

Auditory and tactile signals combine to influence vision during binocular rivalry

Abbreviated title: audio-tactile signals combine to influence rivalry

Claudia Lunghi^{1,2,3}, Maria Concetta Morrone^{3,4} & David Alais⁵

¹ Department NEUROFARBA, University of Florence, Viale Pieraccini, 6, 50139 Florence, Italy.

² Institute of Neuroscience, CNR – Pisa, Via Moruzzi 1, 56124 Pisa, Italy.

³ Scientific Institute Stella Maris (IRCSS), viale del Tirreno 331, 56018 Calambrone, Pisa, Italy

⁴ Department of Translational Research on New Technologies in Medicine and Surgery, University of Pisa, via San Zeno 31, 56123 Pisa, Italy

⁵ School of Psychology, Brennan MacCallum Building, University of Sydney 2006, New South Wales, Australia

Corresponding Author: David Alais, davida@psych.usyd.edu.au

School of Psychology, Brennan MacCallum Building, University of Sydney 2006, New South Wales, Australia

28 pages, 5 figures

Abstract: 250w; Introduction: 466 w; Discussion: 1471 w.

Aknowledgments

This research was supported by the Italian Ministry of Universities and Research (PRIN2009), by EC project “STANIB” (FP7 ERC) and the Australian Research Council, Discovery Project #DP130102336. We thank Erik Van der Burg for helpful comments during the data collection. The authors declare no competing financial interests.

Abstract

Resolution of perceptual ambiguity is one function of cross-modal interactions. Here we investigate whether auditory and tactile stimuli can influence binocular rivalry generated by interocular temporal conflict in human subjects. Using dichoptic visual stimuli modulating at different temporal frequencies, we added modulating sounds or vibrations congruent with one or the other visual temporal frequency. Auditory and tactile stimulation both interacted with binocular rivalry by promoting dominance of the congruent visual stimulus. This effect depended on the cross-modal modulation strength and was absent when modulation depth declined to 33%. However, when auditory and tactile stimuli that were too weak on their own to bias binocular rivalry were combined, their influence over vision was very strong, suggesting the auditory and tactile temporal signals combined to influence vision. Similarly, interleaving discrete pulses of auditory and tactile stimuli also promoted dominance of the visual stimulus congruent with the supra-modal frequency. When auditory and tactile stimuli were presented at maximum strength, but in anti-phase, they had no influence over vision for low temporal frequencies – a null effect again suggesting audio-tactile combination. We also found that the cross-modal interaction was frequency-sensitive at low temporal frequencies, when information about temporal phase alignment can be perceptually tracked. These results show that auditory and tactile temporal processing is functionally linked, suggesting a common neural substrate for the two sensory modalities, and that at low temporal frequencies visual activity can be synchronized by a congruent cross-modal signal in a frequency-selective way, suggesting the existence of a supra-modal temporal binding mechanism.

Introduction

Viewing incompatible images in each eye produces binocular rivalry, an unstable state in which the images perceptually alternate irregularly over time (Blake and Logothetis,

2002; Alais, 2012). Each percept lasts a second or two and alternations continue as long as the rival stimuli are present. Because only one image reaches conscious perception, with the other suppressed from awareness, binocular rivalry is used to investigate the neural correlates of perceptual ambiguity and visual awareness (Logothetis, 1998; Koch, 2007). Rivalry is usually induced by spatial conflict (e.g., orientation, color or spatial frequency: (Yang et al., 1992; Kovacs et al., 1996) or competing visual objects (Tong et al., 1998; Alais and Melcher, 2007; Baker and Graf, 2009). Recently Alais and Parker (2012) showed that spatially matched random patterns rival vigorously when contrast-modulated at very different rates (e.g., factor of four). Vision encodes temporal modulations through two or three broad filters (Mandler and Makous, 1984; Anderson and Burr, 1985; Hess and Snowden, 1992; Johnston and Clifford, 1995; Cass and Alais, 2006) which can efficiently track phase at low temporal frequencies but show non-linear responses above 7-10 Hz (Rogers-Ramachandran and Ramachandran, 1998; Forte et al., 1999), producing a 'frequency doubling' illusion (Kelly, 1966). Poor visual temporal resolution relative to tactile and auditory domains could create difficulties for binding temporal signals from a single multisensory stimulus.

Resolving perceptual ambiguity is an important function of cross-modal interactions (Ernst and Bulthoff, 2004; Alais et al., 2010a; Klink et al., 2012). Recent studies show that non-visual modalities, including audition (Sekuler et al., 1997; Kang and Blake, 2005; Munhall et al., 2009; van Ee et al., 2009; Conrad et al., 2010; Chen et al., 2011), touch (Blake et al., 2004; Maruya et al., 2007; Holcombe and Seizova-Cajic, 2008; Alais et al., 2010b; Lunghi et al., 2010; Lunghi and Morrone, 2013) and olfaction (Zhou et

al., 2010) modulate bistable visual alternations. This study uses temporal frequency rivalry (15 vs. 3.75 Hz) to investigate whether auditory or tactile modulations (also 15 or 3.75 Hz) influence rivalry alternations. We predict the dominant visual temporal frequency will be maintained when the auditory or tactile stimulus matches it, and that visual perception will switch to the other frequency when cross-modal stimulation matches the suppressed temporal frequency. Also, as multisensory interaction helps temporally align signals related to common stimulus events, we expect phase-selective cross-modal influence. Because visual discrimination of phase is poor at high temporal frequencies, we predict phase selectivity only at low temporal frequencies.

Our results confirmed the cross-modal influence on rivalry was phase sensitive for low frequencies only. We also found cooperation between audio-tactile rhythms, as weak auditory and tactile signals strongly influenced rivalry when combined, but strong signals did not when combined in anti-phase. This suggests a common audio-tactile timing mechanism which can modulate vision during ambiguous temporal stimulation.

Materials and Methods

Subjects

Seven naïve observers (two females, average age 25.2 ± 0.8 years), plus author CL, participated in the main experiments. In addition, four naïve observers (three females, average age 28.4 ± 5 years), plus author CL, participated in the experiment in which discrete auditory and tactile pulse stimuli were interleaved in time. All had normal or corrected-to-normal vision, normal stereoacuity and no strong eye preference as defined by perceptual predominance during binocular rivalry. Two observers were reimbursed at \$AU20 for their time.

Ethics Statement

Participants gave written informed consent. The experimental procedure conformed to the declaration of Helsinki and was approved by the local ethics committee (Human Research Ethics Committee (HREC) Low Risk Executive Committee, University of Sydney, Protocol No. 14893) and by the by the ethics committee of the Scientific Institute Stella Maris, Italy.

Apparatus and Stimuli

Visual stimuli were created in MATLAB using PsychToolbox (Brainard, 1997), displayed on a linearized 20-inch LCD monitor (M9177 LCD Apple Cinema: 1280x960 pixels x 60

Hz) and viewed through a mirror stereoscope from a distance of 57 cm. For the experiment on alternating discrete audio-tactile stimuli, the rival stimuli were displayed on a ViewPixx3D monitor (Vpixx Technologies, QC, Canada) and viewed dichoptically through Nvidia 3D vision LCD shutter goggles (Nvidia, CA, USA). The rivaling stimuli were dynamic random-noise sequences (80 frames, 120 x 120 pixels, 3.2° visual angle) filtered into narrow temporal frequency pass-bands (one-third octave full bandwidth), one modulating at 3.75 Hz, the other at 15 Hz, and both were spatially band-pass filtered (centre frequency: 1.08 cyc/deg; full bandwidth: one octave). Spatially matched patterns with a four-fold temporal frequency difference produce vigorous rivalry alternations, as shown elsewhere (Alais and Parker, 2012). To balance relative dominance, the 15 Hz pattern was normalised to maximum contrast and the 3.75 Hz pattern varied (separately for each observer) to achieve approximately equal predominance. Visual stimuli were presented in central vision on a uniform grey background (50.9 cd/m^2) surrounded by a black fixation square to facilitate stable binocular fusion. A small, white, central fixation cross helped stabilise fixation and eye-movements. For the experiment with interleaved discrete auditory and tactile pulses, the visual stimuli were modulated at 3 and 12 Hz.

The auditory and tactile stimuli were amplitude-modulated sine-waves: 50 Hz carrier modulated at 3.75 or 15 Hz (matching the phase and rate of visual modulations) or at 5 and 20 Hz in a mis-matched control experiment. The audio output volume was regulated for each individual to achieve an approximate subjective match of sound and vibration intensities. The intensity was then manipulated by varying the modulation depth ($0 \leq d \leq 1$) of the frequency envelope (FM) modulating the 50 Hz frequency carrier

(FC) according to Equation 1.

$$\text{Stim}(t) = [1 + d \sin (2\pi FMt)]\sin (2\pi FCt)$$

Eq.1

Three modulation depths (d in Equation 1) were compared: 33%, 66% and 100%. For the experiment using interleaved auditory and tactile, the 50 Hz carrier was amplitude-modulated at 3 and 12 Hz according to Eq 1 and stimuli were rectified by annulling the even periods for one modality and the odds for the other one, so that each modality carried frequency information as discrete pulses with a fundamental harmonic at 1.5 and 6 Hz. These signals were presented either as auditory stimuli through headphones (Sennheiser HD 25-SP II; SPL = 72 dB) or as tactile vibrations through a Clark Synthesis Tactile Sound Transducer (TST429 platinum). To make sure the tactile signal was inaudible, observers wore insulating headphones. All signals were well above threshold and were readily perceived even at 33% modulation depth. The tactile driver was located between the video screen and chinrest and observers grasped a rubber ball mounted on a post attached to the driver to receive tactile signals. This was housed in a wooden box so observers could not see their hand during the experiment.

Task and Procedure

Observers tracked their rivalry alternations between the low- and high-frequency modulating patterns using the computer keyboard. One pattern was tinted red and the other green to aid in tracking alternations. Each observer participated in four 240-sec

experimental sessions on different days of 10 experimental conditions, for a total time of 160 minutes. Sessions began with the square fusion frames so observers could adjust the stereoscope for stable binocular fusion and then initiated the rivalry stimuli. Eye of presentation and stimulus colour were randomized and counterbalanced for the two visual temporal frequencies across sessions to control for response bias and eye-preference. As observed in Alais and Parker (2012), colour did not rival independently of temporal frequency.

Three audiovisual and three tactuo-visual conditions were tested. These involved presentation of an auditory or a tactile stimulus of 2.5 s duration at regular intervals of 8 s during a 4-minute rivalry trial. The auditory or tactile envelope frequency always matched one of the visual temporal frequencies (3.75 or 15 Hz) and were delivered in random order, and therefore were unpredictably congruent or incongruent with the temporal frequency perceived by the observer at the moment of presentation (Figure 1A). In separate blocks, the auditory and tactile stimuli were tested at three different modulation depths, 33%, 66% and 100%. Following the same procedure, frequency mismatch conditions were also run in which the auditory and tactile stimuli were higher in frequency than the visual temporal frequencies by 33% (5 & 20 Hz).

Again following the same procedure we tested three tri-modal conditions. In one, the auditory and tactile stimuli both had a 33% modulation depth and the same envelope temporal frequency and phase (matching one of the visual stimuli). In the second, the auditory and tactile stimuli had 100% modulation depth, and both had a envelope frequency matching one of the visual stimuli, but the auditory and tactile

stimuli envelope were both 180° out of phase relative to each other. In the other tri-modal experiment, half-period of tactile and auditory stimuli were presented sequentially in time so that the combined (supra-modal) temporal frequency would match the visual temporal frequency. The order of audiovisual, tactuo-visual and tri-modal conditions was randomized and compared with visual-only control periods.

RESULTS

Presenting auditory or tactile stimuli intermittently during binocular rivalry significantly altered rivalry alternation dynamics. As Figure 1 shows for both auditory and tactile stimulation, the cross-modal stimulus increased the probability of maintaining the visual percept when it was congruent with the current visual percept (Fig. 1C), or increased the probability of a perceptual switch when it was incongruent with the current visual percept (Fig. 1B). For cross-modal signals of 100% modulation depth, paired t-tests showed increased probability of maintaining a percept (Auditory: congruent vs. visual-only, +15.4%, $t(7)=2.8$, $p=0.026$. Tactile: congruent vs. visual-only, +18.9%, $t(7)=4.85$, $p=0.0019$). Conversely, the probability of switching visual percepts during a period of cross-modal stimulation was significantly higher when the cross-modal stimulus was incongruent with the current visual percept (Auditory: incongruent vs. visual-only, +11.3%, $t(7)=2.73$, $p=0.029$. Tactile: incongruent vs. visual-only: +15%, $t(7)=3.29$, $p=0.013$). There was a low probability of switching more than once during a touch period (on the order of 20%). This is not plotted but is equivalent to $1-(\text{probability of maintaining} + \text{probability of switching once})$. This confirms that modulating auditory and tactile stimuli can promote a congruent but unseen visual stimulus from suppression into visual awareness when they match in temporal frequency.

We also examined these effects as a function of cross-modal signal strength by comparing amplitude modulation depth over three levels (100, 66 and 33%). Figure 1 shows the cross-modal influence on binocular rivalry dynamics declines with decreasing signal strength. The strong cross-modal effect obtained with 100% modulation was

reduced for 66% modulation depth, and at 33% modulation neither auditory nor tactile conditions differed significantly from visual-only conditions in probability of switching (Fig. 1B) or maintaining visual percept (Fig. 1C).

We also computed the time-course of the cross-modal effect on rivalry separately for low (3.75 Hz) and high (15 Hz) frequency stimulation. To achieve this, the alternation time series recorded by each observer were divided into segments defined by the cross-modal stimulus onset. An observer's segments were summed and averaged into 250 ms bins (see Figures 2, 3 and 4). Figure 2 shows, for cross-modal stimuli of 100% modulation depth (black symbols), that the probability of seeing the congruent visual stimulus increases over time from the moment of cross-modal onset. This increasing function is similar for auditory and tactile stimulation, but it is slower for low temporal frequency stimulation, increasing progressively and becoming significant at ~ 1.25 s after cross-modal onset for 15 Hz auditory and tactile stimulation (Figure 2C-D) and becoming significant at ~ 2.25 s after cross-modal onset for 3.75 Hz stimulation (Figure A-B). In both conditions the effect continues to build for the duration of the cross-modal stimulus and decays slowly after the offset of cross-modal stimulation. Asterisks indicate the probability of the congruent visual percept exceeding chance on a one-tailed t-test at $p < 0.05$ (double asterisks indicate $p < 0.01$). For cross-modal stimulation of 33% modulation depth the bias in favour of the congruent visual stimulus did not reach statistical significance during cross-modal stimulation (Figure 2, grey symbols). Interestingly, we found a delayed effect (after sound or vibration offset) for 33% modulated cross-modal stimulation, possibly indicating that weak auditory and tactile

signals are less effectively transmitted and the effect takes time to accrue.

We tested whether auditory and tactile stimuli would combine to influence rivalry dynamics more strongly together than separately. In a tri-modal experiment we used the same paradigm but delivered tactile and auditory stimuli together, matched in envelope frequency, phase and duration. We used cross-modal stimuli at 33% modulation because these signals were too weak to bias binocular rivalry on their own (see Fig. 2, grey symbols). Together, however, weak auditory and tactile stimuli exerted a very strong influence on binocular rivalry dynamics similar in magnitude to the auditory and tactile stimulation alone at maximum modulation. The time-course of the tri-modal effect was very similar to that obtained for either auditory or tactile stimuli with 100% envelope modulation depth, building throughout the cross-modal stimulation period, peaking around cross-modal offset, and then declining symmetrically (compare Fig. 3A-C with Fig. 2, black symbols). This result is a novel one indicating that tactile and auditory signals that are matched in envelope frequency and phase can functionally combine to bias visual perception in combination.

To model the summation of the weak auditory and tactile signals we transformed the time-course data in Figure 2 from ‘probability of congruent’ (p , in Equation 2) into d' prime values according to the following formula:

$$d' = 2 * \operatorname{erf}^{-1}(2 * (p - .5)) \quad \text{Eq. 2}$$

With the data now in standardized, we added the auditory and tactile d' values and

converted their sum back to probabilities using the following formula:

$$p = \text{erf}(d')/4 + .5 \quad \text{Eq. 3}$$

This procedure models the combined audio-tactile effect as a linear summation of the individual auditory and tactile effects and is plotted in Figure 3A-C. The linear summation model (continuous line with shaded error bars) is very similar to the observed tri-modal time-course in both amplitude and width and indicates that auditory and tactile temporal signals can combine in a near-linear combination.

To test directly the audio-tactile combination hypothesis, we conducted a tri-modal experiment with auditory and tactile components that were matched in envelope frequency (both 3.75 or 15 Hz) but had opposite phase, meaning they would generate a salient double frequency supra-modal rhythm (a diagram of the auditory and tactile stimuli used in this experiment is reported in Figure 4A). The auditory and tactile signals were presented at maximum amplitude (100%). The results confirm our predictions (Figure 3B-D): the probability of seeing the visual stimulus congruent in temporal frequency with the cross-modal stimulus does not differ from chance at any point for 3.75 Hz stimulation (Figure 3B), while for 15 Hz stimulation an initial, fast bias in favour of the high-temporal frequency visual stimulus is present but no sustained bias (Figure 3C).

We modeled the time course of the tri-modal anti-phase data using the same linear summation analysis described above with one change: we gave the component d' values opposite signs to test whether they would sum to zero. The linear sum model is

plotted in Figure 3B-D (continuous line with shaded error bars) and is not significantly different from chance level at any time point. The close agreement between the observed data and the linear summation prediction further supports the claim that auditory and tactile envelope frequencies combined to form a supra-modal frequency which synchronized visual rivalry perception.

To verify this hypothesis, we performed an additional tri-modal experiment in which periods of pure auditory stimulation were followed by periods of tactile (see diagram in Figure 4B). In this stimuli, the auditory and tactile modulation peaks were temporally interleaved, producing a supra-modal frequency matching the visual temporal frequencies. Using rectified stimuli further allowed us to test whether discrete events carrying temporal frequency information, rather than continuous oscillatory stimulation, can influence temporal frequency binocular rivalry. The results clearly show that the audio-tactile stimuli efficiently combined to bias binocular rivalry (Figure 4C), with the time-course of the effect of promoting dominance of the visual temporal frequency congruent with the combined audio-tactile stimulation being similar to that observed for congruent stimuli in a single modality. This result indicates that auditory and tactile events can combine to bias binocular rivalry, suggesting the existence of a supra-modal temporal mechanism that combines information about temporal patterns across different sensory modalities.

In a final experiment we investigated the temporal frequency selectivity of the cross-modal temporal congruence effects on binocular rivalry dynamics. We returned to the original paradigm of using continuous oscillatory stimulation described above but

used auditory and tactile signals that were 33% higher in temporal frequency (5 and 20 Hz) than the visual stimuli (3.75 and 15 Hz). The cross-modal signals had 100% modulation depth. The results are shown in Figure 5. For low frequency stimulation (5 Hz auditory, Figure 5A and tactile, Figure 5B) the rivalry alternation dynamics did not differ significantly from chance level at any time point in either condition. The lack of a cross-modal effect in this condition indicates that the audiovisual and tactuo-visual interaction is temporal-frequency selective for low frequencies. For the high frequency cross-modal stimulation (20 Hz auditory, Figure 5C and tactile, Figure 5D), a bias towards the high frequency visual stimulus (15 Hz) was observed despite the mismatch in temporal frequency, especially for touch, where the effect reached statistical significance (Figure 5D). This indicates that, for high temporal frequencies, the cross-modal audiovisual and tactuo-visual interaction is not temporal frequency selective. Interestingly, the timecourse of the effect of auditory and tactile stimulation on binocular rivalry for high frequency cross-modal stimulation in this condition builds-up and decays faster compared to that observed for cross-modal stimuli matched in temporal frequency, an effect that is similar to that observed for anti-phase cross-modal stimulation at high temporal frequency, indicating that the lack of selectivity for these frequencies may be related to the poor visual temporal discrimination at high temporal frequencies.

Discussion

We have shown that binocular rivalry from interocular temporal conflict is strongly influenced by tactile and auditory temporal signals. Cross-modal signals temporally congruent with one of the rivaling visual temporal frequencies extended that stimulus' percept duration, if already dominant, or promoted it to dominance if suppressed (Figure 1). Decreasing cross-modal signal intensity reduced the effect, yet weak auditory and tactile stimuli together combined to produce a strong influence (Figure 3A-C), suggesting audio-tactile summation in a common temporal mechanism. In support of this, when discrete auditory and tactile stimuli were alternated in time to create a supramodal frequency matching one of the visual stimuli, a strong binocular rivalry bias was observed in favour of the congruent visual stimulus. In further support, auditory and tactile modulations with maximum amplitude did not influence rivalry when combined in opposite phase (see low temporal frequencies in Figure 3B), strengthening the claim that the source of the effect is supra-modal, after combination of the component signals. We also found that the cross-modal influence is more frequency-selective for low temporal frequencies (3.75 Hz) than for high (15 Hz), as increasing the auditory and tactile modulation rates by 33% eliminated the cross-modal influence on the 3.75 Hz visual stimulus, but still promoted dominance of the 15 Hz visual stimulus. These results agree with evidence showing that vision cannot distinguish temporal phase and frequency differences for frequencies higher than 7-10 Hz due to non-linear mechanisms in visual temporal frequency processing ((Rogers-Ramachandran and Ramachandran, 1998; Forte et al., 1999).

Overall, our results imply close links between three modalities in the temporal

domain in two ways. First, temporal signals in a second modality (tactile or auditory) help resolve temporal ambiguity in vision, and second, rhythmic auditory and tactile signals combine to bias binocular rivalry, with a clear phase and frequency selectivity for low temporal frequency stimulation. The first finding agrees with recent studies showing cross-modal signals help resolve binocular rivalry (van Ee et al., 2009; Conrad et al., 2010; Lunghi et al., 2010; Chen et al., 2011; Lunghi and Alais, 2013; Lunghi and Morrone, 2013) and other visual ambiguities (Blake et al., 2004; Conrad et al., 2012). This is often explained by cross-modal feedback from higher cortical areas modulating early visual processing, similar to feedback accounts of attentional, contextual and semantic modulation of rivalry. However, other cross-modal findings suggest early cortical interactions, such as the tight orientation and spatial frequency tuned tactile influence on spatial rivalry (Lunghi et al., 2010; Lunghi and Alais, 2013). The present results, in showing temporal frequency and phase selectivity and a graded effect of signal intensity suggest an early cross-modal interaction rather than higher-level feedback.

The cross-modal stimulus not only prolonged dominance when the cross-modal stimulus matched the dominant visual percept, but also made a perceptual switch more likely when it matched the unseen stimulus. The effect on the suppressed stimulus is unusual: most studies show that modulatory influences on rivalry (e.g., attention, semantics, context) modulate only the perceptually dominant stimulus (Sobel and Blake, 2002; Meng and Tong, 2004; Chong et al., 2005; but see Hancock and Andrews, 2007; Paffen and Alais, 2012). However, our result could be explained by mutual phase alignment because, even when suppressed from awareness, there is still neural activity

(albeit attenuated) related to a matching temporal frequency in two sensory areas of the brain which, if phase aligned, would boost the suppressed stimulus' salience and make a switch from suppression to dominance more likely. On this view, the cross-modal influence on rivalry is a compulsory, early multisensory interaction rather than feedback from higher levels. That cross-modal stimulation rescues the congruent visual stimulus from rivalry suppression argues against an attentional account because attentional allocation to invisible objects is not thought to be possible. However recent work (Chou and Yeh, 2012) has challenged this by showing that a form object-based attention (same-object advantage) can occur during continuous flash suppression. Moreover, these results cannot be a response bias because the weakest cross-modal modulations failed to influence visual dynamics even though they were still perceptually salient and far above threshold.

Entrainment of coherent neural oscillations has been suggested as a binding and multisensory integration mechanism (Maier et al., 2008; Senkowski et al., 2008), with distributed neural responses bound together when their activity oscillates synchronously. Intrinsic oscillatory activity in unisensory areas can become mutually phase-aligned when driven by matched multisensory input (as in speech, or the cross-modal rivalry conditions used here). A magnetoencephalography study (Luo et al., 2010) found theta-delta (2–7 Hz) activity in primary auditory and visual cortices of human subjects was reset to a mutually aligned phase when watching naturalistic movie samples. This was only observed when auditory and visual streams matched: when mismatched, phase resetting was not observed. Importantly, phase coherence, not

overall power, was modulated by matching audiovisual input. Moreover, their results suggest an active resetting of theta-delta oscillations across sensory areas towards a mutual and preferred phase angle (0° or 180°) that enhances cross-sensory responses because each area receives its input in a high excitability state. This process could strengthen responses to the visual frequency matching the cross-modal stimulus in our experiments, boosting dominance duration or triggering a switch to dominance, and it could do so in real-time on a fine time-scale. In this vein, a cross-modal temporal binding mechanism would rely on the simultaneity and frequency of the cross-modal events rather than on spatial features, that would explain why spatial overlap between the cross-modal stimuli was not a necessary condition for the interaction observed in this study. It would also explain the lack of cross-modal influence with mis-matched stimuli at low temporal frequencies.

Similar phase effects in oscillatory activity are seen in other multisensory combinations. Responses from primary auditory cortex (A1) in awake macaques (Lakatos et al., 2007) are enhanced by somatosensory input through phase resetting of A1 oscillations so that the accompanying somatosensory input arrived at the optimal phase. Another study showed similar effects of visual stimuli modulating auditory oscillations (Kayser et al., 2008). Finally, phase resetting of oscillations in somatosensory cortex due to congruent visual input has been recently observed in rats (Sieben et al., 2013). Overall, given the active alignment of unisensory oscillations from matched multisensory inputs, and connections between early sensory cortices (Rockland and Ojima, 2003; Clavagnier et al., 2004; Beer et al., 2011), we suggest that matched cross-

modal stimuli in our visual rivalry experiments boosted the salience of the matching visual response, biasing the competition underlying rivalry dynamics. Moreover, frequency selective entrainment of neural oscillation has been found in visual (Kanai et al., 2008), somatosensory (Feurra et al., 2011) and parietal (Thut et al., 2011) cortex for alpha and beta frequencies using transcranial stimulation (TMS and tACS). We speculate that in our study cross-modal rhythmic stimulation could entrain visual cortical oscillators in a frequency-selective manner.

The second striking result we report is summation of auditory and tactile temporal signals. Weak audio-tactile modulations with matching frequency and phase combined to strongly influence rivalry while, for low temporal frequencies strong signals with matching frequency but opposite phase did not. Moreover, discrete auditory and tactile temporal events interleaved to produce a supra-modal rhythm matching the visual temporal frequency efficiently interacted with binocular rivalry, promoting dominance of the visual stimulus congruent with the combined audio-tactile temporal frequency. Our results show for the first time that auditory and tactile temporal frequency signals can functionally combine to effectively disambiguate visual perception during binocular rivalry. This functional combination of tactile and auditory temporal signals suggests the existence of a cross-modal temporal binding mechanism of sequential events (rhythm) that can indeed occur between auditory, tactile and visual temporal mechanisms and could be mediated by a supra-modal sense of time that combines the three sensory modalities.

An interesting feature of the probability traces (Figure 2) is the cross-modal effect

promoting dominance of the congruent stimulus develops slowly, taking ~ 1 s to reach significance and rising throughout the 2.5 s sound/touch period. This probably reflects the time-course of adaptation among competing visual neurons, as adaptation is critical in determining perceptual switches. As recently shown (Alais et al., 2010c), contrast sensitivity to rival stimuli slowly changes during a single rivalry period. Initially, sensitivity to the dominant stimulus is high and the suppressed stimulus low. This difference reduces due to adaptation of the dominant response, correspondingly releasing the suppressed response from inhibition, causing convergence towards equal response levels. This makes the visual interpretation increasingly ambiguous and the scope for cross-modal signals to help resolve the conflict increases. Binocular rivalry is therefore well suited to revealing subtle auditory and tactile influences on vision. Indeed, the low frequency phase effects we report provide the first psychophysical evidence of entrained cross-modal synchronization of neural activity, with ambiguous vision biased in a phase-sensitive way by cross-modal stimuli and showing that perceptual oscillations in binocular rivalry are sensitive to synchronization signals.

References

- Alais D (2012) Binocular Rivalry: competition and inhibition in visual perception. *WIREs Cogn Sci* 3:87-103.
- Alais D, Melcher D (2007) Strength and coherence of binocular rivalry depends on shared stimulus complexity. *Vision Res* 47:269-279.
- Alais D, Parker A (2012) Binocular rivalry produced by temporal frequency differences. *Front Hum Neurosci* 6:227.
- Alais D, Newell FN, Mamassian P (2010a) Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving* 23:3-38.
- Alais D, van Boxtel JJ, Parker A, van Ee R (2010b) Attending to auditory signals slows visual alternations in binocular rivalry. *Vision Res* 50:929-935.
- Alais D, Cass J, O'Shea RP, Blake R (2010c) Visual sensitivity underlying changes in visual consciousness. *Curr Biol* 20:1362-1367.
- Anderson SJ, Burr DC (1985) Spatial and temporal selectivity of the human motion detection system. *Vision Res* 25:1147-1154.
- Baker DH, Graf EW (2009) Natural images dominate in binocular rivalry. *Proc Natl Acad Sci USA* 106:5436-5441.
- Beer AL, Plank T, Greenlee MW (2011) Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Exp Brain Res* 213:299-308.
- Blake R, Logothetis NK (2002) Visual competition. *Nat Rev Neurosci* 3:13-21.
- Blake R, Sobel KV, James TW (2004) Neural synergy between kinetic vision and touch. *Psychol Sci* 15:397-402.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433-436.
- Cass J, Alais D (2006) Evidence for two interacting temporal channels in human visual processing. *Vision Res* 46:2859-2868.
- Chen YC, Yeh SL, Spence C (2011) Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry. *Front Psychol* 2:212.
- Chou WL, Yeh SL (2012) Object-based attention occurs regardless of object awareness. *Psychon Bull Rev* 19:225-231.
- Clavagnier S, Falchier A, Kennedy H (2004) Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cogn Affect Behav Neurosci* 4:117-126.
- Conrad V, Vitello MP, Noppeney U (2012) Interactions between apparent motion rivalry in vision and touch. *Psychol Sci* 23:940-948.
- Conrad V, Bartels A, Kleiner M, Noppeney U (2010) Audiovisual interactions in binocular rivalry. *J Vis* 10:27.
- Ernst MO, Bulthoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8:162-169.
- Feurra M, Paulus W, Walsh V, Kanai R (2011) Frequency specific modulation of human somatosensory cortex. *Front Psychol* 2:13.
- Forte J, Hogben JH, Ross J (1999) Spatial limitations of temporal segmentation. *Vision Res* 39:4052-4061.

- Hess RF, Snowden RJ (1992) Temporal properties of human visual filters: number, shapes and spatial covariation. *Vision Res* 32:47-59.
- Holcombe AO, Seizova-Cajic T (2008) Illusory motion reversals from unambiguous motion with visual, proprioceptive, and tactile stimuli. *Vision Res* 48:1743-1757.
- Johnston A, Clifford CW (1995) Perceived motion of contrast-modulated gratings: predictions of the multi-channel gradient model and the role of full-wave rectification. *Vision Res* 35:1771-1783.
- Kanai R, Chaieb L, Antal A, Walsh V, Paulus W (2008) Frequency-dependent electrical stimulation of the visual cortex. *Curr Biol* 18:1839-1843.
- Kang M-S, Blake R (2005) Perceptual synergy between seeing and hearing revealed during binocular rivalry. *Psychologia* 32.
- Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. *Cereb Cortex* 18:1560-1574.
- Kelly DH (1966) Frequency Doubling in Visual Responses. *Journal of the Optical Society of America* 56.
- Klink PC, van Wezel RJ, van Ee R (2012) United we sense, divided we fail: context-driven perception of ambiguous visual stimuli. *Philos Trans R Soc Lond B Biol Sci* 367:932-941.
- Koch C (2007) *The quest for Consciousness: a neurobiological Approach*. Greenwood Village, CO: Roberts & Company.
- Kovacs I, Papathomas TV, Yang M, Feher A (1996) When the brain changes its mind: interocular grouping during binocular rivalry. *Proc Natl Acad Sci U S A* 93:15508-15511.
- Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279-292.
- Logothetis NK (1998) Single units and conscious vision. *Philos Trans R Soc Lond B Biol Sci* 353:1801-1818.
- Lunghi C, Morrone MC (2013) Early Interaction between Vision and Touch during Binocular Rivalry. *Multisensory Research* 26:291-306.
- Lunghi C, Alais D (2013) Touch Interacts with Vision during Binocular Rivalry with a Tight Orientation Tuning. *PLoS One* 8:e58754.
- Lunghi C, Binda P, Morrone MC (2010) Touch disambiguates rivalrous perception at early stages of visual analysis. *Curr Biol* 20:R143-144.
- Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol* 8:e1000445.
- Maier JX, Chandrasekaran C, Ghazanfar AA (2008) Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Curr Biol* 18:963-968.
- Mandler MB, Makous W (1984) A three channel model of temporal frequency perception. *Vision Res* 24:1881-1887.
- Maruya K, Yang E, Blake R (2007) Voluntary action influences visual competition. *Psychol Sci* 18:1090-1098.

- Munhall KG, ten Hove MW, Brammer M, Pare M (2009) Audiovisual integration of speech in a bistable illusion. *Curr Biol* 19:735-739.
- Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. *Int J Psychophysiol* 50:19-26.
- Rogers-Ramachandran DC, Ramachandran VS (1998) Psychophysical evidence for boundary and surface systems in human vision. *Vision Res* 38:71-77.
- Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. *Nature* 385:308.
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31:401-409.
- Sieben K, Roder B, Hanganu-Opatz IL (2013) Oscillatory entrainment of primary somatosensory cortex encodes visual control of tactile processing. *J Neurosci* 33:5736-5749.
- Thut G, Veniero D, Romei V, Miniussi C, Schyns P, Gross J (2011) Rhythmic TMS causes local entrainment of natural oscillatory signatures. *Curr Biol* 21:1176-1185.
- Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21:753-759.
- van Ee Rv, van Boxtel JJA, Parker AL, Alais D (2009) Multisensory Congruency as a Mechanism for Attentional Control over Perceptual Selection. *J Neurosci* 29:11641-11649.
- Yang Y, Rose D, Blake R (1992) On the variety of percepts associated with dichoptic viewing of dissimilar monocular stimuli. *Perception* 21:47-62.
- Zhou W, Jiang Y, He S, Chen D (2010) Olfaction modulates visual perception in binocular rivalry. *Curr Biol* 20:1356-1358.

Figures

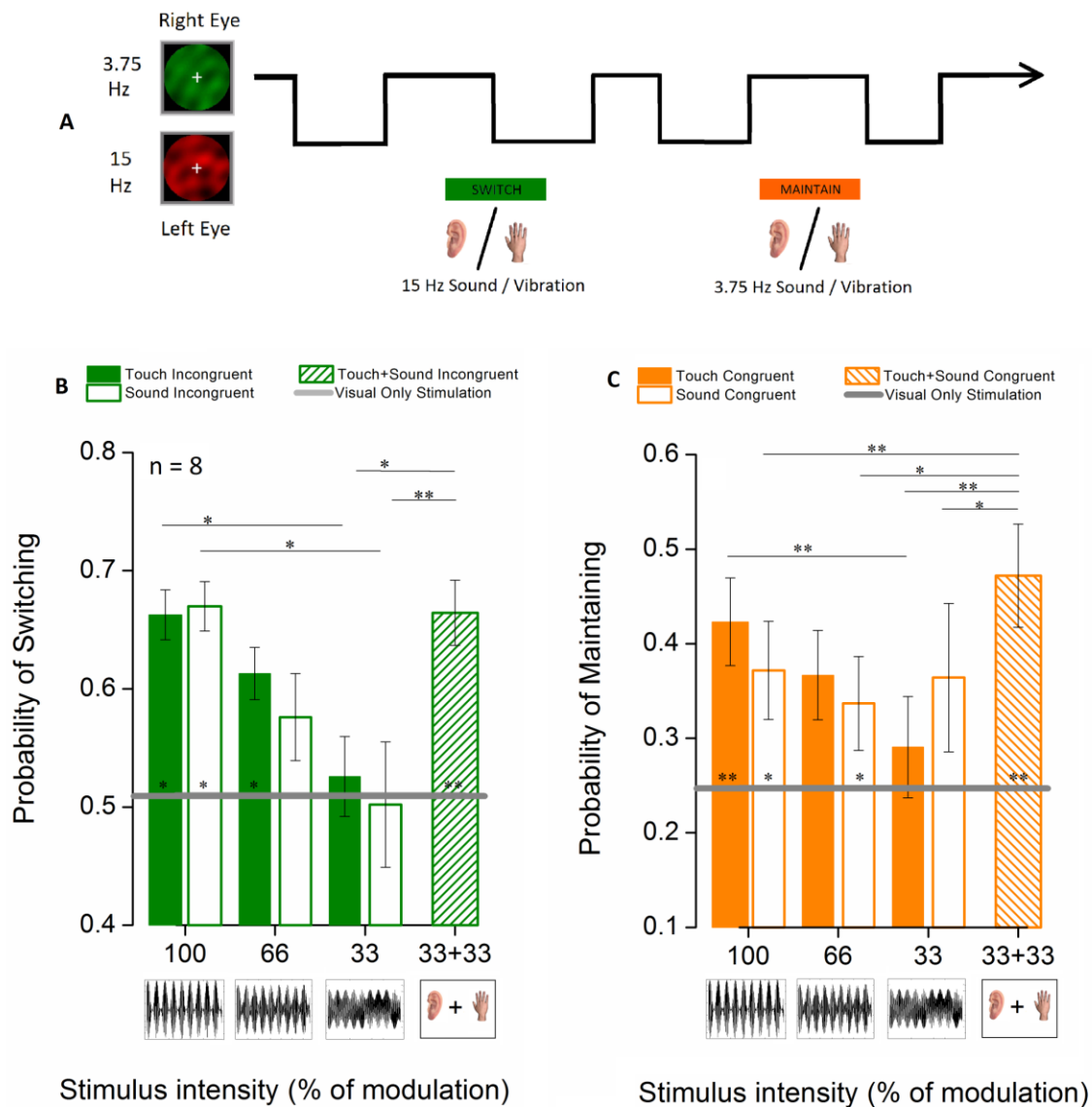


Figure 1. Stimulus intensity dependent influence of tactile and auditory stimulation on binocular rivalry dynamics. **A. Experimental paradigm.** Dynamic visual noise was filtered into narrow temporal frequency pass-bands (3.75 and 15 Hz) and presented dichoptically, leading to rivalrous perceptual alternations. During an extended period of binocular rivalry viewing (240 s), a sound or tactile vibration at 3.75 or 15 Hz (amplitude modulations of a 50 Hz carrier frequency) was presented each 10.6 s for period of 2.6 s. The 8 s between each cross-modal stimulus contained only visual stimulation. Observers

reported the color (red or green) of the dominant visual frequency continuously using keys on the computer keyboard. **B-C. Results.** The probability of switching visual perception once (**B**) or maintaining the same visual perception (**C**) during a 2.6 s period of cross-modal stimulation for tactile (filled bars) and auditory (empty bars) stimulation is plotted for different cross-modal stimulation intensities (100%, 66% and 33% amplitude modulation depth of a 50 Hz carrier) and for a tri-modal condition in which auditory and tactile stimulation were combined at 33% amplitude modulation (diagonally striped bars). The grey lines in panels b and c indicate the mean probabilities of switching or maintaining percept, respectively, during periods of visual-only stimulation. All probabilities were statistically compared using a paired-samples t-test ($N=8$, $\alpha =0.05$, $df=7$), with p values represented by asterisks: * indicates $p\leq 0.05$; ** indicates $p\leq 0.01$. Error bars represent ± 1 standard error of the mean (SEM).

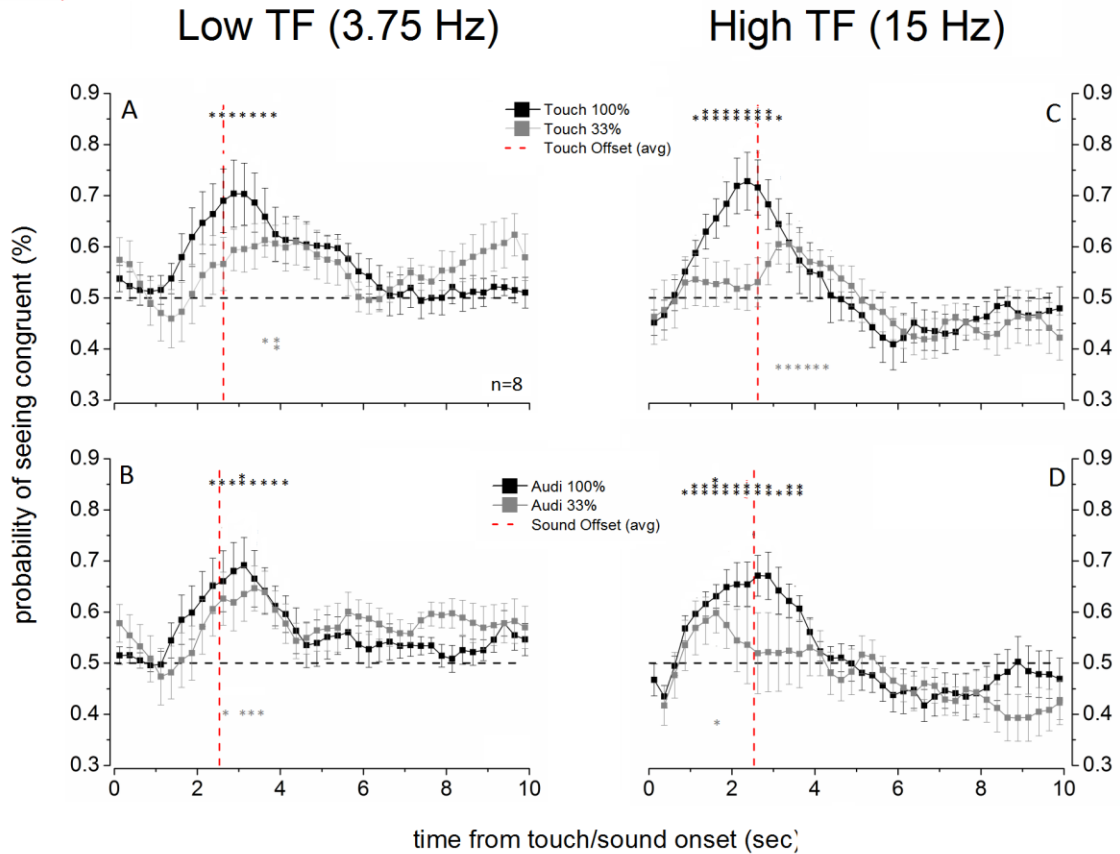


Figure 2. Timecourse of the influence of crossmodal stimulation on bincular rivalry. The plots show how the probability of seeing the stimulus congruent with the crossmodal stimulus varies over time. The effect of cross-modal stimuli at low temporal frequency (panels a and b) and high temporal frequency (panels c and d) are shown separately. To obtain a probability trace, a subject's tracking of perceptual alternations for the 10 seconds following each cross-modal stimulus onset were overlaid and averaged. The data points on each probability trace are the group mean of each subject's percept tracking, averaged every 250 ms from the onset of the cross-modal stimulus. Error bars represent ± 1 SEM, and the vertical dashed red line represents the offset of cross-modal stimulation. For each condition probability traces are reported for cross-modal stimulation at maximum (100%: black symbols) and minimum (33%: grey symbols) modulation depth. The horizontal dashed black line represents chance level (0.5) and therefore no effect of cross-modal stimulation. Every point of the probability trace was

statistically compared with chance level (one-sample t-test, $N=8$, $df=7$), with p values represented by asterisks: * indicates $p \leq 0.05$; ** indicates $p \leq 0.01$.

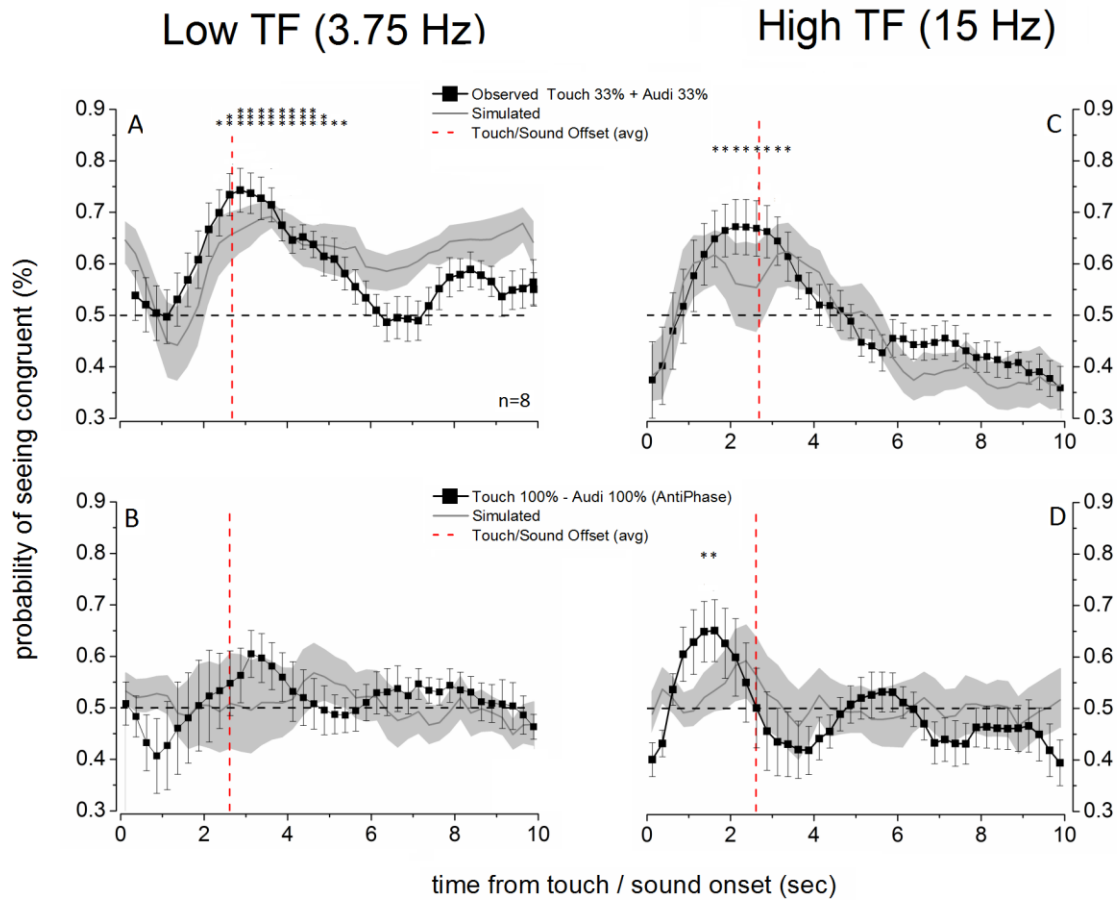


Figure 3. Auditory and tactile stimuli summate to bias binocular rivalry. The timecourse of the probability of seeing the visual stimulus congruent with the 3.75 Hz (left panels) or 15 Hz (right panels) cross-modal stimulus is plotted for the two tri-modal experiments. All error bars show ± 1 SEM. **A,C.** The weak auditory and tactile stimuli (33% modulation depth) presented simultaneously strongly biased binocular rivalry towards the visual modulation congruent with the audio-tactile temporal frequency. Each point on the probability trace was compared against chance level (dashed black line) using a one-

sample t-test ($N=8$, $df=7$), with p values represented by asterisks: * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$. The grey traces simulate a linear summation of the individual auditory and tactile probability traces (figure 2, grey symbols) obtained by transforming the individual traces into d' (Eq. 1), then summing them and reverting the obtained compound d' into probability again (Eq. 2). **B,D.** Combining strong auditory and tactile stimuli (100% modulation depth) did not bias binocular rivalry towards the congruent temporal frequency when the cross-modal stimuli were presented with opposite phases. The grey traces simulate a cancellation of the tactile and auditory effects (Figure 2, black symbols) by transforming the individual auditory and tactile traces for into d' (Eq. 1), subtracting them, and then reverting this difference into a probability again (Eq. 2). The vertical dashed red line represents the offset of crossmodal stimulation.

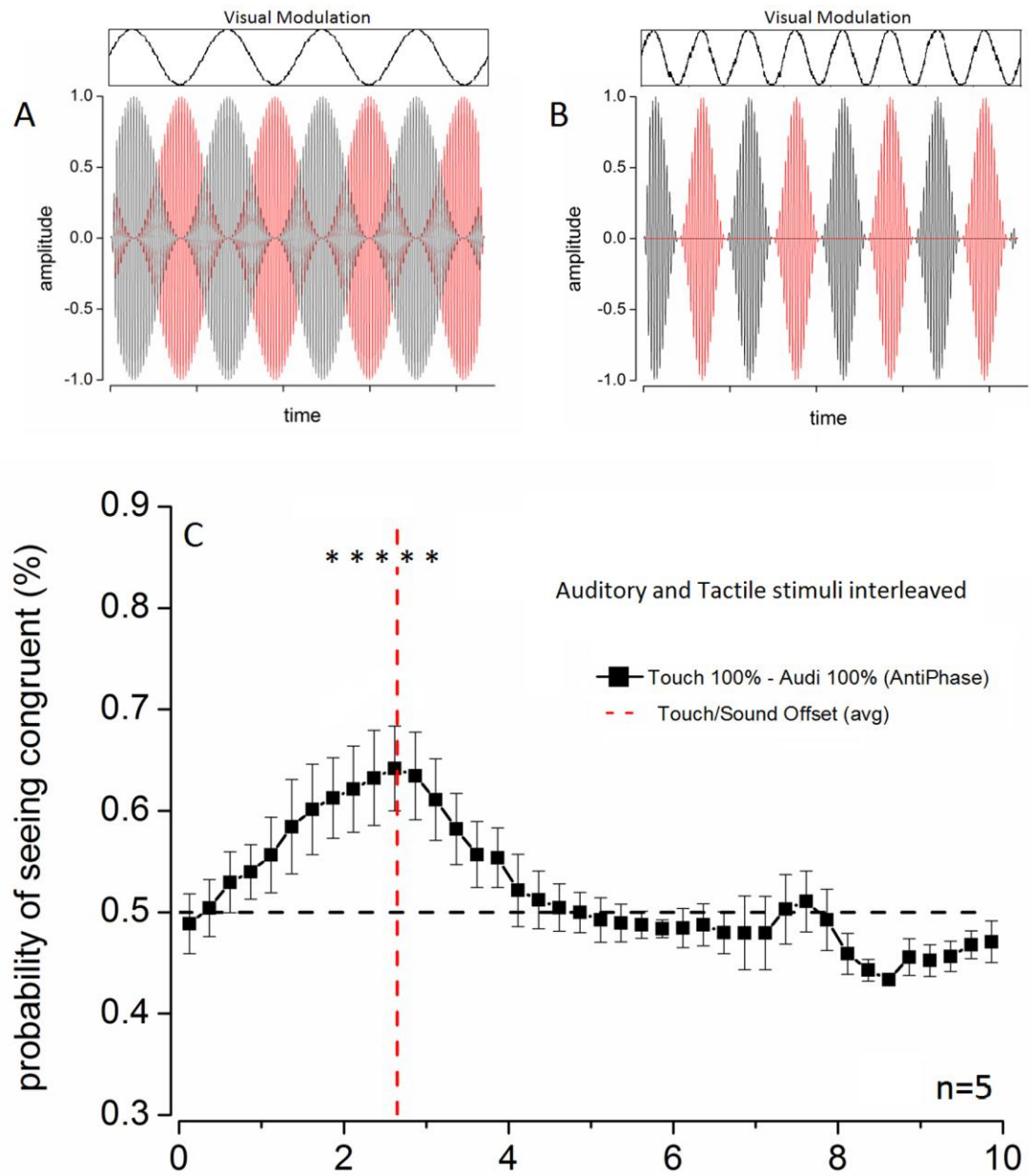


Figure 4. Audio-tactile temporal-frequency combination.

Audio-tactile stimuli used in the tri-modal anti-phase experiments. **(A)** Continuous oscillating amplitude-modulated sounds and vibrations matched in temporal frequency with the visual stimuli in opposite phase. **(B)** Rectified amplitude modulated sounds and vibrations alternated over time. The stimulus in each

modality was obtained by rectifying the envelope shown in A and canceling every second modulation to produce a series of discrete pulses followed by blank periods. When the other modality is rectified in the same way, both can be interleaved over time to form an audio-tactile frequency matching the visual frequency. **(C)** Results of the experiment in which the audio-tactile stimuli described in **(B)** were used. The time-course of the probability of seeing the visual stimulus matching the combined audio-tactile temporal frequency averaged every 250 ms is reported as a function of time from the onset of the cross-modal stimulation. Every point of the probability trace so obtained was statistically compared with chance level (dashed black line, one sample t-test, $N=5$, $df=4$), significance level is represented by asterisks: * indicates $p \leq 0.05$. The red dashed line represents the offset of the cross-modal stimulation.

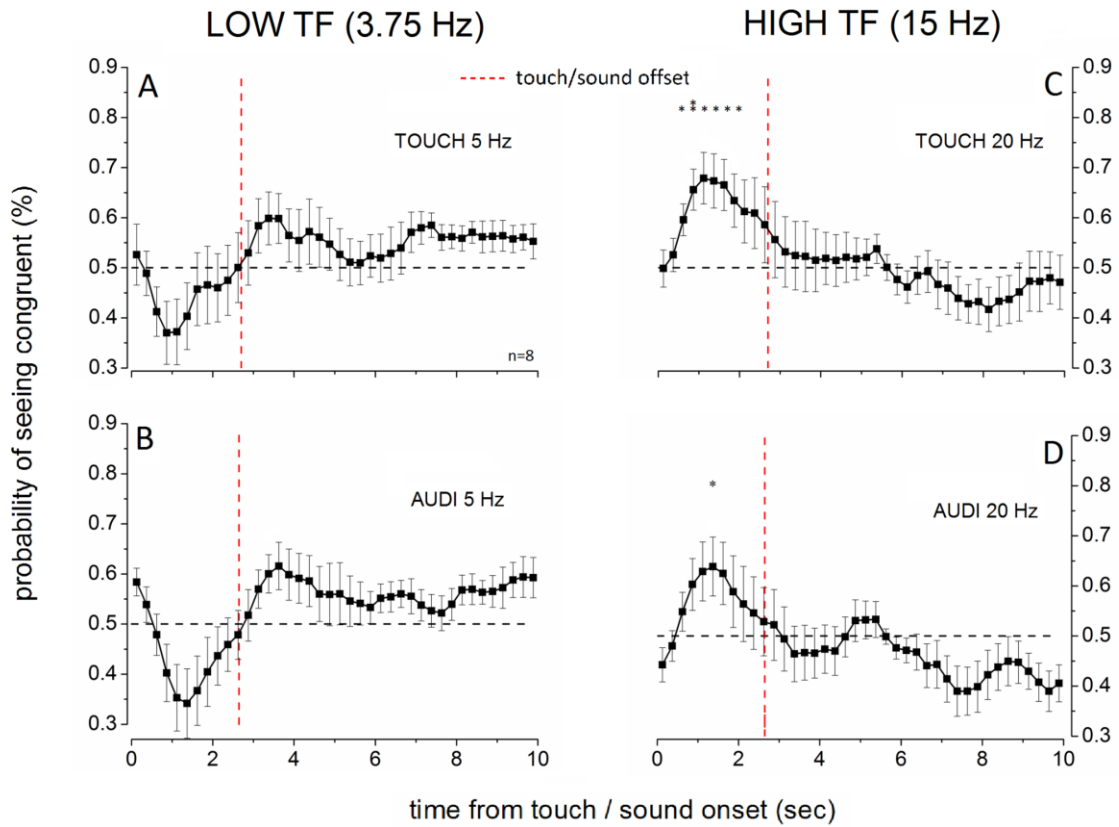


Figure 5. Frequency selectivity of the cross-modal influence on binocular rivalry. Timecourse results from an audiovisual (**A,C**) and a tactuo-visual (**B,D**) experiment in which the cross-modal stimuli were mismatched in temporal frequency by 33% (5 & 20 Hz) relative to the visual stimuli (3.75 & 15 Hz). The probability of seeing the low-frequency visual stimulus during low-frequency cross-modal stimulation (**A,B**) did not differ significantly from chance level (horizontal dashed black line), although tended to decline. The probability of seeing the high-frequency visual stimulus during high-frequency cross-modal stimulation (**C,D**) tended to increase, and did so significantly for high-frequency touch. Each data point was compared against chance level using one-sample t-tests ($N=8$, $df=7$), with p values represented by asterisks: * indicates $p \leq 0.05$; ** indicates $p \leq 0.01$. The red vertical dashed line represents the offset of the cross-modal stimulation.

