41*, 544–563.

Fahle, M. W., Stemmler, T., & Spang, K. M. (2011). How much of the "unconscious" is just pre-threshold? *Frontiers in Human Neuroscience*, *5*, 120.

Farkas, D., Denham, S. L., Bendixen, A., & Winkler, I. (2016). Assessing the validity of subjective reports in the auditory streaming paradigm. *Journal of the Acoustical Society of America*, *139*, 1762–1772.

Fox, R., Todd, S., & Bettinger, L. A. (1975). Optokinetic nystagmus as an objective indicator of binocular rivalry. *Vision Research*, *15*, 849–853.

Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action, but not to perception. *Journal of Neuroscience*, *34*, 1738–1747.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Leipzig, Germany: L. Voss.

Hupé, J.-M., Lamirel, C., & Lorenceau, J. (2009). Pupil dynamics during bistable motion perception. *Journal of Vision*, *9*(7):10, 1–19, doi:10.1167/9.7.10. [PubMed] [Article]

Ketkar, M. D., Wilbertz, G., & Sterzer, P. (2016). Combined fMRI and eye-tracking based decoding of a bistable plaid motion perception. *Perception*, *45*(2 Suppl.), 277.

Kietzmann, T. C., Geuter, S., & König, P. (2011). Overt visual attention as a causal factor of perceptual awareness. *PLoS One*, *6*, e22614.

Klink, P. C., van Ee, R., & van Wezel, R. J. (2008). General validity of Levelt's propositions reveals common computational mechanisms for visual rivalry. *PLoS One*, *3*, e3473.

Knapen, T., Brascamp, J., Pearson, J., van Ee, R., & Blake, R. (2011). The role of frontal and parietal brain areas in bistable perception. *Journal of Neuroscience*, *31*, 10293–10301.

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, *17*, 307–321.

Kondo, H. M., Kitagawa, N., Kitamura, M. S., Koizumi, A., Nomura, M., & Kashino, M. (2012). Separability and commonality of auditory and visual bistable perception. *Cerebral Cortex*, *22*, 1915–1922.

Kornmeier, J., & Bach, M. (2009). Object perception: When our brain is impressed but we do not notice it. *Journal of Vision*, *9*(1):7, 1–10, doi:10.1167/9.1.7. [PubMed] [Article]

Kornmeier, J., & Bach, M. (2012) Ambiguous figures: What happens in the brain when perception changes but not the stimulus. *Frontiers in Human Neuroscience*, *6*, 51.

Ling, S., Hubert-Wallander, B., & Blake, R. (2010). Detecting contrast changes in invisible patterns during binocular rivalry. *Vision Research*, *50*, 2421–2429.

Lorber, M., Zuber, B. L., & Stark, L. (1965). Suppression of the pupillary light reflex in binocular rivalry and saccadic suppression. *Nature*, *208*, 558–560.

Lorenceau, J., & Shiffrar, M. (1992). The influence of

terminators on motion integration across space. *Vision Research*, *32*, 263–273.

Lunghi, C., Morrone, M. C., & Alais, D. (2014). Auditory and tactile signals combine to influence vision during binocular rivalry. *Journal of Neuroscience*, *34*, 784–792.

Macmillan, N. A., & Creelman, C.D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Marx, S., & Einhäuser, W. (2015). Reward modulates perception in binocular rivalry. *Journal of Vision*, *15*(1):11, 1–13, doi:10.1167/15.1.11. [PubMed] [Article]

Micheyl, C., & Oxenham, A. J. (2010). Objective and subjective psychophysical measures of auditory stream integration and segregation. *Journal of the Association for Research in Otolaryngology*, *11*, 709–724.

Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 919–931.

Naber, M., Frässle, S., & Einhäuser, W. (2011). Perceptual rivalry: Reflexes reveal the gradual nature of visual awareness. *PLoS One*, *6*, e20910.

Necker, L. A. (1832). Observations on some remarkable optical phaenomena seen in Switzerland; And on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London Edinburgh Philosophical Magazine and Journal of Science*, *1*, 329–337.

Ooi, T. L., & He, Z. J. (1999). Binocular rivalry and visual awareness: The role of attention. *Perception*, *28*, 551–574.

O'Shea, R. P., Kornmeier, J., & Roeber, U. (2013). Predicting visual consciousness electrophysiologically from intermittent binocular rivalry. *PLoS One*, *8*, e76134.

O'Shea, R. P., Parker, A., La Rooy, D., & Alais, D. (2009). Monocular rivalry exhibits three hallmarks of binocular rivalry: Evidence for common processes. *Vision Research*, *49*, 671–681.

Overgaard, M., & Fazekas, P. (2016) Can no-report paradigms extract true correlates of consciousness? *Trends in Cognitive Sciences*, *20*, 241–242.

Pannese, A., Herrmann, C. S., & Sussman, E. (2015). Analyzing the auditory scene: Neurophysiologic evidence of a dissociation between detection of regularity and detection of change. *Brain Topography*, *28*, 411–422.

Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current Biology*, *22*, 46–49.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Pressnitzer, D., & Hupé, J.-M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, *16*, 1351–1357.

Rajendran, V. G., Harper, N. S., Willmore, B. D., Hartmann, W. M., & Schnupp, J. W. H. (2013). Temporal predictability as a grouping cue in the perception of auditory streams. *Journal of the Acoustical Society of America*, *134*, EL98–EL104.

Rubin, E. (1921). *Visuell wahrgenommene Figuren*. Copenhagen, the Netherlands: Gyldendalske Boghandel.

Spering, M., & Carrasco, M. (2015). Acting without seeing: Eye movements reveal visual processing without awareness. *Trends in Neurosciences*, *38*, 247–258.

Spielmann, M. I., Schröger, E., Kotz, S. A., & Bendixen, A. (2014). Attention effects on auditory scene analysis: Insights from event-related brain potentials. *Psychological Research*, *78*, 361–378.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, *36*, 22–34.

Szalárdy, O., Winkler, I., Schröger, E., Widmann, A., & Bendixen, A. (2013). Foreground-background discrimination indicated by event-related brain potentials in a new auditory multistability paradigm. *Psychophysiology*, *50*, 1239–1250.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*, 753–759.

Tononi, G., Srinivasan, R., Russell, D. P., & Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proceedings of the National Academy of Sciences, USA*, *95*, 3198–3203.

Tsuchiya, N., Frässle, S., Wilke, M., & Lamme, V. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, *19*, 757–770.

Tsuchiya, N., Frässle, S., Wilke, M., & Lamme, V. (2016). No-report and report-based paradigms jointly unravel the NCC: Response to Overgaard and Fazekas. *Trends in Cognitive Sciences*, *20*, 242–243.

van Driel, J., Knapen, T., van Es, D. M., & Cohen, M. X. (2014). Interregional alpha-band synchrony supports temporal cross-modal integration. *NeuroImage*, *101*, 404–415.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences* (unpublished doctoral dissertation). Eindhoven, the Netherlands: Technical University Eindhoven.

van Rhijn, M., Roeber, U., & O'Shea, R. P. (2013). Can eye of origin serve as a deviant? Visual mismatch negativity from binocular rivalry. *Frontiers in Human Neuroscience*, *7*, 1–10.

Wales, R., & Fox, R. (1970). Increment detection thresholds during binocular rivalry suppression. *Perception & Psychophysics*, *8*, 90–94.

Wallach, H. (1935). Über visuell wahrgenommene Bewegungsrichtung. *Psychologische Forschung*, *20*, 325–380.

Warren, R. M., & Gregory, R. L. (1958). An auditory analogue of the visual reversible figure. *American Journal of Psychology*, *71*, 612–613.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung*, *4*, 301–350.

Wheatstone, C. (1838). Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, *128*, 371–394.

Wilbertz, G., van Slooten, J., & Sterzer, P. (2014). Reinforcement of perceptual inference: Reward and punishment alter conscious visual perception during binocular rivalry. *Frontiers in Psychology*, *5*, 1377.

Wilson, H., Blake, R., & Lee, S. (2001). Dynamics of travelling waves in visual perception. *Nature*, *412*, 907–910.

Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., … Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences, USA*, *100*, 11812–11815.