DR. JOHANNA MARGARETE ZUMER (Orcid ID : 0000-0003-0419-3869)

Proposed journal section: Cognitive Neuroscience

Title: The neural mechanisms of audiotactile binding depend on asynchrony

Running title: Neural mechanisms of audiotactile binding

Johanna M. Zumer[a,b,c,d], Thomas P. White[a,b], Uta Noppeney[a,b,c,e]

[a]School of Psychology, University of Birmingham, Birmingham, B15 2TT, United Kingdom

[b]Centre for Computational Neuroscience and Cognitive Robotics, University of Birmingham, Birmingham, B15 2TT, United Kingdom

[c]Centre for Human Brain Health, University of Birmingham, Birmingham, B15 2TT, United Kingdom

[d]School of Life and Health Sciences, Aston University, Birmingham, B4 7ET, United Kingdom

[e]Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, The Netherlands

Corresponding Author: Johanna Zumer, Department of Psychology, Aston University, Birmingham, B4 7ET, United Kingdom; johanna.zumer@gmail.com

Total number of pages: 58

Total number of figures: 5 in the main manuscript and 8 supplementary

Total number of tables: 1 in the main manuscript and 1 in the supplementary

Total number of equations: 0

Total number of words in (i) the whole manuscript (including references and figure legends): 15524; of just main body text: 11144 and (ii) the Abstract: 202

Keywords: EEG; multisensory integration; redundant target effect; temporal integration window.

Abstract

Asynchrony is a critical cue informing the brain whether sensory signals are caused by a common source and should be integrated or segregated. This psychophysics-electroencephalography (EEG) study investigated the influence of asynchrony on how the brain binds audiotactile (AT) signals to enable faster responses in a redundant target paradigm. Human participants actively responded (psychophysics) or passively attended (EEG) to noise bursts, 'taps-to-the-face', and their AT combinations at seven AT asynchronies: 0, ±20, ±70, and ±500ms. Behaviourally, observers were faster at detecting AT than unisensory stimuli within a temporal integration window: the redundant target effect was maximal for synchronous stimuli and declined within a ≤70ms AT asynchrony. EEG revealed a cascade of AT interactions that relied on different neural mechanisms depending on AT asynchrony. At small (≤20ms) asynchronies, AT interactions arose for evoked response potentials (ERPs) at 110ms and ~400ms post-stimulus. Selectively at ±70ms asynchronies AT interactions were observed for the P200 ERP, theta-band inter-trial coherence (ITC) and power at ~200ms poststimulus. In conclusion, AT binding was mediated by distinct neural mechanisms depending on the asynchrony of the AT signals. Early AT interactions in ERPs and theta-band ITC and power were critical for the behavioural response facilitation within a ≤±70ms temporal integration window.

1. Introduction

Imagine sitting outside on a summer evening. Suddenly you hear a buzz and then feel a prick to your skin, as the mosquito lands. You are faster to swat it away because you first heard it

coming. This faster detection of a multisensory event, known as the redundant target effect (RTE) (Miller, 1982; Diederich & Colonius, 2004; Sperdin *et al.*, 2009), illustrates the enormous benefits of multisensory integration.

Importantly, we should integrate signals only if they arise from a common source but segregate them otherwise. Synchrony is a critical cue for indicating whether two signals come from a common source. Multisensory signals need to co-occur within a certain tolerance of asynchrony, termed a temporal integration window (TIW) (Diederich & Colonius, 2004). In particular, the RTE typically follows an inverted U-shape function (Blurton *et al.*, 2015) that is maximal for (near)-synchronous signals and tapers off with increasing asynchrony thereby moulding the TIW. Likewise, observers' perceived synchrony, the emergence of cross-modal biases, and perceptual illusions follow a similar inverted U-shape function with its exact shape varying across different behavioural measures and task-contexts (van Wassenhove *et al.*, 2007; Megevand *et al.*, 2013; Berger & Ehrsson, 2014; Donohue *et al.*, 2015).

At the neural level, multisensory influences have been identified in terms of response enhancements and suppressions, super-additive and sub-additive interactions (Meredith & Stein, 1983; Stanford *et al.*, 2005; Werner & Noppeney, 2010b), shortened neural response latencies (Rowland & Stein, 2007) and altered neural representations (Fetsch *et al.*, 2011; Rohe & Noppeney, 2015; 2016; Aller & Noppeney, 2019; Rohe *et al.*, 2019). Evidence from neuroimaging, neurophysiology, and neuroanatomy has shown that multisensory influences emerge at early and late stages of neural processing (Foxe *et al.*, 2000; Lutkenhoner *et al.*, 2002; Murray *et al.*, 2005; Senkowski *et al.*, 2008; Sperdin *et al.*, 2009; Stekelenburg & Vroomen, 2009; Mercier *et al.*, 2013; Mercier *et al.*, 2015) nearly ubiquitously in neocortex (Schroeder & Foxe, 2002; Ghazanfar & Schroeder, 2006; Lakatos *et al.*, 2007; Werner & Noppeney, 2010a;

Ibrahim *et al.*, 2016; Atilgan *et al.*, 2018). They arise already at the primary cortical level and increase progressively across the sensory processing hierarchy (Foxe & Schroeder, 2005; Bizley *et al.*, 2007; Kayser *et al.*, 2007; Dahl *et al.*, 2009).

This multi-stage and multi-site account of multisensory interplay raises the question of whether the multisensory influences are governed by the same neural mechanisms that are expressed to a variable degree across different asynchrony levels. Alternatively, multisensory interactions may at least to some extent be mediated by different neural mechanisms across asynchrony levels. Further, how do those neural effects relate to the TIW defined by behavioural indices? Moreover, recent neurophysiological studies suggest that multisensory interactions depend on the phase of ongoing neural oscillations and/or rely on mechanisms of phase resetting. For instance, Lakatos *et al.* (2007) showed that a tactile signal can reset the phase of ongoing oscillations in auditory cortices, but only for specific asynchronies.

The current study investigates whether audiotactile (AT) interactions rely on the same or different neural mechanisms across AT asynchrony levels. Unlike previous research that selectively focused only one particular multisensory interaction feature, we assessed multisensory interactions comprehensively for evoked response potentials (ERP), inter-trial coherence (ITC), and induced power responses and related those to the TIW derived from behavioural response facilitation. Given previous unisensory research showing an increase in the TIW along the sensory processing hierarchy (Hasson *et al.*, 2008; Kiebel *et al.*, 2008), we expected that early multisensory interactions potentially in low level sensory areas are confined to narrower temporal integration windows. By contrast, AT interactions that are less sensitive to AT asynchrony and hence not confined to near-synchronous AT signals may occur at later stages in higher order association cortices (Werner & Noppeney, 2011). Participants were presented

with brief airpuff noise bursts, 'taps to the face', and their AT combinations at seven levels of asynchrony: 0, ±20, ±70, and ±500 ms. Because AT interactions may be particularly relevant during low vigilance states, when the responses to a unisensory stimulus may be weak and therefore combine into multisensory enhancement according to the 'principle of inverse effectiveness' (Meredith & Stein, 1983), it is an exciting and to our knowledge unexplored avenue to assess AT integration in the evening. In the psychophysics study observers were instructed to respond to all A, T, and AT events in a redundant target paradigm; in the EEG study a passive stimulation design was used to avoid response confounds. We then compared the multisensory influences in terms of multisensory interactions (i.e. AT + No stimulation ≠ A + T) across AT asynchrony levels for ERPs, ITC, and time-frequency (TF) power responses and characterised their topography across post-stimulus time.

2. Materials and Methods

2.1 Participants. Twenty-five healthy, adult participants with no neurological or sleep disorder were recruited from the local university population (students as well as members of the general public) (N=25, 12 female and 13 male; aged between 18-35 years old). One participant was excluded due to an abnormal finding in the structural MRI. Two participants were excluded from the behavioural analysis, because data were not collected for all conditions. Two different participants were excluded from the EEG analysis, because insufficient EEG data were collected. As a result, we included 22 participants in each of the behavioural and EEG analysis. They gave written informed consent and were compensated with either cash or course credit. The consent forms did not explicitly ask participants if their data may be publicly shared anonymously; hence, the data cannot be shared. Ethical approval for the study was given by the University of

Birmingham Science, Technology, Engineering, and Mathematics Review Committee with approval number ERN_11-0429AP22B.

2.2 Stimulation. Tactile stimulation consisted of a touch to the left side of the face with 200 ms duration. Tactile stimulation to the face was used as an ecologically valid stimulus that requires a rapid response in everyday life. We also chose stimulation to the face (in contrast to hands), as this body location does not require additional processing (e.g. reference frame transformations across the senses) of being potentially crossed relative to body position, thus potentially amenable to a quicker and more automatic route. Furthermore, the auditory association areas that receive feed-forward (layer 4) input from somatosensory stimulation appear to be optimally stimulated by cutaneous stimulation of the head and neck (Fu *et al.*, 2003). The left side was chosen based on previous findings that multisensory integration is enhanced with left-side stimulation and right hemisphere involvement (Giard & Peronnet, 1999; Downar *et al.*, 2000; Molholm *et al.*, 2002; Hoefer *et al.*, 2013). The part of the face touched was on/near the border between the maxillary (V2) and mandibular (V3) divisions of the trigeminal cranial nerve. A fibre optic cable (part of a fibre optic system: Keyence series FS-N, Neu-Isenburg, Germany) was attached to a Lego pneumatic cylinder and driven to move by pressurised air. The tip of this cable (3 mm diameter) was positioned near the face using a flexible plastic snap-together 'goose-neck' pipe that was attached to an adjustable stand. The air pressure changes were controlled by a microcontroller connected via USB to the stimulus computer; communication to the microcontroller was sent via serial port commands in MATLAB (Mathworks, Inc.). The duration of the open valve (i.e. when the diode was extended forward to touch the skin) was set to 200 ms, an ecologically/environmentally valid duration. The fibre optic cable contained a dual fibre: one fibre projected light and the other was a photodiode that

detected the light reflectance; from this, the reflectance dynamics confirmed the exact timing of the touch to the skin. As this was a mechanical air-pressure-driven device, it does not have an immediate on/off time (see Figure S1 for a plot of the reflectance data for one trial). This tactile apparatus was very similar to that used by Leonardelli *et al.* (2015). After the experiment, subjects were queried as to whether they could hear any noises of the tactile device and none reported that they could.

The auditory stimulus (target) was an airpuff noise of 200 ms duration with broadband spectral content (Figure S2 for spectrum plot). The volume of the target was well above threshold for detection but not painfully loud; the volume was stronger on the left channel than on the right (interaural intensity difference) to create the perception of coming from the left. A constant background noise of a recording of a magnetic resonance imaging (MRI) echo-planar imaging sequence (obtained from http://cubricmri.blogspot.co.uk/2012/08/scanner-sounds.html) was played to help mask external noises including those made by the tactile stimulator and for comparison with potential future functional MRI studies. The volume of the background noise, equally loud in both ears, was played at a level comfortable to participants and such that the tactile noises could not be heard. All sounds were presented via E-A-RTone earphone (10 Ohm; E-A-R Auditory Systems) with plastic tube connection (length = 75 cm) to foam ear insert (E-A-RLink size 3A), which also acted as an earplug against external sounds.

2.3 Experimental design. Participants took part in one psychophysics and one EEG session on separate days (typically 4-6 days gap). The experimental design and stimuli were identical across the two sessions. In the psychophysics session participants responded to the first stimulus in a trial irrespective of sensory modality, as fast as possible via a single key board button (i.e. redundant target paradigm). In the EEG session, participants passively perceived the stimuli

without an explicit response in order to examine automatic AT interactions (including during unattentive and drowsy states but excluding sleep stages), in order to avoid motor confounds, and to allow for comparison with sleep, non-responsive patients, etc. The EEG session was acquired one hour before the participant's usual bedtime as part of a varying vigilance study (n.b. sleep data will be reported in a separate communication). To ensure that participants were not asleep, we applied sleep staging and excluded data in actual sleep stages (details below). Hence, in this communication, we focus on multisensory interactions in a low vigilance state that have rarely been studied or reported. Yet, multisensory interactions may be most relevant in low vigilance states to attract observers' attention to salient events in their environment.

In each session, participants were presented with the following ten trial types: no stimulus (or null) condition (N), tactile alone (T), auditory alone (A), and seven audiotactile (AT) conditions varying in asynchrony (-500 ms, -70 ms, -20 ms, 0 ms, 20 ms, 70 ms, 500 ms) where a 'negative' asynchrony refers to A-leading-T (Figure 1a). The audiotactile conditions are referred to by the following abbreviations: AT500, AT70, AT20, AT0, TA20, TA70, TA500, respectively. These asynchronies were chosen to fall either within the behaviourally-defined temporal integration window (TIW) (≤70 ms) based on previous studies (e.g. (Navarra *et al.*, 2007; Harrar & Harris, 2008; Nishi *et al.*, 2014)) or outside the TIW (± 500 ms). Hence, this study focused on a coarse characterisation of the temporal integration window across both A-leading and T-leading asynchronies rather than a fine-grained analysis of asynchronies within a small range (e.g. as in (Naue *et al.*, 2011)). Seven AT, unisensory A, T and null-trials were presented, interleaved randomly with an inter-trial interval uniformly distributed between 2.0 – 3.5 s, including both unisensory and audiotactile conditions with varying asynchronies between the sensory stimuli. Each trial type was presented 100 times in each session. Trials were presented in blocks of 250

trials (roughly 11.75 minutes) over four blocks separated by short breaks. In the EEG session we occasionally shortened the blocks, but still presented 1000 trials in total. In the psychophysics session the AT500 and TA500 conditions were not collected for two participants; thus for behavioural results, only the data from the remaining twenty-two participants are included (after exclusion also of one participant for the afore-mentioned structural MRI abnormality).

Participants kept their eyes closed to obliterate any visual input throughout the experiment. They were seated comfortably with their head stabilised in an adjustable chin rest and were requested to hold their head as still as possible (to promote spatial and temporal consistency of the tactile stimulation over trials).

2.4 EEG recording. EEG data were recorded with a 64 channel BrainProducts MR-compatible cap at 1000 Hz sampling rate, with 63 of the electrodes on the scalp. For all but the first three participants, two additional bipolar electrodes were placed on the face to record horizontal electro-oculargram (EOG) and vertical EOG. For 17 participants, the 64th cap electrode was placed on the participants' back for recording ECG. For the other 8 participants, the 64th electrode was instead placed on the right (unstimulated) cheek for assistance as EOG/electromyogram (EMG). Signals were digitised at 5000 Hz with an anti-aliasing filter of 1000 Hz, then down-sampled to 1000 Hz with a high-pass filter of 0.1 Hz and low-pass filter of 250 Hz, using the software filters available in the BrainProducts acquisition setup; both of the high-pass and low-pass filters had a slope of 12 dB/octave. Electrode impedances were kept below 25 kOhm. Triggers from the stimulus-control computer were sent via LabJack to the EEG acquisition computer.

2.5 Tactile stimulation output: The time course of light reflectance (one example trial depicted in Figure S1) was assessed for each tactile trial to ensure that i. the tactile device actually

touched the skin and ii. to determine the touch onset time (1000 Hz sampling rate). After computing the actual onset of the touch from the light reflectance data, subsequently the exact multisensory onset asynchrony was computed for all multisensory trials. Those that deviated by more than ± 5 ms from the desired asynchrony were discarded. This resulted in 16.8% (± 1.1%) and 16.4% (± 1.2%) of trials rejected for the behavioural and EEG data, respectively (N=24, after excluding the participant with structural MRI abnormality).

2.6 Behavioural analysis. The reaction time data was assessed for multisensory integration effects in two main ways: through the *redundant target effect* (Hershenson, 1962) and the *race model inequality* (RMI) (Miller, 1982). The redundant target effect (RTE) compares the fastest of the unisensory responses against the multisensory response to test whether the response time (RT) is sped up for the multisensory above and beyond the fastest "channel" through unisensory stimulation; this typically will compare the median of one response type against the median of another. The race model, in contrast, assesses the "statistical facilitation" over the whole distribution, as on a given trial, slow processing in one channel may be made up for (in the "race" to process and respond) by fast processing in another; hence, the overall distribution of response times across each unisensory and the multisensory conditions is compared in the race model inequality (RMI). As both RTE and RMI have their relative advantages and limitations, we tested our data in both types of analysis. For all analyses, trials were excluded in which the touch was not actually applied (in an intended T or AT trial) or the actual touch timing was outside the intended asynchrony (in an AT trial).

The median RT within a condition for each participant was computed. The RTE was computed for each participant by subtracting the median RT of the AT condition at a particular level of asynchrony from the median RT of the fastest (A or T) unisensory condition, with the

onset of each unisensory condition adjusted for the particular asynchrony (e.g. $RT_{AT20} - \min(RT_T + 20\text{ ms}, RT_A)$. In addition to the trial exclusion mentioned above, also sensory trials were additionally discarded with no response or with a response time faster than 100 ms or slower than 1 s (occurring in total for an average of 2.7±1.1% of trials across conditions). First, we investigated whether the redundant target effect was different across AT asynchrony conditions, using a one-way repeated-measures (i.e. dependent-samples) ANOVA (rmANOVA) over the seven AT conditions followed by planned post-hoc rmANOVA tests to further narrow the possible asynchrony conditions that drive the overall main effect of asynchrony (i.e. conditions ≤ 70 ms, and if that is significant, then ≤20 ms). Second, we assessed whether the redundant target effect for each condition differed significantly from zero across participants, using a one-sample two-sided t-test.

The race model inequality tests whether the cumulative distribution of processing times for the multisensory condition ($F_{AT}(t)$, for all times t) is less than or equal to the sum of the cumulative distributions for each unisensory condition ($F_A(t) + F_T(t)$, for all times t). Because it is important for the distribution to be computed from all trials including "fast guesses, omitted responses, and outliers" so that the tails of the distribution are computed correctly, we did not omit these trial types but rather accounted for them appropriately, according to the full explanation in Gondan and Minakata (2016). Trials with an omitted response were assigned an RT of ~10 s (plus small jitter), namely a value that is much greater than the maximum trial length. The cumulative distributions for each condition as well as the predicted sum $F_A(t) + F_T(t)$ was computed using the Matlab code RaceModel.m (Ulrich *et al.*, 2007). We modified this function to correct for the "fast guesses" by taking in as additional input the response times for the null "catch" trials, referred to as the "kill the twin correction" as discussed in (Gondan &

Minakata, 2016). The unisensory response times were adjusted (e.g. 20 ms added to the tactile RT) according to the asynchrony of the audiotactile condition (e.g. AT20) against which the unisensory were being compared. The distributions were binned in to deciles. A one-sided t-test was then computed to test if *any* decile bin of the actual cumulative distribution F_{AT} was less than the predicted cumulative distribution of the sum $F_A(t) + F_T(t)$; if so, then the race model is violated and multisensory integration is inferred.

2.7 EEG analysis.

2.7.1 EEG analysis: sleep staging. To ensure that only EEG data was used in which participants were awake, given the passive stimulation design with eyes closed and the evening acquisition, standard sleep scoring was performed using American Academy of Sleep medicine (AASM) 2007 criteria in the FASST open-source software (http://www.montefiore.ulg.ac.be/~phillips/FASST.html) (Leclercq *et al.*, 2011) and custom code in MATLAB (code available upon request). Data were segmented into 30 s chunks and referenced to linked-mastoids. In order to stage the sleep data, two of the authors (J.M.Z. and T.P.W.) studied the AASM manual, were instructed by a local clinical neurophysiologist who specialised in clinical sleep assessments, and worked through example datasets with other researchers who staged sleep as routine in their research (see acknowledgments). After training, the two authors scored the EEG data independently with an initial correspondence of 88%. Differences were discussed and a consensus reached (with correspondence of the consensus to each assessor's scores at 93% and 94%). We assigned standard AASM stages of Awake, Stage Non-REM 1, 2, 3, and REM. Any 30 s chunk that was not scored as Awake was excluded from further analysis. If an individual participant had fewer than 55 trials per condition remaining in

the Awake stage (prior to artefact rejection), the participant was fully excluded. Two participants were excluded for this reason.

We did not further break down the classification to the sub-stages of drowsiness and light sleep, such as using the Hori nine-stage classification for the light stages (Tanaka *et al.*, 1996). Distinguishing between the first two Hori stages based solely on alpha power is more difficult in the subset of participants who do not display obvious/strong alpha waves. We therefore included data combined across states of awareness (i.e. the first two Hori stages), from most alert to very drowsy pre-sleep. We would not have enough data in this study to independently analyse each sub-stage of wakefulness and we did not set out to make this comparison in our design of the study.

There is a remote possibility of REM sleep occurring first without progression through other sleep stages first, although this is most likely to occur in narcolepsy, other sleep disorders, or after REM-sleep deprivation (e.g. (Littner *et al.*, 2005)). In the eight participants in which we recorded EMG activity, we did not observe any REM-like muscle activity. Moreover, none of the participants, based on self-report, showed any signs of sleep disorder or sleep deprivation the night before (sleep duration = 7:40 ± 1:37, hours:minutes). Together, this renders "sleep onset REM" in this dataset extremely unlikely.

2.7.2 EEG analysis: preprocessing: All subsequent EEG data processing (after sleep staging) was performed using the open-source toolbox FieldTrip (Oostenveld *et al.*, 2011) (www.fieldtriptoolbox.org) and custom code in MATLAB (code available upon request). Eye movement artefacts were automatically detected using three re-referenced bipolar pairs ('F7-F8', 'Fp2-FT9', and 'Fp1-FT10') and the VEOG if available. These channels' data were band-pass filtered (1-16 Hz; Butterworth, order 3, two-pass) and transformed to z-values. The exclusion

threshold was set at a z-value of 6 and trials containing these artefacts were excluded. Note that this eye movement artefact rejection was amplitude-based only and did not distinguish between faster eye movements (e.g. saccades) and the slower roving eye movements typical of drowsiness; blinks were minimal as eyes were to remain closed during the stimulus presentation.

EEG data (over all channels for the main analysis) were re-referenced to the average reference, high-pass filtered (0.2 Hz; Butterworth order 3, two-pass), band-stop filtered around the line noise and its harmonics (49-51 Hz, 99-101 Hz, and 149-151 Hz; Butterworth order 4, two-pass), and epoched for each trial. Trials were locked to the onset of the tactile stimulus for tactile and all multisensory conditions and to the auditory or null trigger for A and N conditions, respectively. Initially, the epoch length was from 1500 ms to 2300 ms. Then A trials were shifted ± 500, 70, 20, or 0 ms before being added to a T trial, to create the appropriate A+T combination to contrast with AT trials, hence resulting in variable lengths of pre-stimulus and post-stimulus window lengths, depending on the AT asynchrony.

2.7.3 EEG analysis: multisensory interaction. Multisensory integration in the EEG data was identified in terms of "audiotactile interaction", i.e. the sum of unisensory (A+T) contrasted to the audiotactile plus null (AT+N). The sum of unisensory (A+T) trials was computed for each AT asynchrony level such that the onsets of the auditory and tactile stimuli were exactly aligned to the trials of the AT condition (i.e. we also accounted for the jitter of tactile onsets, see above). Trials from each condition were randomly sub-selected to ensure an equal number of trials per each of the four conditions in a given contrast (A, T, AT, and N). Actual trial numbers for a given participant ranged from a minimum of 47 to a maximum of 92, with a mean over all participants of 66.0 and a median of 66. It is critical to add the null condition (to the multisensory) to account for non-specific effects in a trial such as expectancy of stimulation as

well as random noise. Note that removing a pre-stimulus baseline (or, in the present analysis, 0.2 Hz high-pass filter) for removing drift and DC offset is not sufficient to account for these non-specific effects or 'spontaneous activity'. The argument for a null condition is exemplified in Teder-Salejarvi *et al.* (2002) and further supported and utilised in other multisensory experiments (Talsma & Woldorff, 2005; Bonath *et al.*, 2007; Mishra *et al.*, 2007). Further, the multisensory interaction contrast (with the null-condition included) is equivalent to subtracting the Null condition from each of the stimulus conditions (i.e. an explicit baseline correction): (AT-N)-[(A-N)+(T-N)] = (AT+N)-(A+T). In addition to controlling for the pre-stimulus 'stimulus expectancy' confound (as explained in Teder-Salejarvi *et al.* (2002)), including the Null condition in the interaction contrast also ensures that 'random noise' is averaged out similarly for the sum of the two unisensory and the multisensory + Null sum. We also alleviated 'anticipatory' or 'omission' waves in our data by using a jittered inter-trial interval (uniformly between 2.0-3.5 s).

2.7.4 EEG analysis: multisensory effects on ERP, inter-trial coherence, and time-frequency power. For the evoked response potential (ERP) analysis, EEG data were low-pass filtered (40 Hz; two-pass Butterworth filter order 6). The average over trials within a participant was computed for the combination of conditions A+T and AT+N separately. We assessed the AT interaction within a 500 ms time window, beginning at the onset of the second stimulus.

For time-frequency analysis, EEG data were Fourier transformed with separate parameters for lower (4-30 Hz) and higher (30-80 Hz) frequencies, with zero padding to a 4 s length (applied to both lower and higher frequencies). Sliding time windows of length equal to four cycles (low frequencies) or 200 ms (high frequencies) at a given frequency in steps of 2 Hz (low frequencies) or 5 Hz (high frequencies), after application of a Hanning taper (low frequencies) or multitaper

with +/- 7 Hz smoothing (high frequencies). The complex values were kept for separate analysis of the inter-trial coherence (ITC) (also referred to as phase-locking factor or phase-consistency index) and the time-frequency (TF) power magnitude. Note that the sum of randomly paired individual trials of different condition types (i.e. A+T and AT+N) was computed prior to Fourier transformation so that any cancellation due to phase differences would occur prior to obtaining the Fourier complex value (see Senkowski *et al.* (2007)). The ITC was computed for each 'condition' (with 'condition' here meaning either A+T or AT+N) and subject as the absolute value of the sum of the complex values over 'trials' (where the 'trial' here refers to a sum of individual trials of A and T or of AT and N). We assessed the AT interactions for ITC and TF power separately for 'low frequency' and 'high frequency', within a 1200 ms time window beginning at the onset of the second stimulus and extending to include the low frequency (e.g. alpha and beta) desynchronization / rebound effects. We averaged data across frequencies within each predetermined band (4-6 Hz for theta, 8-12 Hz for alpha, 14-30 Hz for beta) so as to obtain results specific to a band for ease of interpretation.

2.7.5 EEG analysis: statistics across and within asynchronies. Our two main statistical analyses are outlined in Figure 2a. First, we investigated whether the AT "interaction" [(A+T)-(AT+N)] differed *across* the 7 asynchronies in a one-way rmANOVA (Figure 2a.1). Please note that this rmANOVA effectively tests for a three way interaction, i.e. a modulatory effect of asynchrony on AT interactions, with the AT interaction consisting of the 2x2 factors of A (present or absent) and T (present or absent). Second, we also assessed whether the AT interaction was significant (i.e. different from zero) *within* each condition, i.e. separately for each asynchrony level, through a paired (i.e. dependent samples) t-test (Figure 2a.2) (i.e. contrasting

(A+T) versus (AT+N) for each asynchrony). The rmANOVA and t-tests were performed separately for ERP, ITC, and TF power.

To correct for multiple comparisons (over channels and time), we performed non-parametric cluster-based permutation tests for dependent (i.e. paired) samples, with the sum of the statistic (F or t-values) (i.e. max sum) across a cluster as cluster-level statistic and points for a cluster initially detected at an auxiliary uncorrected alpha threshold of 0.05 (Maris & Oostenveld, 2007). All statistical results from power and ITC between 4-30 Hz were further corrected for testing over three frequency bands by dividing the p-value threshold by three (0.05 / 3 = 0.0167).

We illustrate a significant effect in an rmANOVA in a channel X time representation where all significant time points in a channel are highlighted (see Figure 2a, right and Figure 4a as an example). Further, we show an F-map where we show the topography of the F-values averaged across the time-window that includes time points where at least one channel was significant in the rmANOVA (see Figure 2b right and Figure 4b as an example).

Note that any significant finding from the across-asynchrony rmANOVA, which indicates a difference in the AT interaction across asynchronies, need not necessarily appear as a significant finding in the within-asynchrony t-tests, which indicate a strong AT interaction effect for that one asynchrony, and vice versa. However, any correspondence found corroborates both the presence of the within asynchrony AT interaction and its dependence on or selectivity for a particular asynchrony (i.e. significant difference in the AT interaction across asynchrony levels).

2.7.6 EEG analysis: characterisation using component analysis. The rmANOVA reveals AT interaction indices for ERP, inter-trial coherence, time-frequency power that depend significantly on the asynchrony of the AT signals. Because we entered two way AT interactions as dependent variables into a one way rmANOVA, a significant main effect in this rmANOVA effectively

represents a three way (i.e. A x T x asynchrony) interaction effect. In other words, it shows that asynchrony modulates the magnitude of the AT interaction. Because our design included 7 levels of asynchrony, a significant main (resp. three way interaction) effect can arise from various profiles. For instance, it may arise, because a two way AT interaction is only present at AT0, but not at any other asynchrony levels. Alternatively, there may be a positive AT interaction at +70 ms and a negative interaction at -70 ms. To determine which asynchronies drive the main effect in the rmANOVA we first used PCA1 to determine the relative contributions of AT interactions for each of the seven asynchrony levels to the three way interaction effect (for technical details see below). PCA1 thus provides us with a weight for each of the 7 asynchrony levels. Next, we applied these weights (that were derived only from the significant spatiotemporal cluster) to the original complete data set. We then used PCA2 to characterise the spatiotemporal evolution of this effect in terms of the topography and temporal evolution of the first principal component.

In the following we will describe PCA1 and PCA2 in greater methodological depth:

1. PCA1: The rmANOVA revealed a significant main effect as a spatiotemporal cluster. In PCA1, for each asynchrony level we selected the contrast values of the AT interaction in each channel at the time points that were significant in the rmANOVA (i.e. within the spatiotemporal cluster). The individual channel-time points *within* this rmANOVA cluster mask for each asynchrony were reshaped into a 1 X (channel * time points) vector. These seven (one for each asynchrony) vectors were concatenated into a single matrix over asynchronies (7 asynchronies x Masked-channel-time-points). This matrix was entered into the PCA1 that decomposes the matrix into a weighting (mixing) matrix that quantifies the expression over asynchrony of the principal components (PCs), which are in this case the representative masked-channel-time-point vectors; we focussed only on the first (strongest) PC as it is the one explaining the most variance.

The first column of the weighting (mixing) matrix is a 7 (asynchrony) x 1 vector indicating the contributing strength of each asynchrony AT to the main effect in the rmANOVA. For instance, in Figure 2b PCA1 indicates that the F-contrast from the rmANOVA is mainly driven by the TA70 asynchrony. In Figures 4 and 5, we then show the within-asynchrony interaction selectively for the asynchrony that received the greatest absolute weight in the PC.

2. PCA2: We characterised the spatiotemporal evolution of this effect by applying a second PCA. Prior to computing PCA2, we multiplied the weighting corresponding to the 1st PC (1 x 7 asynchrony vector) with a matrix (7 asynchrony X channel*time) containing the data from *all* channel-time points (not just those in the rmANOVA mask) of the audiotactile interactions across the 7 asynchrony levels; this multiplication results in a vector [of size 1 X channel*time] that is a linear combination of the channel*time vectors across the 7 asynchronies, weighted according to their importance determined by PCA1. We reshaped this vector back to the standard matrix [of size channel X time]. This channel X time matrix shows how differences in AT interactions across asynchrony levels (that drove the significant F-contrast in the rmANOVA) evolve over time across all channels. To illustrate this effect we decomposed this channel X time matrix into a dominant topography and its time course using PCA2. We plotted the topography and time course of the first (i.e. strongest) PC (see Figure 2b, right bottom).

We can then compare the topography and time course obtained from PCA2 with the topography and time course of the audiotactile interaction effect for the asynchrony level that received the greatest weight in the PCA1 (Figure 2b left). These comparisons are quantified by a spatial correlation (between topographies) and temporal correlation (of the time courses). We acknowledge that the correlations will be biased to be high as they are partially based on the

same data input; we have computed and presented them to aid the visual comparison, but the values are not to be taken in statistical rigour.

We performed these PCA1 and 2 using custom code in MATLAB and is akin to other two-stage component analyses in which the weighting of a component along a dimension is illustrated (such as networks along frequency bands as in Brookes *et al.* (2012)). In the case of the ERP we performed this analysis separately for the three distinct sub-clusters of significance (Figure 4a). The division into sub-clusters was along temporal (not spatial) boundaries, placed when there was minimal significant channels between the sub-clusters.

3. Results

For the psychophysics study we report the redundant target effect as a behavioural index of audiotactile integration for each asynchrony level and contrasted across asynchronies. For the EEG data we assess how the multisensory interactions (AT+N ≠ A+T) for ERPs, inter-trial coherence (ITC), and time-frequency (TF) power differ across asynchrony levels in a rmANOVA. As described in detail in the methods section we then characterise the spatiotemporal profile of how this AT interaction effect depends on AT asynchrony by applying a first and second stage principal component analysis. Moreover, we report the audiotactile interaction for the asynchrony level that mainly drives the effect in the rmANOVA as indicated by its weight in the first principal component analysis (i.e. the AT interaction for the asynchrony level with maximal PC weight). For completeness and full characterisation of the data, the supplementary materials report all audiotactile interactions for each asynchrony level that was significant when tested independently in t-tests at a particular asynchrony level (i.e. at one of the seven levels of AT asynchrony: 0, ±20, ±70, and ±500 ms (Figure 1a).

3.1 Behavioural results: reaction time facilitation tapered by TIW

We examined the reaction time data for indication of multisensory facilitation through two assessments: the redundant target effect (RTE) and the race model inequality (RMI) (see Methods). As expected, we observed significantly faster (Table S1 for p-values and t-values) median response times for the AT relative to the fastest unisensory condition (i.e. *redundant target effect*) for asynchronies within a ≤ 70 ms window of integration (Figure 1b). Specifically, the significantly faster RTEs (across subjects mean ± SEM) for the different asynchrony levels were: AT70 = 35 ± 6 ms, AT20= 38 ± 5 ms, AT0 = 35 ± 4 ms, TA20 = 33 ± 4 ms, and TA70 = 24 ± 4 ms. Surprisingly, we observed significantly slower response times for the AT500 relative to the unisensory auditory condition, i.e. a negative redundant target effect (across subjects' mean ± SEM) = -16 ± 4 ms. The RTE for the TA500 condition was not significantly different from the unisensory tactile condition (3 ± 4 ms). Note also that the false alarm rate (responding to a null/catch trial) was 0% for all participants.

Furthermore, we found that the redundant target effect differed significantly across the seven asynchrony conditions (one-way repeated-measures ANOVA (rmANOVA); $F_{6,12}=25.4$, p<0.001). As a planned post-hoc test, we further found that, when restricting the conditions to those ≤ 70 ms asynchrony, the redundant target effect differed significantly across the five conditions within ≤ 70 ms (one-way rmANOVA; $F_{4,14}=3.5$, p=0.01). When restricting the comparison across conditions to those ≤ 20 ms, we did not find any significant difference (one-way rmANOVA; $F_{2,16}=1.89$, p>0.1).

Secondly, we tested the data from all seven asynchronies in the race model, which assesses the differences in distributions (not just median) of reaction times. If the reaction time

for *any* decile of the cumulative distribution of the audiotactile RTs is smaller (earlier) than the RT for the same decile of the summed cumulative distribution of the unisensory distributions, then the race model inequality is violated and multisensory response time facilitation (i.e. integration) is assumed. To account for "fast guesses", the sum of the cumulative distributions of the AT and N (null/catch) trials was used instead of just the AT cumulative distribution (see Methods). We chose to compare the first half of deciles (i.e. 10%, 20%, 30%, 40%, and 50%) as typically the RMI is violated in the faster response times; we then applied Bonferroni correction to the p-values to account for both the five decile tests as well as the seven asynchronies (thus a threshold of $p<0.05/35=0.0014$). The five middle AT asynchronies (within ≤ 70 ms) all showed RMI violation for at least one decile; see Table 1 for full results.

In summary, our psychophysics study revealed that audiotactile interactions within a 70 ms temporal integration window (TIW) facilitate stimulus processing and response selection leading to faster response times. Furthermore, the response facilitation varied significantly across synchronies within 70 ms, while for near-synchronous stimuli within 20 ms they seem to be comparable. This behavioural profiles raises the question whether the same neural mechanism mediates AT interactions across all asynchronies, but is attenuated for greater AT asynchronies. Alternatively, different neural mechanisms may be engaged at different AT asynchronies.

3.2 Audiotactile interactions for ERPs: limited to the behavioural TIW

Figure 3a shows the ERPs for the A, T, AT and N conditions. Both tactile-alone (pink) and auditory-alone (green) stimulation evoked a characteristic N100 followed by a P200, while the null condition is a flat baseline. The tactile and auditory stimulation together generate the AT evoked potentials across the different asynchrony levels (Figure 3a, black). While the influences

of both the tactile and auditory evoked responses are clearly visible in the AT responses, we can also observe small deviations from the unisensory responses.

In the following, we investigate whether the "audiotactile interaction" ([AT+N]-[A+T]) significantly varies across asynchronies. If we observe a significant modulation of the AT interactions by asynchrony level in the rmANOVA, we then assess which asynchronies drive the effect based on the weights PCA1. Further we report the AT interaction for the asynchrony level that is associated with the strongest weight in the PCA. Finally, a full discussion of the individual asynchrony results follows at the end of this section.

As shown in Figure 4, the rmANOVA across asynchronies for ERP revealed a near-threshold significant (p=0.065) cluster (shown in yellow) in an early time window (50-150 ms) (Figure 4a in yellow). Details of all statistical findings are in Table S1. The weights of the 1st PC (explaining 39.3% of the variance) (Figure 4b-i) indicate that this effect is mainly driven by AT interactions expressed for synchronous and to some extent also for near synchronous AT stimuli. Indeed, when testing for AT interactions separately for each asynchrony level we observed, for this early time window, a significant AT interaction only for AT synchronous stimuli (AT0) and trends also for the near synchronous stimuli ≤20 ms (i.e. AT±20 ms, see Figure 3b). Further, Figure 4c and 4d indicate that the AT interactions in AT0 evolve during and after the N100 (70-170 ms), in both central and posterior sensors, with A+T being initially more positive and then less negative than AT+N. The topographies and time courses of (c-i) and (d-i) are similar; the topography spatial correlation between these is 0.88 (p<0.001; N channels = 61) and the time course temporal correlation between these is 0.96 (p<0.001; N time points 501).

The rmANOVA also revealed a significant temporally-extensive cluster (p=0.0005), spanning from about 150 ms to the end of the window tested at 500 ms (highlighted in light blue in the

channel-time image in Figure 4a). This later ERP cluster included 3 'subclusters' that were segregated in time, although linked together via a 'bridge' across channel-time space. Because the 3rd subcluster can be attributed to eye artifacts based on it spatiotemporal and asynchrony profile, we do not discuss it further in the main manuscript (for completeness we show it in Figure S3).

The first subcluster extended from 180 to 270 ms (Figure 5a). As shown in Figure 5a, it emerged with a similar spatiotemporal profile and was expressed across asynchrony levels similar to the audiotactile interactions observed for theta-band ITC and TF power; hence, it was plotted alongside these results. For values of correlation between this ERP effect (Figure 5i) and the theta-band ITC (5ii) and theta-band power (5iii), see the section on theta power. The F-values for the modulatory effects of asynchrony (i.e. rmANOVA) were most pronounced over frontocentral electrodes (Figure 5b-i). Note that the concentration at vertex in Figure 5b-i (and b-ii and b-iii) is not artefactual: the F map portrays a signal-to-noise ratio where the conditions differ most (rather than a standard topography or difference of topographies). The first PC (explaining 31.9% of the variance) indicated that this effect was mainly driven by an audiotactile interaction at 70 ms asynchrony (AT70, Figure 5b-i and Figure 3b). Indeed, the spatiotemporal profile of the first PC of PCA2 (Figure 5c-i) and the audiotactile interaction for AT70 (Figure 5d-i) are very similar: they both emerge with a frontocentral topopgraphy and temporal peak around 200 ms (see also Figure 3b for similar effect for TA70). The correlation of the topographies (5c-i with 5d-i) is 0.88 (p<0.001; N = 61 channels) and of time courses is 0.94 (p<0.001; N = 501 time points). As shown in Figure 5d-i, this AT interaction modulated the shape and magnitude of the P200: the P200 peaked and declined earlier for the AT+N (dark blue) relative to A+T (light blue).

The second 'sub-cluster' (Figure 4a, ii) extended from 350 ms to 420 ms. The F-values for the modulatory effects of asynchrony (i.e. rmANOVA) were greatest over occipital electrodes. The weighting across asynchrony levels of PCA1 (explaining 43.6% of the variance) indicates that this later audiotactile interaction effect was most pronounced for near-synchronous conditions ≤ 20ms and particularly for AT20. The further spatiotemporal characterisation of this effect (via PCA2) indicates a topography that varies from front to back and a time course with peak ~400ms (Figure 4c-ii). This spatiotemporal profile was also found for the audiotactile interaction AT20 (Figure 4d-ii) and TA20 (Figure 3b). The topographies and time courses of (c-ii) and (d-ii) are similar; the topography spatial correlation between these is 0.95 (p<0.001; N = 61 channels) and the time course temporal correlation between these is 0.97 (p<0.001; N = 501 time points). Moreover, we note the similarities between the early (100ms; Figure 4-i) and late (400ms; Figure 4-ii) effects in the ERP, both from rmANOVA (and subsequent PCA) and the within-asynchrony AT interactions. Specifically, the topographies from PCA2 between the early and later clusters had a spatial correlation of 0.99 (p<0.001; N = 61 channels) and their time courses a correlation of 0.98 (p<0.001; N = 501 time points) (i.e. correlating Figure 4c-i with 4c-ii); the topographies of the individual asynchronies associated with the early and later ERP clusters had a spatial correlation of 0.65 (p<0.001; N = 61 channels) and the time courses a correlation of 0.86 (p<0.001; N = 501 time points) (i.e. correlating Figure 4d-i with 4d-ii).

Figure 3b shows the ERPs for the sum over A+T (dark blue), sum over AT + N (light blue), and the difference (A+T) − (AT + N), i.e. the audiotactile interaction effects across different asynchrony levels. For ERPs we observed three AT interaction effects that differed in their expression across levels of AT asynchrony (corresponding to the rmANOVA results above).

The first AT interaction effect arose early, at about 100 ms post-stimulus, with a central topography and was significant only for the synchronous condition (Figure 3b, AT0 row). Specifically, a modulation, during and after the N100 (70-170 ms), was found in both central and posterior sensors, with the A+T greater than the AT+N during this time. We note that a trend for this spatiotemporal effect was also observed for the AT20 condition.

The second AT interaction effect emerged at about 200 ms after the second stimulus (latency range: 140-220 ms), was most pronounced over frontocentral electrodes, and was selective for the asynchrony of ±70 ms (Figure 3b, AT70 and TA70 rows). This AT interaction modulated the shape and magnitude of the P200: the P200 occurred earlier and was reduced in amplitude for the AT+N relative to A+T.

The third AT interaction effect, where A+T was more negative than the AT+N, arose later at about 370-400 ms mainly over posterior electrodes for AT asynchrony conditions within a ≤ 20 ms temporal integration window (Figure 3b, AT20, AT0, and TA20 rows). Even though this AT interaction effect was significant only for AT20 and TA20, we observed a qualitatively similar pattern for the synchronous AT0 condition.

In summary, we observed three distinct AT interaction effects for ERPs, all limited to AT asynchrony levels within the behavioural ≤ 70 ms TIW. The AT interactions at 100 ms and 400 ms were expressed mainly for synchronous and near synchronous AT stimuli. The AT interactions at 200 ms were mostly selective for 70 ms asynchrony and, as we will see in the next sections, related to AT interactions expressed in ITC and theta oscillatory power.

3.3 Audiotactile interactions for ITC: selective for ±70 ms asynchronies

The across-asynchrony rmANOVA revealed that the AT interaction for the theta-band ITC differed significantly across the 7 asynchronies (p=0.0005; Figure 5a-ii and Table S1 for statistics details). The topography of the maximal effect was also frontocentral (Figure 5b-ii) in line with the P200 ERP (Figure 5b-i). The first PC from PCA1 (explaining 35.9% of the variance) highlighted that the TA70 condition was the strongest driver (Figure 5b-ii). PCA2 revealed a spatiotemporal profile with a central topography most prominent around 200ms (Figure 5c-ii). Again this spatiotemporal profile was similar to the within-asynchrony audiotactile interactions (in this case, for +70 ms asynchrony, i.e. TA70), which also peaked at about 200 ms with a central topography (Figure 5d-ii) - thereby mimicking the AT interactions we observed for the P200 in the ERP analysis (Figure 5i). The correlation of the topographies (5c-ii with 5d-ii) is 0.90 (p<0.001; N = 63 channels) and of time courses is 0.96 (p<0.001; N = 106 time points). As shown in the supplementary results (Figure S4) we also observed a similar AT interaction effect for -70 ms asynchrony (i.e. AT70). Surprisingly, as seen in both the PCA1 weightings across asynchrony levels (Figure 5b-ii) as well as within-asynchrony multisensory integration effects (Figure S4), the summed 'A+T' ITC was *smaller* than the summed 'AT+N' for the AT70, but *greater* for tactile leading TA70 condition. Thus, the direction of the audiotactile theta-band ITC interaction depends on whether the auditory or the tactile sense is leading. To understand better how this opposite-sign effect in ITC can occur at the same time as the ERP effects (and TFP effects discussed below) being the same-sign in those two asynchronies, we performed a simulation to demonstrate one feasible scenario for the data; see supplementary section 7 for details. In brief, this simulation demonstrates that these data could occur by changes in oscillation amplitude commensurate with the ERP and theta TFP effects but

with opposing effects in (restriction of) phase consistency across trials commensurate with the theta ITC effects. No significant ITC results were found for alpha, beta, or gamma bands. In summary, the AT interactions for the theta-band ITC were most prominent for 70 ms asynchronies and most likely associated with the ERP effects at the same post-stimulus latency and asynchrony conditions.

3.4 Audiotactile interactions for time-frequency power across AT asynchronies

3.4.1 Theta power: The across-asynchrony rmANOVA revealed a marginally significant (see Methods) single cluster (p=0.024; Figure 5a-iii) primarily with frontocentral topography (Figure 5b-iii) and strongest from 200-300 ms (Figure 5a-iii). The first PC from PCA1 (explaining 53.5% of the variance) showed that TA70 (and AT70) mainly drove this difference in audiotactile interactions across asynchrony levels (Figure 5b-iii). PCA2 showed a spatiotemporal profile of this 1st principal component (Figure 5c-iii) with a frontocentral topography peaking around 200-300 ms. Likewise, the audiotactile interaction for the TA70 asynchrony level emerged with a frontocentral topography peaking at about 200-300 ms (Figure 5d-iii; see Table S1 for statistics and also Figure S5 for other asynchronies). These frontocentral AT interactions arose as a result of the AT+N power peak being weaker and decaying earlier relative to the A+T sum. The correlation of the topographies (5c-iii with 5d-iii) is 0.59 (p<0.001; N = 63 channels) and of time courses is 0.79 (p<0.001; N = 106 time points).

Figure 5 also highlights the point that the audiotactile interactions for the P200 ERP, the theta ITC and the theta TFP emerge with a similar spatiotemporal profile and were most pronounced for the ±70 ms asynchrony. Specifically, while the PCA2 topography of the P200 ERP was similar to both the theta ITC (r=0.37; p=0.003; N = 62 channels) and to the theta power (r=0.38;

p=0.002; N = 62 channels) topographies, the theta ITC and theta power topographies were not very similar (r=0.09; p=0.49; N = 62 channels) (comparing across the row of Figure 5c). In contrast to the spatial correlations, the temporal correlations of the PCA2 time courses were similar for P200 ERP and theta power (r=0.35; p=0.01; N = 51 time points), but not between the theta ITC and either P200 (r=0.01; p=0.97; N = 51 time points) or theta power (r=0.05; p=0.75; N = 51 time points); this seems due to the theta ITC interaction peaking earlier (~150 ms) compared to the P200 ERP and theta power interactions, which peak later (~200-250 ms). Comparing across the row of Figure 5d, specifically the spatial correlations of the topographies of the individual asynchronies were r=0.40 (p=0.002), r=0.36 (p=0.005), and r=0.41 (p<0.001) (N= 61 channels for all; between P200 ERP to theta ITC, P200 ERP to theta power, and theta ITC to theta power); the temporal correlations were r=0.28 (p=0.048), r=0.40 (p=0.004), and r=0.78 (p<0.001), respectively (N = 51 time points for all three).

3.4.2 Beta power: The rmANOVA revealed significant differences in audiotactile interactions in beta power across asynchronies (p=0.001; Table S1) in an early (100-350 ms) cluster (Figure S6a) with frontocentral topography (Figure S6b). The weighting across asynchronies showed dependence on (at least) four asynchronies, but not in a pattern that corresponded with either the behavioural results or those from ERP, theta ITC, or theta TFP. The asynchrony dependence highlighted a preference for sense-leading (opposing weightings for tactile leading versus auditory leading) but not in a symmetrical format (Figure S6b). For full details on the beta TFP results, see supplementary section 6.2.

3.4.3 Alpha and gamma power: The rmANOVA across asynchronies did not show any significant differences in the alpha or gamma band for power.

4. Discussion

The current study investigated whether AT integration is mediated by same or different neural mechanisms at different AT asynchrony levels and how these are related to behavioural response facilitation. We thus assessed AT interactions comprehensively for ERPs, ITC, and induced TF power across several levels of AT asynchrony.

Consistent with previous research (Colonius & Diederich, 2004), we observed an inverted U-shape function for the behavioural AT benefit – also termed the "redundant target effect" or "redundant signals effect" (Miller, 1982) – that was maximal for synchronous AT combinations and tapered off with increasing AT asynchrony within a TIW of ≤70 ms (Figure 1b) (Zampini *et al.*, 2005). Likewise, the predictions of the race model were 'violated' for asynchronies ≤70 ms. Both of these analyses suggest that observers experience benefits in audiotactile processing for TIW of ≤70 ms, though some of these reaction time differences may reflect modality-switching costs as a result of random ordering the auditory, tactile and audiotactile conditions (Shaw et al., 2020; Crosse et al., 2019).

At the neural level we observed early AT interactions for evoked responses (ERP) at about 110 ms post-stimulus (Figures 4 and 3b), which dovetails nicely with previous research showing multisensory modulations of the N1 auditory component by visual and tactile stimuli (Foxe *et al.*, 2000; Lutkenhoner *et al.*, 2002; Murray *et al.*, 2005; Sperdin *et al.*, 2009; Stekelenburg & Vroomen, 2009). Critically, the novel finding here was that the early AT interactions were sensitive to the relative timing of the AT stimuli: they were most pronounced for synchronous

AT stimuli and tapered off within a small TIW of ≤20 ms (Figure 4b-i. weights across asynchrony). This temporal precision may be enhanced for interactions of tactile (in particular) with other sensory signals, because tactile latencies are fixed for a particular body location and do not vary depending on the distance of the stimulus from the observer as in audition and vision. The short latency and narrow temporal binding window points towards neural interactions in low level or even primary auditory cortices that may rely on direct connectivity between sensory areas (Fu *et al.*, 2003; Cappe & Barone, 2005; de la Mothe *et al.*, 2006a; Smiley *et al.*, 2007) or thalamic mechanisms (de la Mothe *et al.*, 2006b; Hackett *et al.*, 2007; Cappe *et al.*, 2009) and that increase the saliency of AT events most likely leading to faster and more accurate detection in our psychophysics study.

Later, at about 400 ms post-stimulus, we observed audiotactile ERP interactions that were again most pronounced for synchronous AT stimuli and confined to the TIW of ≤20 ms (Figures 4-ii and 3b). These later interactions may reflect top-down modulatory neural processes in lower regions via feed-back loops (Falchier *et al.*, 2002; Schroeder & Foxe, 2002; Clavagnier *et al.*, 2004). The expression of both early and late ERP interactions followed a U-shape function (Figure 4b-i and 4b-ii) thereby mimicking the asynchrony profile of the redundant target effect that characterised observers' behaviour.

While the ERP effects at ~125 ms and ~400 ms post-stimulus were constrained by classical temporal integration windows, the AT interactions for the P200 ERP component were most pronounced for ± 70 ms AT asynchrony and absent for near-synchronous AT stimulation (see Figures 5b-i and 3b). Both the auditory and the tactile unisensory P200 are thought to be generated in regions previously implicated in audiotactile integration (Foxe *et al.*, 2002; Kayser *et al.*, 2005; Murray *et al.*, 2005; Schurmann *et al.*, 2006) such as the auditory belt area CM or

planum temporale (Godey *et al.*, 2001; Crowley & Colrain, 2004; Smiley *et al.*, 2007) and secondary somatosensory areas (Forss *et al.*, 1994; Disbrow *et al.*, 2001), respectively. Our results show that AT integration facilitates neural processing at about 200 ms post-stimulus: the P200 peaks earlier and/or decays faster for the AT+N sum when compared to the sum of the unisensory A and T conditions (Figures 5d-i and 3b), consistent with multisensory literature, e.g. (Rowland *et al.*, 2007) and consistent with the quicker reaction times in the redundant target effect (AT versus A or T).

   The P200 ERP effects were directly related to AT interactions for theta-band ITC and TFP that emerged with a central topography again at ~200 ms post-stimulus primarily for ± 70 ms AT asynchrony (Figure 5). Critically, whilst the ERP and theta-band power interactions followed a similar temporal profile and topography irrespective of whether the auditory or the tactile stimulus is leading, the ITC effects were inverted for auditory relative to tactile leading stimulation (i.e., the condition weighting for P200 and theta power are in the same direction for AT70 and TA70, whereas the condition weightings for these asynchronies are opposing for theta ITC; Figure 5b). This dissociation between ERP and ITC is mathematically possible and one possible mechanism can be shown to produce this effect in simulation (see supplementary section 7 and comments in Results section 3.3). The selectivity of the P200 and the phase coherence effects for ± 70 ms AT asynchrony may be best accounted for by mechanisms of phase resetting that have previously been implicated in audiotactile and audiovisual interactions in auditory cortices (Lakatos *et al.*, 2007; Kayser *et al.*, 2008; Thorne *et al.*, 2011). From a functional perspective, a preceding tactile stimulus may reset the phase in auditory cortices and thereby facilitate the localization of an auditory stimulus that is presented 70 ms later. Likewise, a preceding auditory stimulus may provide an alert to facilitate tactile processing and possible

avoidance actions. Not only have tones been shown to elicit responses in somatosensory cortex (Borgest & Ermolaeva, 1975; Liang *et al.*, 2013), but also an *inhibitory* multisensory interaction by auditory stimulation was found in cat somatosensory area SIV (Dehner *et al.*, 2004) and auditory projections were found to inhibitory interneurons in cat SIV (Keniston *et al.*, 2010). In summary, our P200 ERP and theta-band ITC and power results are supported by evidence of bidirectional audiotactile integration, especially to association cortices, and of directional asymmetries in the AT interaction (Cecere *et al.*, 2017).

In summary, AT interactions were mediated by two distinct neural mechanisms depending on AT asynchrony: (i) ERP effects at ~100 and ~400 ms were most pronounced for (near) synchronous AT signals and followed a U-shape function across asynchronies, mimicking the temporal binding window at the behavioural level and (ii) ERP effects primarily driven by the ± 70 ms AT asynchronies were reflected in the P200, theta ITC, and theta power and may be mediated by mechanisms of phase resetting.

We also observed AT interactions for induced beta oscillatory power that were expressed across a set of asynchrony levels. As shown in Figure S7 (bottom row), both auditory and tactile stimuli suppressed beta oscillatory power (event-related desynchronization; ERD) at about 200-400 ms, related to a release from inhibition (Neuper & Pfurtscheller, 2001), followed by a rebound in power beyond baseline levels from about 600 ms – 1200 ms post-stimulus (event-related synchronisation; ERS), related to resetting and recovery (Pfurtscheller & Lopes da Silva, 1999; Neuper & Pfurtscheller, 2001). Beta-band AT interaction effects differed significantly across asynchronies only in the early (100-300 ms) time window (Figure S6). In the late (~1000 ms) window we observed AT interactions in the beta rebound for a few specific asynchronies, but no significant difference across asynchronies (Figure S7 and discussed further in the

supplementary discussion). The early effects were supported by several asynchronies and indicated a possible dependence on which sense came first, as the weighting across conditions (Figure S6b) flipped for A-leading versus T-leading. This is consistent with other studies that have shown sense-leading dependencies (e.g. Cecere et al., 2017) although novel here for beta-band effects. In contrast to the AT interactions in ERP, ITC and theta TF power, the AT interactions for beta power were not limited to the ± 70 ms temporal integration window but included the AT500 condition. The AT interactions for beta power may thus generally reflect non-specific mechanisms of multisensory priming or attention by which a preceding A (or T) signal may alert the observer to imminent touch (or sound) events, in light of the debate as to whether cross-modal stimuli with asynchronies up to 500-600 ms may be actually *integrated* or whether the first stimulus (only) primes and/or draws exogenous (spatial) cross-modal attention (Macaluso *et al.*, 2001; McDonald *et al.*, 2001; Stein *et al.*, 2010). Alternatively, the AT interactions for beta power may rely on several mechanisms depending on the AT asynchrony level, in which case the topography shown in Figure S6b and S6c may reflect the combination of these three (or more) mechanisms.

Caveats, limitations, and considerations:

Whilst this study did not directly compare the dual AT conditions against dual AA or TT conditions, it has been shown (Forster *et al.*, 2002) that the reaction times to visual-tactile dual stimuli were faster than for dual tactile or dual visual, indicating that the redundant target effect reflects special multisensory processing above and beyond that of integrating two stimuli (of the same sense). In the same way, regarding the EEG data, we did not explicitly contrast multisensory against dual unisensory conditions (e.g. AT+AT versus AA+TT) and thus are not accounting for general neural mechanisms of processing two stimuli versus one at a time.

However, we argue that by contrasting different multisensory asynchrony conditions against each other (i.e. rmANOVA applied to each the reaction time and the EEG data), we mitigate this interpretational ambiguity.

The tactile and auditory stimuli in this study were approximately aligned in time and space; however, other parameters were not matched, such as frequency (the tactile was a single on/off push and the sound's broadband spectrum is depicted in Figure S2). Frequency in particular (Butler et al., 2012; Yau et al., 2010) has been shown to be involved in preattentive audiotactile coupling. While these factors may have boosted or hindered integration in this study, they were not varied across the asynchrony conditions and so should not affect our main finding that AT interaction effects depend on asynchrony.

Studying multisensory integration in a passive paradigm at low vigilance prior to sleep onset is a novel approach associated with new insights and limitations. A passive paradigm enables us to study neural effects not confounded by explicit response selection processes. Indeed, in particular the confounding effects of decision making and response selection on neural interactions in fMRI have been widely recognized (e.g. (Werner & Noppeney, 2010a; Lee & Noppeney, 2011). For instance, Werner and Noppeney (2010a) have shown that multisensory integration effects in prefrontal cortices mainly reflect multisensory interactions at the decisional level that are only evoked when observers need to perform explicit categorization tasks. To study automatic multisensory interactions it is therefore important to use passive paradigms and relate neural effects to behavioural effects that are observed in associated psychophysics experiments (e.g. see also (Lee & Noppeney, 2011) and (Lee & Noppeney, 2014) for similar approach). Yet, studying behaviour and neural processing in separate experimental sessions is limited in the way that statistical brain-behaviour correspondences cannot be determined based on inter-trial

variability but only based on inter-condition or inter-subject variability. Our study thus focused on the former and demonstrates that even though the neural interaction effects occurred in a different cognitive context than the behavioural effects, some of them followed the same inverted U-shape profile across asynchrony levels. These results suggest that automatic neural effects that can be observed in the absence of explicit responses may potentially be responsible for the time window of integration at the behavioural level. Future studies recording EEG in an active redundant target paradigm are needed to directly link neural and behavioural effects based on inter-trial variability.

Moreover, as far as we know this is the first study that explored audiotactile interactions in human observers in a low vigilance state prior to sleep. It moves beyond previous research that has manipulated cognitive state in terms of various forms of attention (Talsma *et al.*, 2010) and may even provide new links with neurophysiological studies in anesthetized non-human primates (e.g. (Noel *et al.*, 2019)). For instance, cross-modal phase-resetting mechanisms have been reported both in anaesthetised (e.g. (Lakatos *et al.*, 2007)) and awake (e.g. (Romei *et al.*, 2012)) non-human ad human primates. Yet, low vigilance states come with their own challenges. First, we need to ensure that participants are indeed in the same low vigilance state rather than moving into sleep. We have ensured this with the help of sleep staging. Second, ideally we should have included an additional cognitive state during day time in a passive paradigm or even record EEG in an active paradigm. Considering unisensory processing as affected by vigilance states, it is well-known that early, bottom-up neural/ERP processes remain intact even when progressing from drowsiness to stages 1 and 2 of Non-REM sleep, while later neural responses are largely abolished. We thus can feel confident that at least the ~ 100 ms ERP effects observed here in awake drowsiness are not significantly altered compared to alertness. Third, we acknowledge

there could be greater variability in response times as participants drift to deeper drowsiness (e.g. (Jagannathan *et al.*, 2018)); however the presentation of trial types (asynchronies) were fully randomised in both behavioural and EEG setups and so should not cause a bias in either dataset. Fourth, it may be that different asynchrony conditions in our study (e.g. the more temporally congruent / synchronous ones) may draw attention and therefore be more alerting and induce a different vigilance state in different asynchrony conditions (e.g. review of attention and multisensory congruency in Talsma et al., 2010). Due to the different baseline vigilance states for the behavioural and EEG experiments, this congruency-vigilance interaction may play a distinct effect on the inverted U-shape, emphasising it more in the EEG data with a lower vigilance baseline. A future large scale study comparing AT interactions across different cognitive states should provide an even more comprehensive assessment of AT interactions across different cognitive states.

Conclusions:

To conclude, this psychophysics-EEG study demonstrates that AT integration is mediated by different neural mechanisms depending on AT asynchrony: for (near) synchronous AT signals, AT interactions were observed for early and late ERPs. For ±70 ms AT asynchrony interactions were expressed in middle latency ERP, theta ITC and theta power. Finally, across AT asynchrony levels even beyond the behavioural integration window we observed AT interactions for beta-band power that result from modulations of early and late (rebound) effects. This diversity of temporal profiles demonstrates that distinct neural mechanisms govern a cascade of multisensory integration processes depending on AT asynchrony.

Acknowledgements

Conflict of interest

None.

Author Contributions

U.N. and J.M.Z. formulated and planned the experiments. J.M.Z. carried out the experiments. J.M.Z., T.P.W., and U.N. carried out data analysis. J.M.Z. and U.N. prepared the manuscript. U.N. supervised the project.

Data availability statement

We apologise that we will not be able to upload the data in an open repository, because the ethics consent forms did not include this possibility.

Abbreviations

A = auditory; AT = audiotactile; EEG = electroencephalography; EMG = electromyogram; EOG = electrooculogram; ERP = evoked response potential; ITC = inter-trial coherence; MRI =

magnetic resonance imaging; N = null; PC = principal component; PCA = principal component analysis; rmANOVA = repeated measures analysis of variance; RT = reaction time; RTE = redundant target effect; T = tactile; TF = time-frequency; TIW = temporal integration window.

References:

Aller, M. & Noppeney, U. (2019) To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLoS Biol*, **17**, e3000210.

Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K.C. & Bizley, J.K. (2018) Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron*, **97**, 640-655 e644.

Berger, C.C. & Ehrsson, H.H. (2014) The fusion of mental imagery and sensation in the temporal association cortex. *J Neurosci*, **34**, 13684-13692.

Bizley, J.K., Nodal, F.R., Bajo, V.M., Nelken, I. & King, A.J. (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cereb Cortex*, **17**, 2172-2189.

Blurton, S.P., Greenlee, M.W. & Gondan, M. (2015) Cross-modal cueing in audiovisual spatial attention. *Atten Percept Psychophys*, **77**, 2356-2376.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.J. & Hillyard, S.A. (2007) Neural basis of the ventriloquist illusion. *Curr Biol*, **17**, 1697-1703.

Borgest, A.N. & Ermolaeva, V. (1975) [Functional organization of pathways transmitting auditory signals in the somatosensory zone of the cat cerebral cortex]. *Neirofiziologiia*, **7**, 476-485.

Brookes, M.J., Liddle, E.B., Hale, J.R., Woolrich, M.W., Luckhoo, H., Liddle, P.F. & Morris, P.G. (2012) Task induced modulation of neural oscillations in electrophysiological brain networks. *NeuroImage*, **63**, 1918-1930.

Cappe, C. & Barone, P. (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur J Neurosci*, **22**, 2886-2902.

Cappe, C., Morel, A., Barone, P. & Rouiller, E.M. (2009) The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cereb Cortex*, **19**, 2025-2037.

Cecere, R., Gross, J., Willis, A. & Thut, G. (2017) Being First Matters: Topographical Representational Similarity Analysis of ERP Signals Reveals Separate Networks for Audiovisual Temporal Binding Depending on the Leading Sense. *J Neurosci*, **37**, 5274-5287.

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.

Colonius, H. & Diederich, A. (2004) Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci*, **16**, 1000-1009.

Crowley, K.E. & Colrain, I.M. (2004) A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin Neurophysiol*, **115**, 732-744.

Dahl, C.D., Logothetis, N.K. & Kayser, C. (2009) Spatial organization of multisensory responses in temporal association cortex. *J Neurosci*, **29**, 11924-11932.

de la Mothe, L.A., Blumell, S., Kajikawa, Y. & Hackett, T.A. (2006a) Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J Comp Neurol*, **496**, 27-71.

de la Mothe, L.A., Blumell, S., Kajikawa, Y. & Hackett, T.A. (2006b) Thalamic connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J Comp Neurol*, **496**, 72-96.

Dehner, L.R., Keniston, L.P., Clemo, H.R. & Meredith, M.A. (2004) Cross-modal circuitry between auditory and somatosensory areas of the cat anterior ectosylvian sulcal cortex: a 'new' inhibitory form of multisensory convergence. *Cereb Cortex*, **14**, 387-403.

Diederich, A. & Colonius, H. (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept Psychophys*, **66**, 1388-1404.

Disbrow, E., Roberts, T., Poeppel, D. & Krubitzer, L. (2001) Evidence for interhemispheric processing of inputs from the hands in human S2 and PV. *J Neurophysiol*, **85**, 2236-2244.

Donohue, S.E., Green, J.J. & Woldorff, M.G. (2015) The effects of attention on the temporal integration of multisensory stimuli. *Front Integr Neurosci*, **9**, 32.

Downar, J., Crawley, A.P., Mikulis, D.J. & Davis, K.D. (2000) A multimodal cortical network for the detection of changes in the sensory environment. *Nat Neurosci*, **3**, 277-283.

Falchier, A., Clavagnier, S., Barone, P. & Kennedy, H. (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci*, **22**, 5749-5759.

Fetsch, C.R., Pouget, A., DeAngelis, G.C. & Angelaki, D.E. (2011) Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci*, **15**, 146-154.

Forss, N., Salmelin, R. & Hari, R. (1994) Comparison of somatosensory evoked fields to airpuff and electric stimuli. *Electroencephalogr Clin Neurophysiol*, **92**, 510-517.

Forster, B., Cavina-Pratesi, C., Aglioti, S.M. & Berlucchi, G. (2002) Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp Brain Res*, **143**, 480-487.

Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C. & Schroeder, C.E. (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res Cogn Brain Res*, **10**, 77-83.

Foxe, J.J. & Schroeder, C.E. (2005) The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, **16**, 419-423.

Foxe, J.J., Wylie, G.R., Martinez, A., Schroeder, C.E., Javitt, D.C., Guilfoyle, D., Ritter, W. & Murray, M.M. (2002) Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J Neurophysiol*, **88**, 540-543.

Fu, K.M., Johnston, T.A., Shah, A.S., Arnold, L., Smiley, J., Hackett, T.A., Garraghty, P.E. & Schroeder, C.E. (2003) Auditory cortical neurons respond to somatosensory stimulation. *J Neurosci*, **23**, 7510-7515.

Ghazanfar, A.A. & Schroeder, C.E. (2006) Is neocortex essentially multisensory? *Trends Cogn Sci*, **10**, 278-285.

Giard, M.H. & Peronnet, F. (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci*, **11**, 473-490.

Godey, B., Schwartz, D., de Graaf, J.B., Chauvel, P. & Liegeois-Chauvel, C. (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clin Neurophysiol*, **112**, 1850-1859.

Gondan, M. & Minakata, K. (2016) A tutorial on testing the race model inequality. *Atten Percept Psychophys*, **78**, 723-735.

Hackett, T.A., De La Mothe, L.A., Ulbert, I., Karmos, G., Smiley, J. & Schroeder, C.E. (2007) Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J Comp Neurol*, **502**, 924-952.

Harrar, V. & Harris, L.R. (2008) The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Exp Brain Res*, **186**, 517-524.

Hasson, U., Yang, E., Vallines, I., Heeger, D.J. & Rubin, N. (2008) A hierarchy of temporal receptive windows in human cortex. *J Neurosci*, **28**, 2539-2550.

Hershenson, M. (1962) Reaction time as a measure of intersensory facilitation. *J Exp Psychol*, **63**, 289-293.

Hoefer, M., Tyll, S., Kanowski, M., Brosch, M., Schoenfeld, M.A., Heinze, H.J. & Noesselt, T. (2013) Tactile stimulation and hemispheric asymmetries modulate auditory perception and neural responses in primary auditory cortex. *NeuroImage*, **79**, 371-382.

Ibrahim, L.A., Mesik, L., Ji, X.Y., Fang, Q., Li, H.F., Li, Y.T., Zingg, B., Zhang, L.I. & Tao, H.W. (2016) Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition. *Neuron*, **89**, 1031-1045.

Jagannathan, S.R., Ezquerro-Nassar, A., Jachs, B., Pustovaya, O.V., Bareham, C.A. & Bekinschtein, T.A. (2018) Tracking wakefulness as it fades: Micro-measures of alertness. *NeuroImage*, **176**, 138-151.

Kayser, C., Petkov, C.I., Augath, M. & Logothetis, N.K. (2005) Integration of touch and sound in auditory cortex. *Neuron*, **48**, 373-384.

Kayser, C., Petkov, C.I., Augath, M. & Logothetis, N.K. (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci*, **27**, 1824-1835.

Kayser, C., Petkov, C.I. & Logothetis, N.K. (2008) Visual modulation of neurons in auditory cortex. *Cereb Cortex*, **18**, 1560-1574.

Keniston, L.P., Henderson, S.C. & Meredith, M.A. (2010) Neuroanatomical identification of crossmodal auditory inputs to interneurons in somatosensory cortex. *Exp Brain Res*, **202**, 725-731.

Kiebel, S.J., Daunizeau, J. & Friston, K.J. (2008) A hierarchy of time-scales and the brain. *PLoS Comput Biol*, **4**, e1000209.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A. & Schroeder, C.E. (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, **53**, 279-292.

Leclercq, Y., Schrouff, J., Noirhomme, Q., Maquet, P. & Phillips, C. (2011) fMRI artefact rejection and sleep scoring toolbox. *Comput Intell Neurosci*, **2011**, 598206.

Lee, H. & Noppeney, U. (2011) Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc Natl Acad Sci U S A*, **108**, E1441-1450.

Lee, H. & Noppeney, U. (2014) Temporal prediction errors in visual and auditory cortices. *Curr Biol*, **24**, R309-310.

Leonardelli, E., Braun, C., Weisz, N., Lithari, C., Occelli, V. & Zampini, M. (2015) Prestimulus oscillatory alpha power and connectivity patterns predispose perceptual integration of an audio and a tactile stimulus. *Hum Brain Mapp*, **36**, 3486-3498.

Liang, M., Mouraux, A., Hu, L. & Iannetti, G.D. (2013) Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nat Commun*, **4**, 1979.

Littner, M.R., Kushida, C., Wise, M., Davila, D.G., Morgenthaler, T., Lee-Chiong, T., Hirshkowitz, M., Daniel, L.L., Bailey, D., Berry, R.B., Kapen, S., Kramer, M. & Standards of Practice Committee of the American Academy of Sleep, M. (2005) Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep*, **28**, 113-121.

Lutkenhoner, B., Lammertmann, C., Simoes, C. & Hari, R. (2002) Magnetoencephalographic correlates of audiotactile interaction. *NeuroImage*, **15**, 509-522.

Macaluso, E., Frith, C. & Driver, J. (2001) (Response to) Multisensory Integration and Crossmodal Attention Effects in the Human Brain. *Science*, **292**, 1791-1791.

Maris, E. & Oostenveld, R. (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*, **164**, 177-190.

McDonald, J.J., Teder-Sälejärvi, W.A. & Ward, L.M. (2001) Multisensory Integration and Crossmodal Attention Effects in the Human Brain. *Science*, **292**, 1791-1791.

Megevand, P., Molholm, S., Nayak, A. & Foxe, J.J. (2013) Recalibration of the multisensory temporal window of integration results from changing task demands. *PLoS One*, **8**, e71608.

Mercier, M.R., Foxe, J.J., Fiebelkorn, I.C., Butler, J.S., Schwartz, T.H. & Molholm, S. (2013) Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. *NeuroImage*, **79**, 19-29.

Mercier, M.R., Molholm, S., Fiebelkorn, I.C., Butler, J.S., Schwartz, T.H. & Foxe, J.J. (2015) Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation. *J Neurosci*, **35**, 8546-8557.

Meredith, M.A. & Stein, B.E. (1983) Interactions among converging sensory inputs in the superior colliculus. *Science*, **221**, 389-391.

Miller, J. (1982) Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol*, **14**, 247-279.

Mishra, J., Martinez, A., Sejnowski, T.J. & Hillyard, S.A. (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J Neurosci*, **27**, 4120-4131.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E. & Foxe, J.J. (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res*, **14**, 115-128.

Murray, M.M., Molholm, S., Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., Schroeder, C.E. & Foxe, J.J. (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex*, **15**, 963-974.

Naue, N., Rach, S., Struber, D., Huster, R.J., Zaehle, T., Korner, U. & Herrmann, C.S. (2011) Auditory event-related response in visual cortex modulates subsequent visual responses in humans. *J Neurosci*, **31**, 7729-7736.

Navarra, J., Soto-Faraco, S. & Spence, C. (2007) Adaptation to audiotactile asynchrony. *Neurosci Lett*, **413**, 72-76.

Neuper, C. & Pfurtscheller, G. (2001) Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates. *Int J Psychophysiol*, **43**, 41-58.

Nishi, A., Yokoyama, M., Ogawa, K., Ogata, T., Nozawa, T. & Miyake, Y. (2014) Effects of Voluntary Movements on Audio-Tactile Temporal Order Judgment. *Ieice T Inf Syst*, **E97d**, 1567-1573.

Noel, J.P., Ishizawa, Y., Patel, S.R., Eskandar, E.N. & Wallace, M.T. (2019) Leveraging Nonhuman Primate Multisensory Neurons and Circuits in Assessing Consciousness Theory. *J Neurosci*, **39**, 7485-7500.

Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.M. (2011) FieldTrip: Open source software for
advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*,
2011, 156869.

Pfurtscheller, G. & Lopes da Silva, F.H. (1999) Event-related EEG/MEG synchronization and
desynchronization: basic principles. *Clin Neurophysiol*, **110**, 1842-1857.

Rohe, T., Ehlis, A.C. & Noppeney, U. (2019) The neural dynamics of hierarchical Bayesian causal
inference in multisensory perception. *Nat Commun*, **10**, 1907.

Rohe, T. & Noppeney, U. (2015) Cortical hierarchies perform Bayesian causal inference in multisensory
perception. *PLoS Biol*, **13**, e1002073.

Rohe, T. & Noppeney, U. (2016) Distinct Computational Principles Govern Multisensory Integration in
Primary Sensory and Association Cortices. *Curr Biol*, **26**, 509-514.

Romei, V., Gross, J. & Thut, G. (2012) Sounds reset rhythms of visual cortex and corresponding human
visual perception. *Curr Biol*, **22**, 807-813.

Rowland, B.A., Quessy, S., Stanford, T.R. & Stein, B.E. (2007) Multisensory integration shortens
physiological response latencies. *J Neurosci*, **27**, 5879-5884.

Rowland, B.A. & Stein, B.E. (2007) Multisensory integration produces an initial response enhancement. *Front Integr Neurosci*, **1**, 4.

Schroeder, C.E. & Foxe, J.J. (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res*, **14**, 187-198.

Schurmann, M., Caetano, G., Hlushchuk, Y., Jousmaki, V. & Hari, R. (2006) Touch activates human auditory cortex. *NeuroImage*, **30**, 1325-1331.

Senkowski, D., Gomez-Ramirez, M., Lakatos, P., Wylie, G.R., Molholm, S., Schroeder, C.E. & Foxe, J.J. (2007) Multisensory processing and oscillatory activity: analyzing non-linear electrophysiological measures in humans and simians. *Exp Brain Res*, **177**, 184-195.

Senkowski, D., Schneider, T.R., Foxe, J.J. & Engel, A.K. (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci*, **31**, 401-409.

Smiley, J.F., Hackett, T.A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D.C. & Schroeder, C.E. (2007) Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol*, **502**, 894-923.

Sperdin, H.F., Cappe, C., Foxe, J.J. & Murray, M.M. (2009) Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front Integr Neurosci*, **3**, 2.

Stanford, T.R., Quessy, S. & Stein, B.E. (2005) Evaluating the operations underlying multisensory

integration in the cat superior colliculus. *J Neurosci*, **25**, 6499-6508.

Stein, B.E., Burr, D., Constantinidis, C., Laurienti, P.J., Alex Meredith, M., Perrault, T.J., Jr.,

Ramachandran, R., Roder, B., Rowland, B.A., Sathian, K., Schroeder, C.E., Shams, L., Stanford,

T.R., Wallace, M.T., Yu, L. & Lewkowicz, D.J. (2010) Semantic confusion regarding the

development of multisensory integration: a practical solution. *Eur J Neurosci*, **31**, 1713-1720.

Stekelenburg, J.J. & Vroomen, J. (2009) Neural correlates of audiovisual motion capture. *Exp Brain Res*,

198, 383-390.

Talsma, D., Senkowski, D., Soto-Faraco, S. & Woldorff, M.G. (2010) The multifaceted interplay between

attention and multisensory integration. *Trends Cogn Sci*, **14**, 400-410.

Talsma, D. & Woldorff, M.G. (2005) Selective attention and multisensory integration: multiple phases of

effects on the evoked brain activity. *J Cogn Neurosci*, **17**, 1098-1114.

Tanaka, H., Hayashi, M. & Hori, T. (1996) Statistical features of hypnagogic EEG measured by a new

scoring system. *Sleep*, **19**, 731-738.

Teder-Salejarvi, W.A., McDonald, J.J., Di Russo, F. & Hillyard, S.A. (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Res Cogn Brain Res*, **14**, 106-114.

Thorne, J.D., De Vos, M., Viola, F.C. & Debener, S. (2011) Cross-modal phase reset predicts auditory task performance in humans. *J Neurosci*, **31**, 3853-3861.

Ulrich, R., Miller, J. & Schroter, H. (2007) Testing the race model inequality: an algorithm and computer programs. *Behav Res Methods*, **39**, 291-302.

van Wassenhove, V., Grant, K.W. & Poeppel, D. (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, **45**, 598-607.

Werner, S. & Noppeney, U. (2010a) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J Neurosci*, **30**, 2662-2675.

Werner, S. & Noppeney, U. (2010b) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb Cortex*, **20**, 1829-1842.

Werner, S. & Noppeney, U. (2011) The contributions of transient and sustained response codes to audiovisual integration. *Cereb Cortex*, **21**, 920-931.

Zampini, M., Brown, T., Shore, D.I., Maravita, A., Roder, B. & Spence, C. (2005) Audiotactile temporal

order judgments. *Acta Psychol (Amst)*, **118**, 277-291.

Table 1: Race model statistics

Decile			AT500	AT70	AT20	AT0	TA20	TA70	TA500
10%	p-value		0.917	0.002	**<0.001**	**<0.001**	**0.001**	0.013	0.130
	t-value		1.437	-3.204	-4.693	-4.303	-3.679	-2.404	-1.183
20%	p-value		0.963	**0.001**	**<0.001**	**<0.001**	**<0.001**	0.005	0.110
	t-value		1.880	-3.462	-5.619	-4.089	-4.461	-2.866	-1.251
30%	p-value		0.988	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.001**	0.042
	t-value		2.417	-4.848	-6.275	-4.561	-3.920	-3.816	-1.819
40%	p-value		0.979	**<0.001**	**<0.001**	0.010	**0.001**	0.003	0.028
	t-value		2.159	-5.351	-4.789	-2.526	-3.775	-3.133	-2.020
50%	p-value		0.994	**<0.001**	**<0.001**	0.053	0.003	0.006	0.060
	t-value		2.780	-5.501	-4.009	-1.689	-3.031	-2.790	-1.625

Table 1: Statistical results from comparing the race model inequality for each decile of reaction time distribution over the seven asynchronies. P-values in bold are significant after Bonferroni correction.

Figure captions

Figure 1. Experimental design and behavioural results. (a) Each row depicts the onsets of the auditory stimulation (indicated by loudspeaker) and tactile stimulation (indicated by face) for each of the 10 conditions including the null (N), auditory alone (A), tactile alone (T) and the seven AT conditions with asynchrony: 0, ±20, ±70, and ±500 ms. The wavy line at the bottom indicates the continuous MRI background noise. (b) Median reaction times versus asynchrony. Each black dot represents an individual participant's median reaction time; the white cross is the mean over participants. Negative asynchronies indicate auditory-leading; the leftmost and rightmost datasets indicate the A and T conditions, respectively.

Figure 2. Overview of analyses stream. (a) Statistics: within each asynchrony, the response time course (ERP, TF power, or ITC) for A alone (appropriately shifted in time to match a particular asynchrony) and T alone are summed, as are the time courses for AT (for a given asynchrony) and N. The difference of these, i.e. the audio-tactile "interaction" is computed for each asynchrony. (a.1) These AT interactions computed for each asynchrony are compared across asynchronies in a repeated-measures (dependent-samples / paired) ANOVA (see Methods for details). (a.2) Furthermore, the AT interaction within each asynchrony is tested with a paired (dependent samples) t-test. (b) To assess the relative contributions of conditions to any significant modulatory effects of asynchrony on AT interaction (i.e. three way interaction) in the rmANOVA as well as to relate the within-asynchrony assessment of AT interaction to the across-asynchrony assessment, two PCAs were sequentially performed. All AT interaction values from channel-time points in the significant cluster found in the rmANOVA were reshaped into a vector, one for each asynchrony, and then concatenated over asynchrony. This 7 X masked-channel-time matrix was entered in to PCA1, from which the first component was extracted. The first component's weighting indicated the contribution of each asynchrony level to the effects revealed in the rmANOVA. This weighting across asynchrony levels (i.e. 1 X 7 asynchrony levels) was also multiplied with the original (non-masked) 7 X channel-time matrix to obtain the pattern across channel-time (1 x channel-time vector). This vector was reshaped back to a matrix of channels X time which reflects the differences across the 7 asynchrony levels in AT interactions over time and channels that drove the significant effects in the initial rmANOVA. Using PCA2 we decomposed this channels X time matrix into the topographies and their time courses. We plotted the topography and time course of the first (i.e. dominant)

principal component alongside the within-asynchrony AT integration effect for the asynchrony level which had the strongest condition weighting from PCA1.

Figure 3: Evoked response potentials. (a) Evoked response potentials for N, A, T, and AT conditions for the following sets of sensors: frontocentral ['Fz' 'Cz' 'F1' 'F2' 'FC1' 'FC2' 'C1' 'C2'] (the ring of electrodes centred on FCz, where the P200 effect is strongest) and posterior ['CP5' 'POz' 'Pz' 'P3' 'P4' 'C4' 'O1' 'O2' 'P7' 'PO7'] (where the later 400 ms effect in ±20 ms asynchrony is strongest). The A evoked response is shifted by the appropriate asynchrony to align with the auditory onset in the corresponding AT condition. (b) Each row shows the audiotactile interaction for a particular level of AT asynchrony. *Left columns:* ERPs of the sum of the auditory and tactile (A+T), the sum of audiotactile plus null (AT+N), and the audiotactile interaction, i.e. the difference ([A+T]-[AT+N]). Shaded grey areas indicate the timing of significant AT interactions at p < 0.05 corrected at the cluster level for multiple comparisons across electrodes and time points within a 500 ms window starting with the second stimulus and limited by the black dashed line. *Right column:* Topographies of the sums: A+T, AT+N, and (A+T)-(AT+N) for time windows of significant AT interactions. The time windows are relative to the onset of the second stimulus. A black star over an electrode indicates that it is part of a significant cluster.
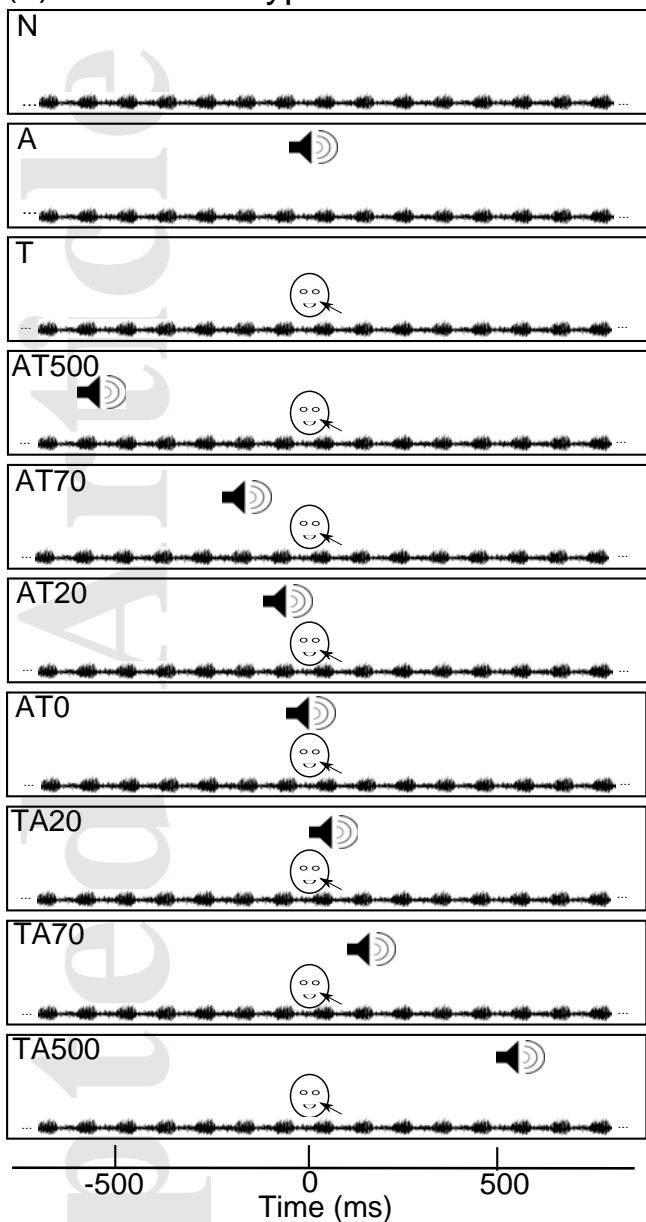
Figure 4. Early and late ERP effects. (a) Channel X time matrix. Channels are arranged from frontal to occipital (top to bottom). Highlighted are channel-time points that are part of the clusters of the modulatory effect of asynchrony on AT interaction (i.e. three way interaction). For p-values, see Table S1. The second cluster is divided into three sub-clusters based on three

time windows (analysed in separate PCAs). This figure focuses on the time windows (i) 30-150 ms and (ii) 350-420 ms. (See Figure 5 for time window 150-300 ms and Figure S3 for time window 450-500 ms.) (b) (left) Topography of rmANOVA F values averaged across time windows (i) 50-135 ms and (ii) 360-410 ms (n.b. small deviations from the time windows in A were allowed to avoid time points with only few significant channels). (right) The weights (obtained from PCA1) indicate which asynchrony levels drove the three way interaction, i.e. the differences in AT interactions across asynchronies, in the rmANOVA. A strong weight is obtained for (i) AT0 in the early time window and (ii) AT20 in the later time window. (c) The topographies and time courses of the component (obtained from PCA2) that contribute most strongly to the across-asynchrony differences in AT interactions. The topographies in the early and later time windows are similar but with opposite polarities and evolve with similar time courses. (d) The topography and time course of AT interaction contrast for the asynchrony level with the greatest absolute weight in PCA1 (sub-figure (b)) is shown for (i) AT0 in the early time window and (ii) AT20 in the late time window. The larger black dots indicate the sensors included in the significant AT interaction cluster (i.e. A+T versus AT+N) for a particular asynchrony; the grey shaded area indicates the temporal extent of this significant cluster. For ERP interaction effects for each asynchrony level, see Figure 3b. Please note that, whereas (b) shows F-values (i.e. statistical values), (c) and (d) show contrast values (i.e. differences in voltages across conditions).
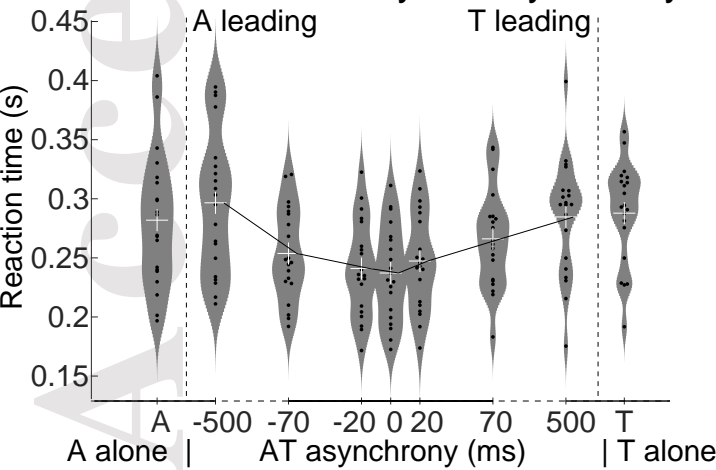
Figure 5. Same layout as for Figure 4, except that the columns are from the following data: (i) P200 ERP effect, (ii) theta-band ITC effect, and (iii) theta-band power effect. Note the x-axis differences between the sub-plots (0-500 ms for ERP and 0-1.2 s for theta ITC and power). (a)

Channel X time matrix. Highlighted are channel-time points that are part of the clusters of the modulatory effect of asynchrony on AT interaction (i.e. three way interaction). For p-values, see Table S1. (b) (left) Topography of rmANOVA F values averaged across time windows (i) 180-270 ms, (ii) 0-440 ms, and (iii) 100-370 ms. (n.b. small deviations from the time windows in (a) were allowed to avoid time points with only few significant channels). (right) The weights (obtained from PCA1) indicate which asynchrony levels drove the three way interaction, i.e. the differences in AT interactions across asynchronies, in the rmANOVA. A strong weight is obtained for (i) AT70 for ERP (ii) TA70 for theta ITC, and (iii) TA70 for theta power. Note that the second largest weight for ITC and power was for the AT70 asynchrony, same as the ERP. (c) The topographies and time courses of the component (obtained from PCA2) that contribute most strongly to the across-asynchrony differences in AT interactions. Note that all three have temporal peaks around 250 ms. (d) The topography and time course of AT interaction contrast for the asynchrony level with the greatest absolute weight in PCA1 (sub-figure (b)) is shown for (i) AT70 in the P200 ERP, (ii) TA70 for theta ITC, and (iii) TA70 for theta power. The larger black dots indicate the sensors included in the significant AT interaction cluster (i.e. A+T versus AT+N) for a particular asynchrony; the grey shaded area indicates the temporal extent of this significant cluster. Note the similarity in both topographies and in time courses of the AT interaction. For interaction effects for each asynchrony level, see Figures 3b, S4, and S5 for ERP, ITC, and power, respectively. Please note that, whereas (b) shows F-values (i.e. statistical values), (c) and (d) show contrast values (i.e. differences across conditions).

(a) Trial type schematic

N

A

T

AT500

AT70

AT20

AT0

TA20

TA70

TA500

-500   0   500
Time (ms)

(b) Reaction times by AT asynchrony

A leading   T leading

Reaction time (s)

0.45
0.4
0.35
0.3
0.25
0.2
0.15

A | -500 -70 -20 0 20 70 500 | T

A alone | AT asynchrony (ms) | T alone

(a) Statistics

Response time course

Interaction (A+T)-(AT+N)

A.1 Repeated-measures ANOVA across asynch.

7 AT asynchronies

A=auditory stimulus
T=tactile stimulus
N=null / no stimulus
C=Channels
PC=principal component

A.2 t-test within asynch.

F-map mask

Mask the data
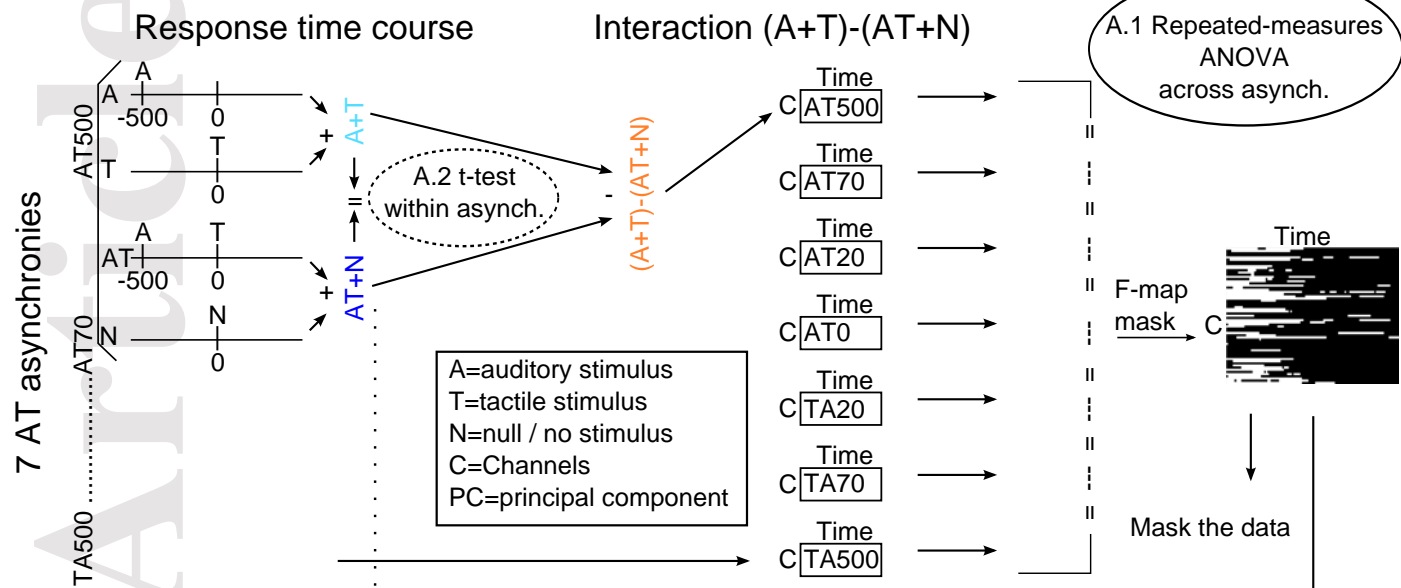
(b) Characterisation with principal component analysis (PCA)

Using only channel-time points within the mask

7 AT asynchronies

Mask C*Time

concatenate

Mask C*Time

PCA1

PC

Mask C*Time

1st PC

F-values are driven by these distinctions across asynchronies

F map

A.2 t-test result: TA70

Select asynchrony with strongest weighting

A+T
AT+N
(A+T)-(AT+N)
T onset
A onset

Weightings applied to original Channel*Time data over 7 asynchronies and reshaped
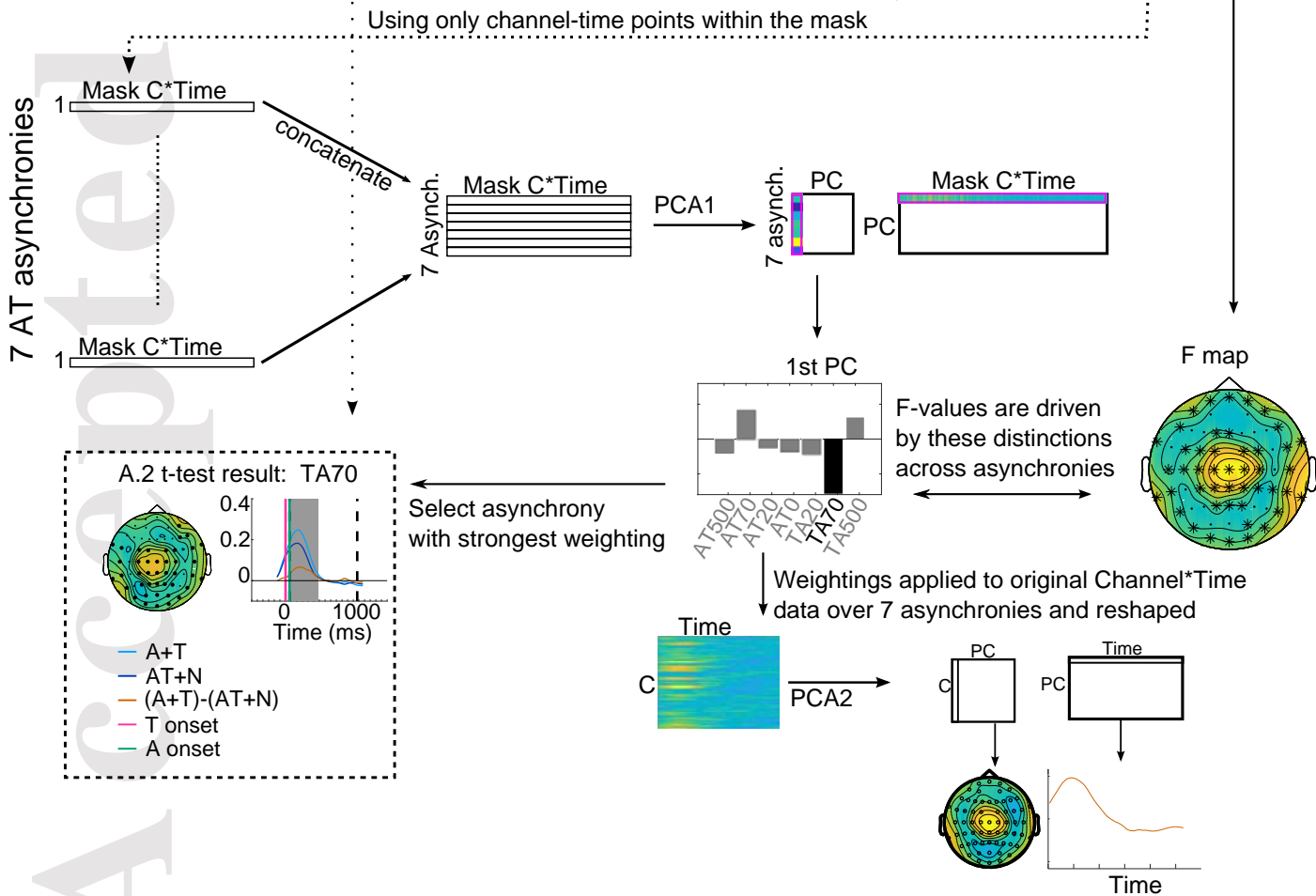
PCA2

Evoked response potentials (ERP), as function of AT asynchrony

(a) ERP per condition

(b) ERP audiotactile interactions

Figure legend:
- N (gray)
- T (pink)
- A (green)
- AT (black)
- A+T (light blue)
- AT+N (dark blue)
- (A+T)−(AT+N) (orange)
- T onset (pink)
- A onset (green)

Row labels (left): AT500, AT70, AT20, AT0, TA20, TA70, TA500

Columns: Frontocentral, Posterior (for panel a); Frontocentral, Posterior, Topography (for panel b)

Topography time windows: 120–270 ms, 370–420 ms, 70–170 ms, 340–400 ms, 140–220 ms

Topography labels: A + T, AT + N, (A+T)−(AT+N)

Difference of mV scale: −3 to 3

x-axis: Post-tactile stimulus time (s), ranging −0.5 to 1.0

y-axis (panel a): uV, −2 to 6
y-axis (panel b): −5 to 5

(a) Mask of ANOVA

ERP

(i) 30-150 ms

(ii) 350-420 ms

Channels: F, C, O

Time (ms), 0 to 500

(b) F map and asynchrony weighting of PCA1

(i) p=0.065, 50-135 ms, F value 0 to 2.5
scaled mV: AT500, AT70, AT20, AT0, TA20, TA70, TA500

(ii) 360-410 ms, F value 0 to 5
scaled mV: AT500, AT70, AT20, AT0, TA20, TA70, TA500

(c) PCA2: Spatio-temporal profile of first component

(i) scaled mV -0.3 to 0.3; scaled mV -0.1 to 0.1, Time (ms) 0 to 500

(ii) scaled mV -0.25 to 0.25; scaled mV -0.1 to 0.1, Time (ms) 0 to 500

(d) AT interaction for asynchrony with highest weight in PCA1

(i) AT0, mV -3 to 3, 70-170 ms

(ii) AT20, mV -3 to 3, 370-420 ms

A+T
AT+N
(A+T)-(AT+N)
T onset
A onset

(a) Mask of ANOVA

(i) ERP — 150-300 ms

Channels: F, C, O — Time (ms) 0 to 500

(ii) theta ITC — all points in mask — Time (ms) 0 to 1000

(iii) theta TFP — all points in mask — Time (ms) 0 to 1000

(b) F map and asynchrony weighting of PCA1

(i) F value 0–6 — 180-270 ms — scaled mV: AT500, AT70, AT20, AT0, TA20, TA70, TA500

(ii) F value 0–9 — 0-440 ms — scaled coh.: AT500, AT70, AT20, AT0, TA20, TA70, TA500

(iii) F value 0–5 — 100-370 ms — scaled mV^2: AT500, AT70, AT20, AT0, TA20, TA70, TA500

(c) PCA2: Spatio-temporal profile of first component

(i) scaled mV −0.2 to 0.2 — scaled mV, Time (ms) 0 to 500

(ii) scaled coh. −0.25 to 0.25 — scaled coh., Time (ms) 0 to 1000

(iii) scaled mV^2 −0.2 to 0.2 — scaled mV^2, Time (ms) 0 to 1000

(d) AT interaction for asynchrony with highest weight in PCA1

(i) mV −3 to 3 — 120-270 ms — AT70, Time (ms) 0 to 500

(ii) Diff. of Coh. −0.15 to 0.15 — 0-510 ms — TA70, Time (ms) 0 to 1000

(iii) mV^2 −1 to 1 — 0-470 ms — TA70, Time (ms) 0 to 1000

Legend: A+T, AT+N, (A+T)-(AT+N), T onset, A onset