Journal of Experimental Psychology: General

What You See Depends on What You Hear: Temporal Averaging and Crossmodal Integration Lihan Chen, Xiaolin Zhou, Hermann J Müller, and Zhuanghua Shi

Online First Publication, September 13, 2018. http://dx.doi.org/10.1037/xge0000487

 \mathbf{Y}_1 \mathbf{Y}_2 \mathbf{D}_2 is the set of \mathbf{Y}_1 and \mathbf{A}_3 and \mathbf{A}_4 and \mathbf{D}_4 Crossmodal Integration

Lihan Chen and X_{\pm} and \mathcal{A}_{\pm} . $\mathbb{P}_{\mathbf{u}}$ University University

Hermann J. Müller Ludwig Maximilian University of Munich and Birkbeck and Birkbeck and Birkbeck and Birkbeck College, University of London

 \mathcal{A} and \mathcal{A} Ludwig Maximilian University of Munich

In our multisensory N our more on a visual information than on auditory in \mathcal{N} , we one visual input for temporal processing. One typical demonstration of this is that the rate of auditory flutter assimilates the rate of concerns of concerns visual flicker. To date, however, this auditory dominance effect has largely \sim و المعروضي العموم من ²⁰⁰رمي الرياضي العربي من العربي بهروس منهوري مي العراقيون الرواحي والمعروفي would have a similar impact on visual temporal processing, where information is extracted from the ω auditory sequence that comes to influence visual timing, and $\mathcal{W}_{\mathcal{M}}$ and visual temporal rates are integrated together in quantitative terms. We investigated these questions by assessing, and modeling, the influence of a task-irrelevant auditory sequence on the type of "Ternus apparent motion": group motion versus element motion. The type of motion seen critically depends on the time interval between the two Ternus display frames. We found that an irrelevant auditory sequence preceding the Ternus display modulates the visual interval, making observers perceive either more group motion or more element motion. This biasing effect manifests whether the auditory sequence is regular or irregular, and it is based on a summary statistic extracted from the sequential intervals: the audiovisual intervals: the audiovisual interaction depends on \mathcal{S} the discrepancy between the mean auditory and visual intervals: if it is intervals: if it becomes too large, no interaction occurs—which can be quantitatively described by a partial Bayesian integration model. σ verall, our findings reveal and perceptual perceptual perceptual averaging principle that may underlie complex σ audiovisual interactions in many everyday dynamic situations.

 $Keywords: p = p = p \perp p \perp p$, and a perceptual and motion, multisensory interaction, $p = p \perp p$ Bayesian integration

Most stimuli and events in our everyday environments are multisensory. It is thus no surprise that our brain of the our brain of the outer σ المداحي المعلومات والحرينات المعديات بال فعلال المعالم المناسبة conflict. One typical such phenomenon, in a performance we have a performance we have a performance we have a
The performance we have a performance we have a performance we have a performance we have a performance we have entriloquism effect (Alexandrich, 2013; Occelli, 2013; Occelli, Occelli, Occelli, Occelli, Occelli, Occelli, O \mathcal{L} , \mathcal{L} , \mathcal{L} , \mathcal{R} \mathcal{L} , \mathcal{L} , 2012; \mathcal{L} \mathcal{L} , \mathcal{L} , 200₉; Slutsky \mathcal{R} $R_{\rm c}$, ~ 2001 , $^{\rm W}$ perceive the ventrical ventriloguistic as coming as coming from the model is a dummy as if it was the dummy that it was the dummy that it was the dummy that is defined as speaking. Of note in the present context, and integration integrations α has not only been demonstrated in spatial localization, but also in the temporal domain. In contrast to the dominance of vision in audiovisual spatial perception, audition dominates temporal processing, such as in reduced as intervals. As an example, think of the intervals. As an example, think of think of \mathbf{A} how \mathbb{N} , we tend to \mathbb{N} and \mathbb{N} and \mathbb{N} are movements coordinates coordinates coordinates coordinates coordinates coordinates coordinates coordinates \mathbb{N} dinating a musical passage, or Morse code flashes emanating from a naval ship. In fact, neuroscience evidence evidence evidence evidence evidence ϵ

Lihan Chen, Center for Brain and Cognitive Sciences and School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, and Key Laboratory of Machine Perception (Ministry of $E_\mathrm{eff} = \frac{1}{2} \left(\frac{1}{2} \frac{1}{\omega_\mathrm{eff}} \right)$ سند ال $E_\mathrm{eff} = \frac{1}{2} \left(\frac{1}{2} \frac{1}{\omega_\mathrm{eff}} \right)$ and $E_\mathrm{eff} = \frac{1}{2} \left(\frac{1}{2} \frac{1}{\omega_\mathrm{eff}} \right)$ Sciences and School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, Key Laboratory of Machine Perception (Ministry of Education), and P \mathbb{R} \mathbb{R} \mathbb{R} institute for \mathbb{R} Brain Research, Peking University; Hermann J. Müller, Department Psycholo $g_{\mu\nu}$ ی سازم دیگر سازی \blacksquare در ایران ایران ایران ایران ایران میکنس می سازی \blacksquare logical Sciences, Birkbeck College, University of London; Zhuanghua Shi, \mathbf{D}_max and \mathbf{v}_max of \mathbf{L}_max and \mathbf{V}_max \mathbf{L}_max and \mathbf{L}_max of \mathbf{L}_max

 \mathcal{A} and \mathcal{A} and \mathcal{A} and \mathcal{A} and \mathcal{A} and \mathcal{A} and \mathcal{A}

 \pm . The study has been presented as a talk at the 17th International at the 17th International international \pm Multipensory Research Forum (IMF), \sim 2016, Supply 2017, Supply study was supported by grants from the Natural Science Foundation of $C_{\rm tot}$ ($C_{\rm tot}$ 31200760, 12113600, and 61627600), Deutsche Forschult SH1666 3/1 and π SH166 3/1 and π schultz SH166 3/1 and π senschaftleraustaustauschen (proWA). The data and the source code of source code of Λ statistical analysis and modeling are available at [https://github.com/](https://github.com/msenselab/temporal_averaging) [msenselab/temporal_averaging.](https://github.com/msenselab/temporal_averaging)

Correspondence concerning this article should be addressed to Lihan Chen, School of Psychological and nitiveAs442.cj11c7F11information for time estimation is encoded in the primary auditory σ cortex for both visual and auditory events (Kanai, Lloyd) $\mathcal{L}_{\mathcal{L}}$, 2011). This is consistent with the percep-that the perceptual system automatically abstracts temporal structure from rhythmic visual sequences and represents this structure using an auditory code (Guttman, Gilroy, & Blake, 2005).

 \mathbf{A}_c , and competitive demonstration of \mathcal{N}_c and \mathcal{N}_c rhythm influences visual tempo is $\mu_{\text{max}}^{\text{w}}$ and *auditory driving effect* (Boltz, 2017; Gebhard & Mowbray, 1959; Knox, 1945; Shipley, 1 964): the phenomenon that variations in auditory flutter rate may be may be notice the rate of perceived visual flickers in \mathcal{E}_max in the rate of perceived visual flicker. This is influence, though, is dependent on the disparity between the au-disparity between the au-disparity between the ditory and visual rates [\(Recanzone, 2003\)](#page-13-1). Quantitatively, this section of the section of the section of the s
And the section of t influence has been described by a Bayesian model of audiovisual integration (Roach, Heron, $\&$ assumes that \mathcal{L}^{W} , 2006), \mathcal{L}^{W} , \mathcal{L}^{W} the brain takes into account prior \mathcal{P}^W the discrepancy the discrepancy of between the auditory and visual rates in determining the degree of $\mathcal{N}_{\mathbf{z}}$. auditoris und integration. Auditory driving is a robust effect that A generalizes across different types of tasks, including temporal adjustment and production (Myers, Cotton, & Hilp, 1981) and $p \cdot \sigma$ and discrimination (Welch, Dution A, $\&$ \Rightarrow \rightarrow 1 \rightarrow 1 \rightarrow), and it may even be seen in the effect of one single auditory in the effect of one single auditory intervals. ω on a subsequent visual interval interval interval $(\mathbb{R}^n, \mathbb{R}^n)$ and \mathbb{R}^n are \mathbb{R}^n 201 .

It should be noted, however, that also has primarily driving has primarily \mathcal{F}_max been investigated using regular rhythms, the implicit assumption being that the mean auditory rate influences the mean visual rate. \cdots \cdots \cdots \cdots \cdots \cdots *ensemble coding* (**A** \cdots /, 2011; A_z (\sim , 2001) substitution of perceptual averaging can be rapidly can be replaced can be rapidly can be replaced by A_z

accomplished even from a set of variant objects or events; for example, we can get can get can get the average size of application of applications in application of applications in application of \mathcal{P}^W supermarket display, or the average tempo of a piece of music. With regard to the present context, audiovisual integration, it remains an open question how the average tempo in audition quantitatively influences the temporal processing of visual e^{\pm} events—an issues prominent as that becomes prominent as the mechanisms underlying perceptual averaging processes themselves are still a سم المحمد المركزي المناسبة. تما المحمد ا
المحمد المحمد المحم lying the representation of magnitudes (e.g., visual numerosity and temporal durations) are non-linear rather than linear (Allan $\&$ \ldots , 1, 1, Dehaene, Izard, Spelke, Book, Spelke, Sp \sim \sim 2003). It has also been reported that, in temporal bisections of the model (x_i, y_i) comparing one interval with two reference intervals y_i, y_i subjective middle middle middle middle middle short and one long reference long reference \mathcal{N}_c

distinct percepts of visual apparent motion: *element* or *group* $\frac{1}{\sqrt{w}}$ motion, where the type of apparent motion is mainly determined by $\frac{1}{\sqrt{w}}$ by the visual interval interval (ISI) between the two displays \mathbb{R}^N frames (with other stimulus settings being fixed). Element motion is typically observed with short ISI (e.g., 0 ms), and group motion with long $\left(1, 2, 0, \ldots, 1, 1, \ldots, 1\mathbf{A} \right)$. When two beeps are presented in temporal proximity to, or synchronously with, the two visual frames, the beeps can systematically bias threshold between the types of visual between the two types of visual between the apparent motion: either toward toward element motion (if the auditory $\frac{1}{\sqrt{2}}$ interval, is shortered than the visual interval interval intervals in the visual intervals of the visual intervals of $\frac{W_{\rm{out}}}{\sigma}$ motion (if ISIA is longer than the visual interval; λ is λ et al., 2010). Similar temporal ventriloquism effects have also been found with $\frac{W_{\rm tot}}{W_{\rm tot}}$ other tasks, such as temporal order judgments (for a review, $\frac{N}{2}$, \ldots , $\frac{N}{2}$, \ldots \ldots & \ldots , 201). Here, \mathbb{R}^N is the Ternus temporal. ventriloquism paradigm by presenting a $\mathbb{W}_{\mathbb{R}^n}$ whole sequence of bee prior to the Ternus display frames, in addition to the two beautiful to th paired with Terminal frames (see Figure 1C; recall that previous call that previous \mathcal{N}_{max} studies had presented just the latter two beeps) to examine the influence of the temporal averaging of auditory intervals on visual apparent motion.

Experiment $1/\frac{W}{\pi}$ of the first instance, to demonstrate, the first instance, to an auditory driving effect using this new paradigm. In Experiment 2, $^{\prime\prime\prime}$, $^{\prime\prime\prime}$, we went on the with irregard with irregard with irregular auditory sequences would have a similar impact on visual impact on visual impact on \mathcal{W} apparent motion. In Experiment 3, $^{\mathrm{W}}$, we manipulate the variability $^{\mathrm{w}}$ of the auditory sequence to examine for (and quantify) influences of the variability of the auditory intervals on visual apparent motion. In Experiment 4, we further determined which the determined which types of $\mathcal{W}_{\mathbf{z},\mathbf{z}}$ temporal averaging statistics, the AM or the auditory \mathbf{A} or the auditory \mathbf{A} intervals, influences visual Ternus apparent motion. And Experimental Termus apparent motion. And Experimental ment 5 was designed to rule out a potential conformation of the potential conformation of the potential confor " بعد الموسودية المتحديث المستعمل المستعمل المستعمل المستعمل المستعمل المستعمل المستعمل المستعمل المستعمل المس Ternus motion percept—in the cross-modal temporal averaging. Finally, we are the computation to identify the computational model the computational model that best $W_{\rm{max}}$ describes the cross-modal temporal interaction: mandatory full Bayesian integration versus partial integration (Ernst & Banks, 2002 , \lt , \lt , \lt , \lt , \lt , \lt , 200).

Mate is a d Met d

Pa_{tc}_{at}

A $t_{\text{max}} = \frac{1}{1-\frac{1}{2}}$ (21, 22, 1, 12, 12, in Eq. 1. 1) ages ranging from 18–33 years) took part in the main experiments. $\mathbf A$ l observers had normal or corrected-to-normal vision and reported normal hearing. The experiments were performed in compliance with the institutional guidelines set by the $\mathbf{A}_{\perp 2}$ and $\mathbf{A}_{\perp 2}$ \mathbf{A}_{max} s Committee of the Department of Psychology, Peking University Un versity (approximation proved proved proved proved proved proved proved proved proved $(2012-03)$ 01] –). All observers provided written in formed conservers provided consent according consent according to \sim to the institutional guidelines prior to participating $\mathbb{M}_{\geq 0}$ and were paid were pa for the time on \mathbb{Z}_{2n} and \mathbb{Z}_{2n} The number of participants records 1 and 2 \mathbf{w} based on the effect size in our previous study of the temporal of th Termina version ventriloridage (Shi et al., 2010), where the pairing of the pairing of $\mathcal{L}_{\mathcal{A}}$ auditory between the visual Termula \mathcal{N}_{max} with the visual Termula displays \mathcal{N}_{max} *d* greater than 1 for the modulation of the Ternus motion percept. We thus used a conservative effect size of 0.25 and a power of $^{\rm W}$ and 0 for the estimation and recruited more than the estimated sample s ize (or 1 participants). Given the effects we are at the effec examine turned out to be quite reliable, we used a standard sample reliable, we used a standard sample sample s' is 12 participants in Experiments 4 and 5.5 \pm 4.5 \pm

A a_ratus and St

 $T_{\rm eff}$ experiments were conducted in a dimensional literature: $\sqrt{M_{\rm eff}}$ and $\sqrt{M_{\rm eff}}$ $\mathcal{F}^{(1)}$ cabin. Visual stimuli were presented in the central region of $\mathcal{F}^{(1)}$ -22 -in. Crath monitor (FD 225P, Qing Dao, China), $\frac{W}{\sqrt{2}}$ resolution $1,024 \times 100$ and a refresh rate of ϵ . $\mathcal{W} = \mathcal{W}$ ing distance was 57 cm, maintained by using a chin rest. ${\bf A}$ visual Ternus display consisted of two stimulus frames, each two stimulus frames, each α containing two black disks (local containing $(0.24 \times l)^{-2}$ containing and separate and separate ration between disks. 1.6° and 3° of visual angles, respectively, respectively, respectively, α presented on a gray background (16.1 cd/m²). The two \mathcal{N}_{max} is shared one element location at the center of the monitor, while $\mathcal{W}_{\mathbf{x}}$ containing two other elements located at horizontal located at horizontal positions relative to the center ℓ_1 for 1 , ℓ_2 frame was was was was was was was was the center of ℓ_1 presented for 0 ms; the interval $(1,1)$ interval \mathcal{N} interval the interval \mathcal{N} $t_{\rm max}$, 0.230 ms, where $\mathcal{L}^{\mathbf{w}}$ is a step size of 30 ms. \ldots sound between $(1000$ Hz, 65 dB, 30 ms) $\frac{W}{\sqrt{2}}$ dB, 30 ms $\frac{1}{2}$ delivered via an M-Audio card (Delta 1010, Bei Jing, China) to an

headset (Philips SHM1900, Bei Jing, China). To ensure accurate $t_{\rm max}$ timing of the auditory and visual stimuli, the duration of the duration of the visual stimuli, the visual stimul stimuli and the synchronization of the auditory and visual stimuli \mathcal{W}_{σ} and monitoring via the monitoring pulses \mathcal{E}_{σ} is vertical synchronization pulses. T experimental program with $\sum_{i=1}^{N} \frac{1}{i} \sum_{j=1}^{N} \frac{1}{j} \sum_{$ Δ , Ma) and the Psychophysics Toolbox (Braince (A,λ)).

E e_j e_j ta De

P act ce. $\int_{\mathbb{R}^n}$ to the formal experiment of the formal experiment, $\int_{\mathbb{R}^n}$ familiarized with visual Ternus displays of either typical elements of either typical elements of \mathcal{C} $\begin{pmatrix} w' & w' & w' \\ w' & w' & w' \end{pmatrix}$ of $(0, 1)$ or $w' = 1$ or typical group $(1, 1)$ 200 ms) in a practice block. The problem to discriminate the discriminate to discriminate the discri $t_{\rm{max}}$ types of apparent motion by pressing the left or the right or the ri mouse button, respectively. The mapping button, respectively. The mapping but the map \mathcal{W} and type of motion was contracted across participants. Durants \mathbf{P} across \mathbf{P} ing practice, when a response was made that was made that with with with $\frac{w}{w}$ inconsistent with with the typical motion percept, immediate feedback appeared on the s showing the typical response (i.e., e.e., element or group \mathcal{C} tion). The practice session continued until the participant reached \mathcal{L} conformity of 95 \mathbf{A} participants achieved this criterion with \mathcal{L} 120 t_{max} , t_{max} to the two extreme ISIs used (50 and 260 ms, respectively) gave rise to non-model percepts of either elements of either elements of either elements of eith motion or group motion. **Pretest.** For $e = 1$, the transition threshold between the transition e^{W} is element and group motion was determined in a pretest session $\mathbf A$ $\frac{1}{\sqrt{M}}$ began with the presentation of a central factor of A

 $(\gamma_{\rm eff})$ visual Ternus apparent motion and formal experimental e iments, as \mathcal{N}^W as fitting the corresponding cumulative Gaussian cumulative Gaussian cumulative Gaussian p sychometric functions. Based on the psychometric functions, $\overline{\mathcal{N}}$ could then estimate the discrimination variability of Ternus appar- ϵ ent motion (i.e., σ_m) based on the standard deviation of t cumulative Gaussian function. The parameters of the Bayesian models (see Bayesian modeling section below) were estimated by minimizing the prediction errors using the R optim function. Our raw data, together with the source code of statistical analyses and source code of statistical and α Bayesian modeling, are available at the github repository: [https://](https://github.com/msenselab/temporal_averaging) والموا**لية** التي توجد من مأتي المدينة التي توجد من

Re \cdot t

E $e \leftarrow e + 1$ and 2: **B**_i **Regular** and **I**_i**e** and **E** A dt $2I$ te_r a A te_rt eV_{is} a M t Perce t

$\begin{array}{l} \mathcal{N}^{(n)}(\mathbb{R}^{n},\mathbb{R}^{n},\mathbb{R}^{n})\cong\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb{R}^{n}\times\mathbb$

These results are interesting in two results are interesting to \mathcal{F}_c mandatory, full Bayesian integration (see the Bayesian Modeling section below for details), auditory-interval variability should affects the weights of the crossmodal temperature ℓ integration (Buus, ℓ 1 , singlet al., \sim 2013), $\mathcal{N}_{\rm tot}$ the influence less ning the influence less ence of the auditory interval. Accordingly, the slopes of the slopes of the slopes of \mathbf{A}_c the fitted lines in 2^{μ} would be expected to be flatter under the flatter under the flatter under the flat high compared with the low \mathbb{W}_{α} condition, yielding and interaction, yielding and interaction, \mathbb{W}_{α} between means interval and contribution of the fact that the fact that the fact that the fact that the fact th nonsignificant suggests that the ensemble mean of the auditory integrated with the visual integrated with the visual integrated with the visual integrated with $\binom{W}{N}$ return to this point in the Bayesian Modeling section). Second, the down was down as a shift of the low, compared with the shift of the low, $\frac{W}{\sqrt{N}}$ \mathcal{L}^{∞} condition indicates that the perceived auditory mean intervals of \mathcal{L}^{∞} $(\omega_{\perp\omega,\lambda\omega}$ influences the audio-visual integration) is actually not the ω \mathbf{A} , we may we make the manipulated account of this shift may be shown of this shift may be shown in this shift may be derive from the fact that the auditory sequences \mathcal{M}_{c} that the auditorial contribution \mathcal{M}_{c} have a lower GM than the sequences with low variance, the sequence \mathcal{N} is the sequence, that is the sequence of \mathcal{N} perceived ensemble mean is likely geometrically encoded. Experiment $\sqrt{\frac{W}{\omega}}$ was designed to address this (potential) confound by directly comparing the effects of ensemble coding based on the \mathcal{A} versus \mathcal{A} and \mathcal{A}

E e_1 , e_2 **t** 4: Pe e_2 **t** a A e_1 **A** d**t** \cancel{q} I te_d a A_{ssimilate the V_{ist} a I te_d a T_c a_d the} **GM Rat e** i T a t e AM

In Experiment 4, we compare three types of auditory sequences of auditory sequence three types of auditory sequences. in our audiovisual Ternus apparent motion paradigm: a baseline sequence, and $\Lambda_{\rm S}$ sequence, and a GeoM sequence. The PSEs sequence. The PSEs sequence. The PSEs sequence $\mu^{\mathbf{w}}$ - 1. (

 $\hat{I}_{full} = wI_a + (1-w)I, TF2850x.2110T9.51wTF 11Tf.61Tf900925821Tf1.0546Tm, TF101Tf$ *.460829 f)-55 316.179II* 1(w)Tj

effect, that is, a dominant influence of the last interval prior to the last interval the Ternus frames. Using a Bayesian integration approximately \mathcal{N} showed that the behavioral responses are behavioral responses are best particles are best particles by particles cue integration, rather than by full integration. Thus, our results reveal the processing—in particular, the temporal averaging—to

 $\epsilon_{\rm eff}$ eraging of the auditory sequence (regardless of its regularity) that it is regularity that $\epsilon_{\rm eff}$ exerted a great influence on the visual interval.

Te a Aea a d Ge et c d

- S , \overline{S} , $\overline{S$ time perception. *Trends in Cognitive Sciences, 17,* 556–564. [http://dx](http://dx.doi.org/10.1016/j.tics.2013.09.009). http://dx. 11/dx. 11/dx. 11/dx. 11/dx. http://dx. 11/dx. 1 .doi.org/10.1016/j.tics...2013.009.jp/
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science, 145,* 1 2. 1 0. $\frac{1}{2}$, 1 0. $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$
- $S_{\rm max}$, β , $\mathbf{A}_{\rm n}$ & $R_{\rm max}$ and spatial dependence, G. (2001). The spatial dependence of the spatial dependence of dency of the ventriloquism effect. *NeuroReport: For Rapid Communication of Neuroscience Research, 12, 7–10. http://dx.doi.org/Neuroscience Research, 12, 7–10. http://dx.doi.o* 00001. $-200101220-0000$

 μ , V. (2005). A theory of magnitude: Common cortical metrics of magnitudes: t_1, t_2, t_3, t_4, t_5 . *Trends in Cognitive Sciences, 7,* \blacksquare ..., 12.