Journal of Mathematical Psychology 99 (2020) 102443

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Audiovisual integration in depth: Modeling the effect of distance and stimulus effectiveness using the TWIN model

Nathan Van der Stoep [a,*], Hans Colonius [b], Jean-Paul Noel [c], Mark T. Wallace [c,d,e,f], Adele Diederich [g]

[a] *Department of Experimental Psychology, Helmholtz Institute, Utrecht University, The Netherlands*
[b] *Department of Psychology, Oldenburg University, Oldenburg, Germany*
[c] *Vanderbilt Brain Institute, Vanderbilt University Medical Center, Vanderbilt University, Nashville, TN 37235, USA*
[d] *Department of Psychology, Vanderbilt University, Nashville, TN 37235, USA*
[e] *Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN 37235, USA*
[f] *Department of Psychiatry, Vanderbilt University Medical Center, Nashville, TN 37235, USA*
[g] *Life Sciences & Chemistry, Jacobs University, 28759 Bremen, Germany*

ARTICLE INFO

Article history:
Received 10 February 2020
Received in revised form 23 August 2020
Accepted 25 August 2020
Available online 19 September 2020

Keywords:
Multisensory integration
Auditory
Visual
Depth
Space
Temporal binding window
Modeling

ABSTRACT

Integrating information across our various sensory modalities results in striking behavioral benefits. This integration depends on a variety of factors, among which are the effectiveness of the stimuli and the relative timing between them. Both of these factors physically vary as a function of the distance between stimuli and the observer: intensity decreases as a function of distance for both auditory and visual stimuli, while the relative timing of energy arriving at the sensory organs differs due to differing transmission speeds. As a result, the depth at which multisensory stimuli are presented is likely to be an important factor in the gain that is derived from integrating them. Here, we use a computational approach – the Time-Window-of Integration (TWIN) framework – to examine differences in simultaneity judgments and reaction times to audiovisual stimuli presented at two depths. Using the TWIN model, we tested whether the observed behavior could be explained solely on the basis of differences in peripheral processing times, on the basis of changes in the temporal binding window (TBW), or by a combination of both factors. The results indicated that a model allowing for different TBWs for near and far space best accounts for the observed data in the majority of participants. However, the best overall model (regardless of the number of parameters) was a model containing both distance-dependent peripheral processing times and TBWs. Interestingly, TBWs were found not to expand from near to far space, but rather to get smaller. Taken together, the results indicate that distance is an additional factor in multisensory integration, above its impact on relative timing and intensity.

1. Introduction

The integration of sensory input from multiple senses is known to lead to an array of behavioral benefits (Calvert, Spence, & Stein, 2004; Spence, & Driver, 2004; Stein, & Stanford, 2008). Multisensory integration (MSI) can lead to more accurate (Odegaard, Wozny, & Shams, 2015), more precise (Alais, & Burr, 2004; Ernst, & Banks, 2002; Rohde, van Dam, & Ernst, 2016), and faster (Gondan, Niederhaus, Rösler, & Röder, 2005; Miller, 1982, 1986) detection and localization responses. Two classical principles that govern MSI describe the need for stimuli from different sensory modalities to be presented in close spatial and temporal proximity (within a certain spatial and/or temporal binding window (TBW)). That is, sound and light have to be presented from roughly the same location (Meredith, & Stein, 1996; Spence, 2013) at roughly the same time (Chen, & Vroomen, 2013; Meredith, Nemitz, & Stein, 1987; Spence, & Squire, 2003; Vroomen, & Keetels, 2010) to elicit robust multisensory gain. What seems most important with regard to the principle of temporal alignment is approximate physiological synchrony (i.e. no or a small delay between when stimuli from different senses trigger a response, both behaviorally and neurophysiologically). For behavior, this principle has been labeled the principle of congruent effectiveness and describes the need for stimuli to be more or less equally effective in triggering a response (Otto, Dassy, & Mamassian, 2013). A third principle is based on the neurophysiological observation that the amount of multisensory enhancement (i.e., the absolute number of spikes) is inversely

---

[*] Correspondence to: Langeveld building, Room H0.26, Heidelberglaan 1, 3584 CS, Utrecht, The Netherlands

E-mail address: N.VanderStoep@uu.nl (N. Van der Stoep).

https://doi.org/10.1016/j.jmp.2020.102443
0022-2496/

related to the effectiveness of the unisensory component stimuli (Meredith, & Stein, 1986).

The translation of these physiological principles to behavior is non-trivial (see Otto et al., 2013, for "behavioral principles" that differ from the physiological ones). For example, an overall inverse effectiveness pattern has been demonstrated for audiovisual speech perception, but somehow the benefits at intermediate signal-to-noise ratios stand out (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). Furthermore, in terms of multisensory facilitation of response times, the opposite of an inverse effectiveness pattern was observed: facilitation increased as stimulus intensity increased (Leone, & McCourt, 2013). Various studies have shown that the relative effectiveness and reliability of auditory and visual stimuli are important for maximum multisensory facilitation (Hershenson, 1962; Leone & McCourt, 2013; Otto et al., 2013; Raab, 1962). In fact, many variables intuitive in behavior, such as "depth"/distance from the observer, do not have (as far as we know) direct codes at the level of single neurons (i.e., depth being a combination of intensities, binocular disparity, auditory reverberation, and more). Here, we attempted to add to the characterization of the "principles"/governing MSI at the behavioral level by establishing whether depth, above and beyond its impact on peripheral processing times (i.e., intensity) and relative timing of cues, is an important player in MSI (see Van der Stoep, Serino, Farnè, Di Luca and Spence, 2016, for a recent review).

Both the temporal relation between auditory and visual inputs and their relative effectiveness change as a function of distance (Spence & Squire, 2003; Van der Stoep, Nijboer, Van der Stigchel, & Spence, 2015; Van der Stoep, Serino et al., 2016). For example, the same visual stimulus viewed at a greater distance has a smaller retinal image size, and information from the same auditory and visual stimuli arrive at the sensory organs with less energy when presented at a greater distance. Given this fact that stimulus effectiveness decreases as the distance between the stimulus and the observer increases, one may expect to find changes in multisensory integration with increasing distances according to the principle of inverse effectiveness. Indeed, it was recently shown that MSI of auditory and visual stimuli that were presented in far space resulted in larger multisensory gain relative to stimuli that were presented in near space (Van der Stoep, Van der Stigchel, Nijboer and Van der Smagt, 2016). Interestingly, when same change in retinal image size and intensity caused by the increase in distance from near to far space was applied to stimuli presented in near space, there was no increase in multisensory gain. Thus, only the combination of an increase in distance and a decrease in retinal image size and intensity resulted in increased multisensory gain.

The distance between a stimulus and an observer not only affects the effectiveness (i.e., intensity) of those stimuli, but also the temporal relation of synchronously presented audiovisual stimuli due to the differing speeds of transmission of sound and light energy in air (e.g. Alais, & Carlile, 2005; Arnold, Johnston, & Nishida, 2005; Di Luca, 2014; Engel, & Dougherty, 1971; Kopinska, & Harris, 2004; Sugita, & Suzuki, 2003). Given these differences, auditory and visual information will arrive at the sensory organs at different times when presented synchronously at a large distance from the observer (Spence & Squire, 2003; Vroomen & Keetels, 2010). In accordance with the temporal principle, such delays will affect various aspects of how these stimuli are integrated. For example, the distance between paired audiovisual stimuli and the observer modulates the audiovisual delay that is perceived as being synchronous (i.e. the point of subjective simultaneity, PSS, e.g. Alais & Carlile, 2005; Engel & Dougherty, 1971; Kopinska & Harris, 2004; Lewald, & Guski, 2004; Sugita & Suzuki, 2003). Humans seem to be able to take into account the difference in arrival times when asked to judge whether or not sound and light were presented simultaneously. However, when they have to judge whether they perceive them as being synchronous not taking distance into consideration, the PSS shifts when the distance between the stimuli and the observer increases. It was recently observed that the size of the TBW also changes with increases in distance. Although one may expect that larger delays between auditory and visual events that appear farther away cause an increase in the size of the TBW, the opposite was observed (Noel, Lukowska, Wallace and Serino, 2016; Noel, Modi, Wallace and Van der Stoep, 2018). That is, the TBW was larger for audiovisual events that were presented in peripersonal/near as compared to extrapersonal/far space.

Thus, the distance from which we observe audiovisual sources seems to affect MSI in several ways. Unclear to date is how changes in timing and stimulus effectiveness contribute to changes in MSI with increasing distance. To investigate this, both response time (RT) and simultaneity judgment (SJ) data from Noel, Modi et al. (2018) was analyzed within the time-window-of-integration framework (TWIN model; (Colonius, & Diederich, 2004; Diederich, & Colonius, 2015). The TWIN model is a quantitative stochastic framework that has previously been successful in describing multisensory behavioral data and can provide unique insights into multisensory processes through the various parameters of the model (Colonius & Diederich, 2004). For example, using the TWIN model it has been demonstrated that the probability of MSI is smaller despite a larger TBW in the elderly relative to younger individuals (Noel, De Niear, Van der Burg and Wallace, 2016), a result of a slowing of peripheral processing (Diederich, Colonius, & Schomburg, 2008).

In the current modeling study, we focused on two parameters: unisensory peripheral processing times (which are affected by stimulus effectiveness) and the size of the TBW. Three different TWIN model versions were fitted to the SJ and RT data to investigate the role of these two parameters in modulating MSI in near and far space: (1) A distance-dependent stimulus effectiveness model with different unisensory peripheral processing times for near and far space but with a single TBW size for both distances, (2) a distance-dependent TBW model with identical peripheral processing times but different TBW sizes in near and far space, and finally (3) a distance-dependent TBW and stimulus effectiveness model where both peripheral processing times and the size of the TBW were allowed to vary with stimulus distance. The performance of these distance-dependent models was compared to a distance-independent TWIN model to see whether any distance-dependent parameters had additional value in describing multisensory behavior for stimuli in near and far space.

2. Methods

2.1. Description of the original study

The final dataset from Noel, Modi et al. (2018) and Noel, Serino and Wallace (2018), available at https://osf.io/b6hvs/) (Noel, & Van der Stoep, 2017) consisted of data from 32 participants (16 females, mean age = 21.11 years, range = 18–26 years). Participants took part in two tasks: a simultaneity judgment (SJ) and a multisensory redundant signal task (RST). Auditory stimuli were 40 ms of a 3000 Hz pure tone presented at 75 dB(A) and visual stimuli consisted of blue Light Emitting Diodes (LED) that were turned on for 40 ms. Stimuli could be presented from six different locations, three in near and three in far space (at a central location and at 26° to the left or right of the central location, see Fig. 1). Audiovisual stimuli were always spatial aligned and were presented at different stimulus onset asynchronies (SOA, ±350, ±250, ±150, ±50, or 0 ms with negative SOAs indicating that the sound was presented first).

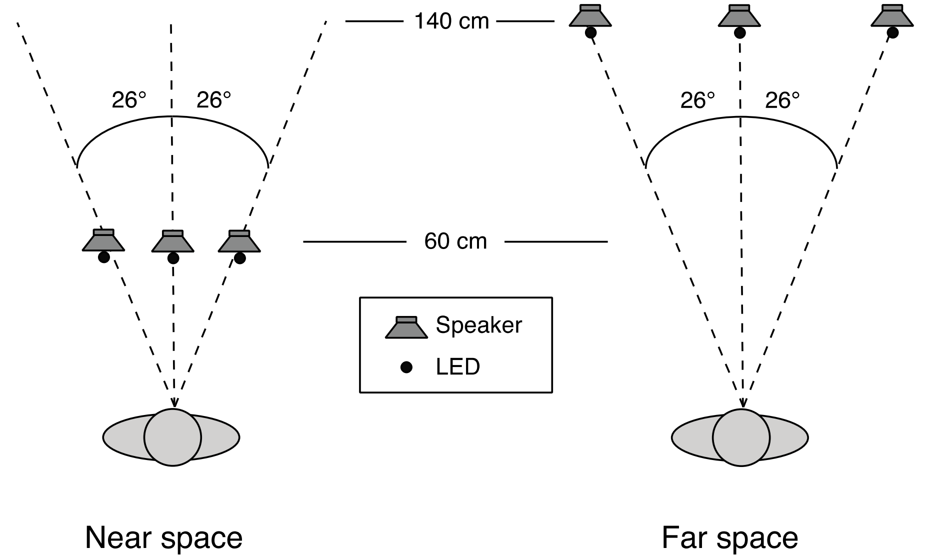

Fig. 1. The experimental setup that was used in the study by Noel, Modi et al. (2018) and Noel, Serino et al. (2018).

The SJ task consisted of 30 trials per SOA for the left, 30 trials per SOA for the right, and 15 trials per SOA for the central location. Stimuli were presented either in near or far space (blocked). In total, the SJ task contained 1350 trials (675 trials at each distance). Participants had to indicate whether they perceived the A and V stimuli were presented simultaneous or not by pressing one of two buttons.

The RST task required an implicit spatial discrimination. That is, participants had to press a button as fast as possible when A, V, or AV stimuli appeared at one of the peripheral locations (Go trials) and withhold their response when a target appeared at the central location (No-go trials). There were 40 auditory-only and 40 visual-only Go trials (20 left and 20 right per condition), and 40 AV Go trials (again 20 left, and 20 right) per SOA. There were 20 unisensory No-go trials and 10 AV No-go trials per SOA. In total, this task consisted of 550 trials.

2.2. Time-Window-of-Integration (TWIN) model

The time-window-of-integration (TWIN) model (Colonius & Diederich, 2004; Diederich, & Colonius, 2004) assumes that a multisensory (e.g., audiovisual) stimulus triggers a race mechanism in early peripheral sensory pathways (*first stage*), followed by a compound stage of converging sub-processes comprising integration of the input and preparation of a response.[1] This *second stage* is defined by default: it includes all subsequent, possibly temporally overlapping, processes that are not part of the peripheral processes in the first stage. Thus, the TWIN model has two separate sources of facilitation: "statistical facilitation" (i.e. probability summation) in the first stage, and co-activation in the second stage. The outcome of each stage added up in such a way that there can be, for example, facilitation in the first and second stage, but it could also occur that inhibition in the second stage evens out any facilitation from the first stage. The central assumption of the model concerns the temporal configuration needed for multisensory interaction to occur:

> TWIN assumption: Multisensory interaction occurs *only* if the peripheral processes of the first stage all terminate within a given temporal interval, the 'time window of integration'.

Thus, the time window acts as a filter determining whether afferent information delivered from different sensory organs is registered close enough in time to trigger multisensory integration. Passing the filter is necessary but not sufficient for multisensory interaction to occur, since the amount of interaction may also depend on many other aspects of the stimulus set, in particular the spatial configuration of the stimuli. The amount of multisensory interaction manifests itself in an increase, or decrease, of second stage processing time. Although this amount does not directly depend on the stimulus onset asynchrony (SOA) of the stimuli, temporal tuning of the interaction occurs because the *probability of integration* is modulated by the SOA value.

The race in the first stage is represented by two statistically independent, non-negative random variables V and A, standing for the peripheral visual and auditory processing times, respectively. With τ as SOA value and ω as integration window width parameter, the TWIN assumption for multisensory integration to occur, denoted by I, is formally defined by

$$I(\tau, \omega) = \{\max(V, A + \tau) < \min(V, A + \tau) + \omega\}. \tag{1}$$

for the redundant signal task. The dependence of I on τ and ω is usually not made explicit in writing $\Pr(I)$ for the probability of integration to occur. Note that while τ is a physical parameter given in the experimental setup, ω is a model parameter that must be estimated once distribution functions for A and V have been specified.

Response time. Writing S1 and S2 for first and second stage processing times, respectively, overall expected response time in the multisensory condition with SOA equal to τ, denoted as $\mathrm{E}[RT_V, A + \tau]$, is computed by conditioning on the events of integration occurring (I) or not occurring (C):

$$\begin{aligned}\mathrm{E}[RT_{VA,\tau}] &= \mathrm{E}[S_1] + P(I)\,\mathrm{E}[S_2|I] + [1 - P(I)]\,\mathrm{E}[S_2|C]\\ &= \mathrm{E}[S_1] + \mathrm{E}[S_2|C] - P(I) \times \Delta.\\ &= \mathrm{E}[\min(V, A + \tau)] + \mu - P(I) \times \Delta.\end{aligned} \tag{2}$$

Here, μ is shorthand for E[S2|C] and ΔI stands for the difference E[S2|C] − E[S2|I]. Thus, the term Pr(I) × ΔI is a measure of the expected amount of multisensory interaction in the second stage, with positive ΔI values corresponding to facilitation, and negative ones to inhibition. Obviously, event I cannot occur in the unimodal condition, thus the expected unimodal reaction times are

$$\mathrm{E}[RT_V] = \mathrm{E}[V] + \mathrm{E}[S_2|C] \quad \text{and}$$

---

[1] The model is not a biophysically realistic model of neural processing, but rather an abstract psychological model of behavior.

$$\mathrm{E}[RT_A] = \mathrm{E}[A] + \mathrm{E}[S_2|C],$$

for the visual and auditory stimulus, respectively. Although the model developed so far allows a number of empirically testable predictions (see, e.g., Diederich, & Colonius, 2007), fitting the model to data requires that the distributions for A and V are specified. Following previous studies (e.g., Diederich, & Colonius, 2008), we assume exponential distributions for A and V . Moreover, adding a Gaussian component to the second stage S2 results in RT distributions as a mixture of ex-Gaussian distributions, which are quite common in RT modeling (see Luce, 1986).

Simultaneity judgments. In order for the TWIN model to predict judgments of simultaneity as well, some further assumptions are required, in analogy to the extension for temporal order judgments introduced in Diederich and Colonius (2015). Typically, data from an audiovisual simultaneity judgment (SJ) task are presented in the format of the relative frequency of responding "visual and auditory stimulus are simultaneous" given the specific stimulus onset asynchrony τ between the visual and auditory stimulus. These frequencies yield an estimate for the psychometric function

$$\Psi(\tau) = \Pr(\text{Response “visual and auditory stimulus simultaneous”} \mid \tau).$$

The extended TWIN model assumes that a participant responds "simultaneous" under two conditions. First, when integration takes place this results in a unitary "event" preventing the participant from perceiving which modality was presented prior to the other one. Alternatively, when no integration occurs the participant perceives the onset of the auditory prior to the visual or vice versa; in both cases there is a tendency to respond "simultaneous" with a certain probability β. Thus, the psychometric function is

$$\Psi(\tau) = \Pr(I) + [1 - \Pr(I)]\,\beta = \beta + (1 - \beta)\Pr(I)$$

Inserting the definition of I (Eq. (1)) yields

$$\begin{aligned}\Psi(\tau) &= \beta + (1-\beta)\Pr(\max(V, A+\tau) < \min(V, A+\tau) + \omega)\\ &= \beta + (1-\beta)\,[\Pr(A+\tau < V < A+\tau+\omega)\\ &\quad + \Pr(V < A+\tau < V+\omega)].\end{aligned} \tag{3}$$

Obviously, more elaborate mechanisms for generating simultaneity judgments are conceivable (see García-Pérez, & Alcalá-Quintana, 2015). For example, the tendency to respond "simultaneous" may depend on the order in which the stimulus onsets have been perceived leading to two different β probabilities. In order to determine Ψ, computation of $\Pr(I)$ is required. Denoting the distribution functions of V and A by F_V and F_A, respectively, the probabilities for the events leading to integration can be written as

$$\begin{aligned}&\Pr(A+\tau < V < A+\tau+\omega)\\ &\quad = \int_0^\infty \{F_V(a+\tau+\omega) - F_V(a+\tau)\}\, dF_A(a)\\ &\Pr(V < A+\tau < V+\omega)\\ &\quad = \int_0^\infty \{F_A(v+\omega-\tau) - F_A(v-\tau)\}\, dF_V(v).\end{aligned} \tag{4}$$

Once distribution functions F_V and F_A have been chosen, the psychometric function is completely specified, so that the parameters of the distributions and the value of β can be estimated by fitting the model to both response and simultaneity judgment data. In the following, we assume $F_A(t) = 1 - \exp[-\lambda_A\ t]$ and $F_V(t) =$ 1-exp[-$\lambda_V\ t$] for $t \geq 0$ and positive parameters λ_A and λ_V. Derivations for the mean response times and psychometric functions are found in Appendix A.

For the general case, the parameters to be estimated from the data are listed in Table 1, including value restrictions as used in the estimation procedure. The value restrictions are purely heuristic and based on prior experience. They serve to keep the estimation method "on track". Three versions of the model were fitted to the data with either separate or combined near and far space parameters. The optimization routine *fminsearchbnd* of Matlab (Mathworks, 2019) was used to estimate the parameters. This routine is similar to the standard *fminsearch* routine (Lagarias, Reeds, Wright, & Wright, 1998), except that the range of the parameters can be predetermined. Importantly, all parameters are estimated anew for each model. Thus, although the distance-dependent stimulus effectiveness model and the distance-dependent TBW model are nested models with respect to the combined model, this does not necessarily lead to the best fit for the combined model. See Table 2 for a full description of all parameters included in each model and Fig. 2 for a schematic depiction of each model.

Table 1
Parameters and their restrictions for the estimation routine.

Parameters	Restriction limits (in ms)	
$1/\lambda_V$	10–200	Mean peripheral processing time for visual stimulus
$1/\lambda_A$	5–200	Mean peripheral processing time for auditory stimulus
μ	50–400	Mean central processing time
ω_{RT}	20–600	Window of integration for RT task
ω_{SJ}	20–600	Window of integration for SJ task
Δ	−50–200	Amount of multisensory interaction
β	0–1	Probability of a response bias in the SJ task

Model fit. To compare the fit of the different model versions, we determined both the Root Mean Square Error of Approximation (RMSEA) and the Akaike Information Criterion (AIC). The RMSEA is an absolute measure of fit taking into account the number of parameters in a model (Schubert, Hagemann, Voss, & Bergmann, 2017). We report this measure of fit as it provides information about which model best describes the observed behavior. In addition to this absolute measure of fit, the AIC was calculated for each version of the model and each participant separately using the following formula: AIC = χ2 + k(k + 1) - 2df. This relative measure of goodness of fit can be used to compare the different versions of the model taking into account the number of parameters in the model. The AIC measure penalizes for the number of parameters in the model to allow for a fairer comparison between models with a different number of parameters. Using the AIC puts the model with the most parameters at a disadvantage (in this case the distance-dependent stimulus effectiveness and TBW model, model 3). Thus, while RMSEA details which model describes best the observed data (i.e., the goal is to account for data), AIC describes which is the best model (i.e., the goal is the built a 'good' model). A version of the TWIN model without any distance-dependent parameters (see Table 2) was used as the base model to which the other models were compared. Given that the behavioral data contained an effect of distance, we wanted to select the distance-dependent model that best described the data with the ultimate goal to investigate whether distance-dependent effects were due to changes in stimulus effectiveness, the TBW, or both. Therefore, we calculated the difference in AIC between each distance-dependent model and the distance-independent model to determine the amount of information loss relative to the distance-independent model:

$$\Delta\mathrm{AIC}_i = AIC_i - AIC_0 \tag{5}$$

With i indicating the version of distance-dependent TWIN model and AIC$_0$ the AIC of the distance-independent TWIN model (also always the model with the lowest AIC value). This way, we could analyze whether any of the distance-dependent models had

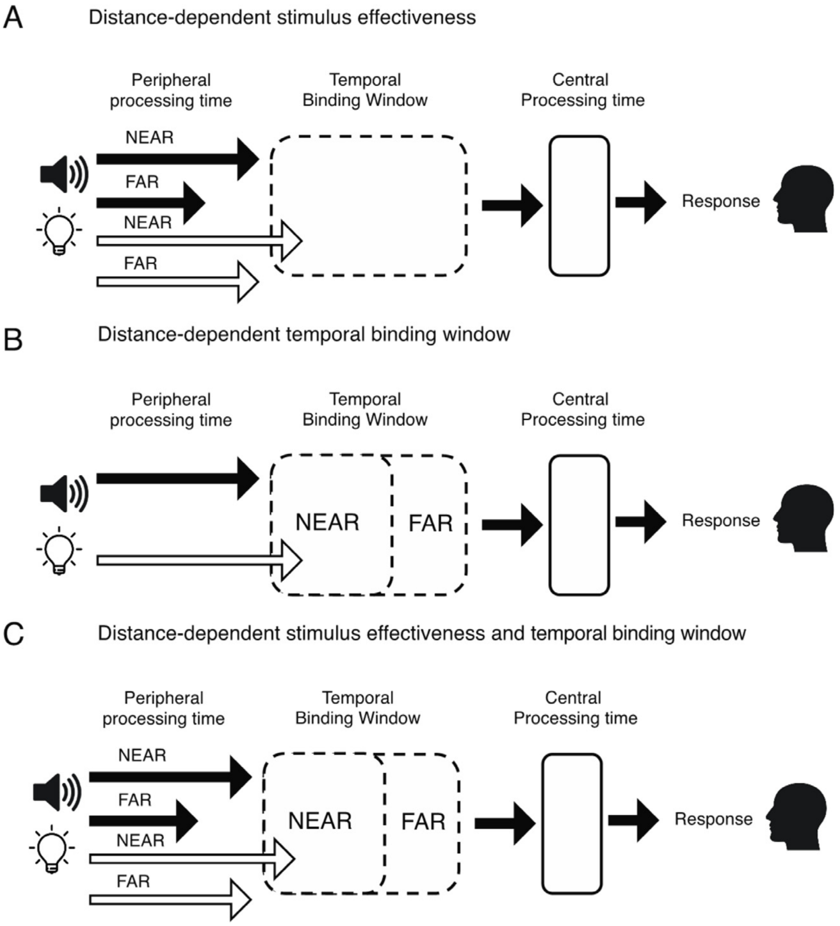

Fig. 2. Schematic depictions of each TWIN model version that was used to fit the SJ and RT data in this study. (A) A distance-dependent stimulus effectiveness model with a fixed TBW for near and far space, (B) a distance-dependent TBW model with a fixed peripheral processing time for near and far, and (C) a model in which both peripheral processing time and the TBW could vary between near and far space.

Table 2
Parameters included in each model version.

Model version	Parameters
Distance-independent model (baseline)	$1/\lambda V, 1/\lambda A, \omega RT, \omega SJ, \mu, \Delta, \beta$
Distance-dependent stimulus effectiveness	$1/\lambda V_N, 1/\lambda A_N, 1/\lambda V_F, 1/\lambda A_F, \omega RT, \omega SJ, \mu, \Delta, \beta$
Distance-dependent temporal binding window	$1/\lambda V, 1/\lambda A, \omega RT_N, \omega \mathrm{SJ}_N, \omega \mathrm{RT}_F, \omega SJ_F, \mu, \Delta, \beta$
Distance-dependent stimulus effectiveness and TBW	$1/\lambda V_N, 1/\lambda A_N, 1/\lambda V_F, 1/\lambda A_F, \omega \mathrm{RT}_N, \omega \mathrm{SJ}_N, \omega RT_F, \omega SJ_F, \mu, \Delta, \beta$

Note: Subscript *N* or *F* indicates separate parameter estimation for near or far space, respectively.

additional value beyond the distance-dependent TWIN model. As a rule of thumb, models with ΔAIC $>$ 10 have basically no support for being a plausible candidate model, models with ΔAIC between 4 and 7 have some support, and models with ΔAIC $\leq$ 2 have substantial support (Burnham, & Anderson, 2004).

For the analysis of how often a certain model was the best model for a participant, the model resulting in the lowest ΔAIC value was chosen as the best fitting model for each participant. This way, for each model, the percentage of participants for whom a certain model was the best fit could be determined. Additionally, we analyzed the parameters of the best model (e.g. peripheral and central processing time and TBW size).

3. Results

3.1. The best fitting model

Overall, the TWIN models provided a reasonable fit to the RT and SJ data (see Fig. 3A and B). First, as a measure of absolute fit (as opposed to a measure indexing the most parsimonious model), the Root Mean Square Error of Approximation (RMSEA) was calculated for each model (see Fig. 4A). In terms of RMSEA, the combined model (see Fig. 2C) was best (M = 0.196, SD = 0.109). A Bayesian repeated measures ANOVA was used to compare the RMSEA between models. This analysis showed a significant effect of Model version ($\mathrm{BF}_{10} = 5.711 \times 10^8$). As expected (given the increased complexity of the distance-dependent stimulus effectiveness and TBW model), post-hoc comparisons between the four models indicated that (1) all distance-dependent models outperformed the distance-independent model (all $\mathrm{BF}_{10} > 3350$), (2) the combined model outperformed the other two distance-dependent models (all $\mathrm{BF}_{10} > 22$), and (3) the evidence for or against a difference between the stimulus effectiveness model and the TBW model was inconclusive (BF_{10} = .754). The latter means that it was not possible to differentiate between the stimulus effectiveness model and the TBW model given the data. These results indicate that a model that takes into account both changes

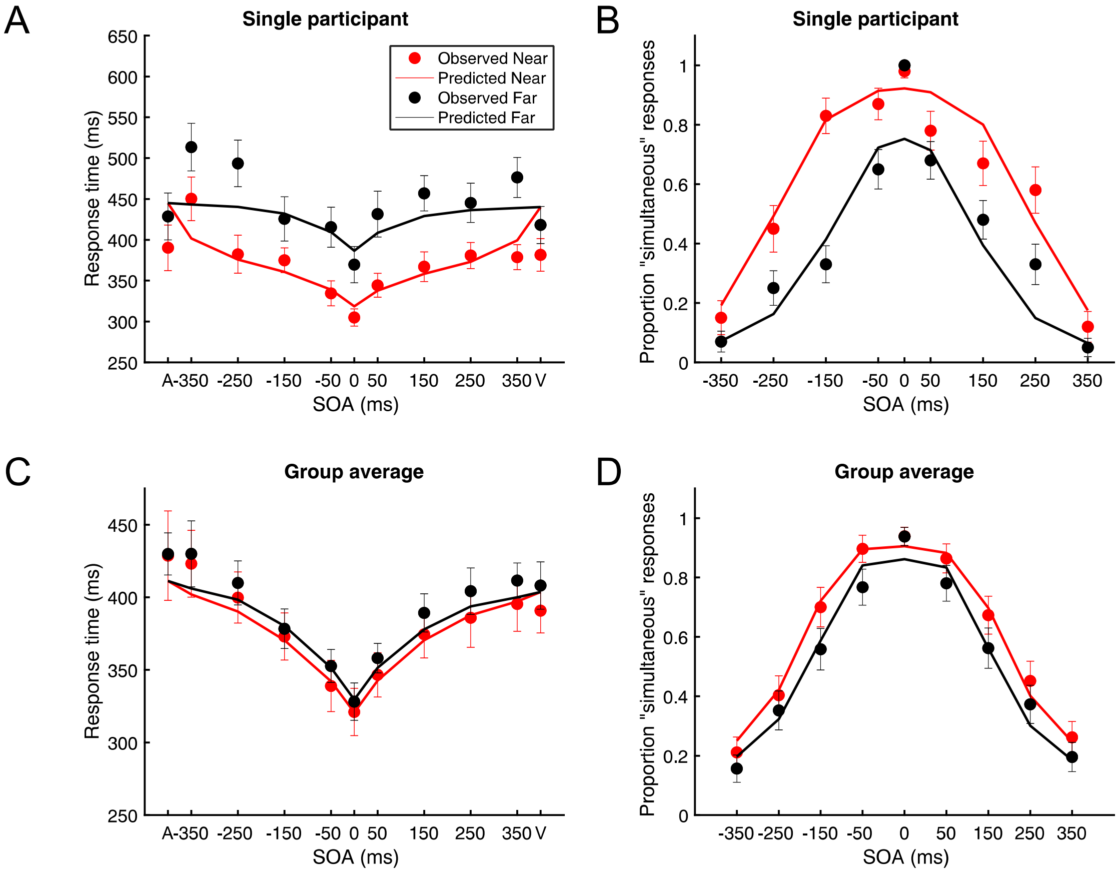

Fig. 3. Behavioral results (dots) and model fit (solid lines) for the RT task (A) and SJ task (B) in near (red) and far space (black) of a single participant with the best fit for the best fitting model according to ΔAIC criterion (the TBW model). The group average behavioral results (dots) and the group average fit (solid line, based on the average of the individual fits) of the best fitting model according to the ΔAIC criterion (the TBW model) for the RT (C) and SJ (D) task for both distances. Error bars indicate standard errors. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

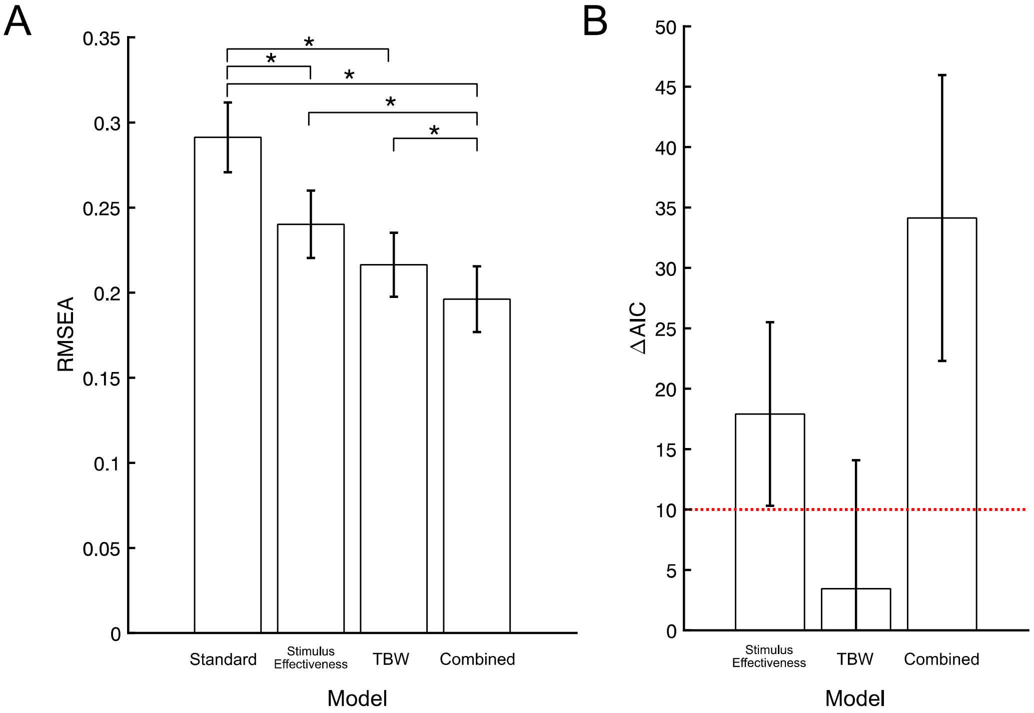

Fig. 4. The average goodness of fit for each model (RMSEA) (A), and the ΔAIC relative to the distance-independent TWIN model (B). Asterisks indicate meaningful differences between models (BF > 5, corrected for multiple comparisons). The dashed red line in panel B indicates the ΔAIC = 10 boundary. Models above this line basically have no support for being a good candidate model relative to the distance-independent baseline model. Error bars indicate standard errors.

in stimulus effectiveness and changes in TBW size with increasing distance best described the data.

To assess whether any of the models could be considered close fitting models, the RMSEA was compared to the close fit RMSEA cut-off of 0.05 using one-sample t-tests (for normally distributed RMSEA data) and Wilcoxon signed rank tests (for not normally distributed RMSEA data; Kelley, & Lai, 2011). All models were worse than close fitting (all $p < .05$, i.e. their RMSEA was significantly larger than 0.05). However, the ratio between our sample size (N = 32) and number of parameters (ranging from 7 to 11) may not have been optimal for testing the closeness of fit of these models.

Next, to estimate which model is the most parsimonious account of the putative neurocognitive mechanisms at play given the near and far RT and JS data, the performance of the distance-dependent models was compared taking into account the number of parameters in each model.

3.2. The most parsimonious model

Taking into account the number of parameters in evaluating each model's performance, the distance-independent TWIN model showed the best fit (i.e. the lowest average AIC value). This may come as no surprise given that this model has the lowest number of parameters. Since we were interested in modeling distance-dependent effects (as evident from the behavioral data), we investigated which distance-dependent model was the most informative using the distance-independent model as a baseline. The ΔAIC values indicated that the distance-dependent TBW model had merit relative to the standard TWIN model. The TBW model was the only model with support as a plausible model with an average ΔAIC of 3.435 (SE = 10.649) and the other two distance-dependent models having an average $\Delta\text{AIC} > 10$ (see Fig. 4, Burnham & Anderson, 2004).

Although the stricter approach of comparing the ΔAIC between models highlights a different model as the best model compared to the RMSEA measure, it provides further evidence that stimulus effectiveness alone cannot account for differences in multisensory behavior between near and far space. Overall, these results indicate that the behavioral data can be best described when a distance-dependent TBW is included in the model (see Fig. 2B and C).

3.3. Model performance per individual

Even though the distance-dependent TBW model had the lowest ΔAIC value, this model did not result in the best fit for all of the participants. To gain insight into how often a certain model outperformed the other models at an individual level, we determined the percentage of participants whose data was best fitted by a certain model based on the lowest ΔAIC value of the three distance-dependent models.

Based on the ΔAIC, the distance-dependent TBW model version (Fig. 2B) resulted in the best fit for the majority of the participants (62.5%, N = 20). The distance-dependent stimulus effectiveness model (Fig. 2A) resulted in the best fit for 37.5% (N = 12) of the participants. The combined model version (distance-dependent stimulus effectiveness + TBW, Fig. 2C) did not prove to be the best fit for any of the participants based on ΔAIC. This may in part be the result of the fact that the AIC measure heavily penalizes for the number of parameters in each model and the combined model version contains the most parameters. Based on the RMSEA, the combined model resulted in the best fit for the majority of the participants (50%, N = 16), followed by the distance-dependent TBW model (31.25%, N = 10), and the distance-dependent stimulus effectiveness model (18.75%, N = 6). The distance-independent model did not result in the best fit for any of the participants.

The ΔAIC and RMSEA values for each participant and model can be found in supplementary Table S1 and S2, respectively. The parameter values of the best fitting model (according to AIC the distance-dependent TBW, see Fig. 2B) of each participant are shown in supplementary Table S3. The group average of the behavioral data and the group average fit of the distance-dependent TBW model can be found in Fig. 3.

3.4. Parameters of the distance-dependent TBW model

Given that the distance-dependent TBW model was the best model according to the AIC criterion, we further analyzed the details of the parameters in this model (see figure S1 for details of group averages of each parameter for each model).

Peripheral processing time: In the model, a peripheral processing time for visual input (Mean $\lambda_V = 83.479$ ms, SD = 36.221) and a peripheral processing time for auditory input (Mean $\lambda_A = 91.043$ ms, SD = 38.954) were estimated regardless of distance (processing times were estimated for near and far space together). There was no difference between these unisensory peripheral processing times ($\text{BF}_{01} = 3.145$). This means that, on average, the unisensory stimulus properties used in the experiment did not result in differences in unisensory stimulus effectiveness.

Temporal binding window: The size of the TBW was estimated separately for near and far space, and separately for the SJ and the RT task. A Bayesian paired t-test for the SJ task indicated that there was a significant difference between the estimated TBW size for near ($M = 217$, $SE = 13$) and far space ($M = 177$, $SE = 12$, $\text{BF}_{10} = 62.114$).

For the RT task, there was no difference in the estimates of the size of the TBW between near (M = 139, SE = 18) and far space (M = 189, SE = 33, BF01 = 3.017). This observation seems to be in contrast with the behavioral findings of the temporal window of fastest detection (TWFD) by Noel, Modi et al. (2018) and Noel, Serino et al. (2018), which showed a larger TWFD in near than in far space. However, the method with which the TWFD was determined in Noel, Modi et al. (2018) and Noel, Serino et al. (2018) was different from determining the TBW in the current study which may explain the different results. In Noel, Modi et al. (2018) and Noel, Serino et al. (2018) the TWFD was taken as an estimate of the TBW. The TWFD was calculated by fitting a Gaussian distribution to the normalized and inverted pattern of RTs in the AV condition across all SOAs per participants. The SD of the fitted distribution was taken as the TWFD.

We also compared the size of the estimated TBW between the SJ and RT task separately for near and far space. The TBW was smaller for the RT task (M = 139, SE = 18) than for the SJ task (M = 217, SE = 13) in near space ($\text{BF}_{10} = 10.447$), but did not significantly differ in far space ($\text{BF}_{01} = 4.968$, RT: M = 189, SE = 33, SJ: M = 177, SE = 12)

Central processing time, multisensory gain, and bias: The estimated central processing time of the distance-dependent TBW model was 320 ms (*SD* = 68) and the estimated multisensory gain was 62 ms (*SD* = 51). The estimate of the response bias in the SJ task was 0.093 (*SD* = 0.035), indicating that there was barely any bias.

4. Discussion

Studies have shown that the behavioral changes stemming from the interaction between auditory and visual information is dependent on the distance from which that information is perceived (for reviews, Previc, 1998; Spence, Lee, & Van der Stoep, 2017; Van der Stoep et al., 2015; Van der Stoep, Serino et al., 2016; Van der Stoep, Van der Stigchel et al., 2016). However, it is unclear whether these distance-dependent modulations of multisensory integration are due to a true distance (i.e., spatial) effect, or result from alterations in the intensity of stimuli (Krueger Fister, Stevenson, Nidiffer, Barnett, & Wallace, 2016) and/or differences in their relative timing. In this study, we used a modeling approach to further investigate data from a recent study demonstrating changes in multisensory temporal processing for stimuli presented in near and far space (Noel, Modi et al., 2018). We tested whether the reported larger TBWs in near as compared to far space are the result of the brain somehow changing its sensitivity to temporal differences between auditory and visual input based on the stimulus-observer distance, or whether these changes in TBW size can be explained simply by changes in stimulus effectiveness (affecting peripheral processing times).

The results of our modeling approach indicate that the observed changes in multisensory temporal processing of stimuli in near and far space appear to be explained by separate TBW widths for near and far space in the majority of participants. This finding supports the notion that the brain deals differently with the temporal structure of audiovisual events depending on the distance between stimuli and the observer (Noel, Lukowska et al., 2016; Noel, Modi et al., 2018). Although the different measures (AIC vs RMSEA) used to compare the models indicated different models as being the best model, both measures indicated that at least a distance-dependent TBW was needed to explain the behavioral data. In addition, the comparison of the RMSEA between models shows that distance-dependent estimates of unisensory peripheral processing times can provide additional explanatory power in describing the observed behavior. Altogether, these findings highlight that changes in the TBW between near and far space cannot be simply explained by differences in stimuli intensity. However, the driving factor or functional reason for the TBW to change with distance remains unclear.

One possibility is that the reliability of sensory signals is distinctly impacted by the distance from which we perceive them. For example, spatial resolution increases with proximity, and thus perhaps temporal factors are underweighted as stimuli get close. Although this could explain these changes in TBW width conceptually, it does not provide any functional or mechanistic explanation of this phenomenon. Alternatively, one could argue that temporal windows of integration emerge from our experience with audiovisual events at different distances (Wallace, & Stein, 2007). Namely, the visual angle/retinal position of visual input and the interaural time and level differences of auditory input change more rapidly when an audiovisual object moves from left to right in close proximity (as opposed to at a greater distance). In addition to the purely physical changes in stimulus size as viewed by the sensory organs, the more rapid changes may increase spatial and temporal uncertainty in near relative to far space.

One may wonder whether the TBWs described in the current models and the behavioral data reflect the temporal tuning of multisensory neurons (in, for example, the superior colliculus, or in cortical multisensory brain regions such as the superior temporal sulcus, Meredith et al., 1987; Noesselt, Bergmann, Heinze, Münte, & Spence, 2012). Temporal binding seems to arise from the fact that discharge trains of multisensory neurons that are evoked by auditory and visual sensory input overlap (Parise, & Ernst, 2016), resulting in the, more or less, simultaneous activation of a multisensory neuron (Meredith et al., 1987). In the context of response times, this principle has been described as the principle of congruent effectiveness (Otto et al., 2013): stimuli from different modalities that result in equal reaction times will result in the greatest multisensory gain (presumably because these signals overlap in space and time in the brain). Say, for example, an auditory stimulus triggers a neural response earlier than a certain visual stimulus in near space (resulting in unequal effectiveness). The neural response to the auditory and visual stimuli should then become more temporally aligned and, after a certain distance, more misaligned when the same stimuli are presented farther from the observer (e.g. due to the lower intensity of the sound at the ear and the lower speed of sound in air). This could provide the basis for an explanation not only as to how stimulus effectiveness but also how the distance from which we perceive sensory input can alter multisensory temporal processing.

Given that more distant stimuli are generally also of lower intensity and results in a smaller retinal image, one could argue that TBWs are smaller due to the fact that discharge trains evoked by weaker unisensory stimuli simply do not overlap as much in time as with stronger stimuli. In contrast, given that less intense stimuli result in weaker and more variable responses, one could also argue that TBWs should become larger. The latter observation is in line with behavioral findings of larger TBWs for weaker stimuli compared to stronger stimuli (Krueger Fister et al., 2016). Thus, so far, it is difficult to come up with a mechanistic explanation considering discharge trains evoked by stimuli at a single cell level to account for the observed changes in TBW size with increases in distance.

Another factor that may play a role in distance dependent changes in the TBW, is the increase in the difference in arrival times of sound and light resulting from the differing speeds of sound and light in air (Spence & Squire, 2003). A well-known effect of this is the observation of an increasing delay between lightning and thunder with increasing distance. Although previous studies of audiovisual synchrony perception have shown that this delay can play a role in the perception of simultaneity (e.g. Alais & Carlile, 2005; Sugita & Suzuki, 2003), the difference in arrival times for the distances used in the current study are most likely negligible. That is, the estimated arrival time difference between auditory and visual stimuli was ∼1.75 ms for the near distance (60 cm) and ∼4.08 ms for the far distance (140 cm). It is highly unlikely that the 2.33 ms change in auditory and visual arrival times between near and far space alone can explain the observed 40 ms difference in the size of the TBW. However, besides these arrival time differences, changes in unisensory signal strength with increases in distance could contribute differences in unisensory processing times. A systematic manipulation of unisensory response time differences at various distances could provide insights into the contribution of these factors to distance-dependent changes in the TBW. So far, however, it seems unlikely that these low-level stimulus properties explain the changes in TBW between near and far space suggesting that the brain has distance-dependent modes for multisensory integration.

It is not unthinkable that there are other ways in which the TBW can change based on the distance of sensory input besides changes in basic stimulus properties or arrival times. However, this would at least require the brain to be able to distinguish between different depths in the auditory and visual domain (Foley, 1980; Knapp, & Loomis, 2003; Kolarik, Moore, Zahorik, Cirstea, & Pardhan, 2016; Zahorik, Brungart, & Bronkhorst, 2005), and a certain degree of plasticity of the TBW. Both have been demonstrated before, but of particular interest are the findings of a narrowing of the TBW after feedback training (Powers, Hillock, & Wallace, 2009), recalibration of the point of subjective simultaneity (Vroomen, Keetels, de Gelder, & Bertelson, 2004), and studies indicating distance-specific impairments in visuospatial attention (Aimola, Schindler, Simone, & Venneri, 2012; Halligan, & Marshall, 1991; Van der Stoep et al., 2013). Whereas the former two findings demonstrate plasticity of multisensory temporal processing resulting from simple (cognitive) feedback or exposure to asynchronous multisensory stimuli, the latter could also be taken to suggest that the brain indeed has distance-specific systems or modes of processing.

If the brain can indeed flexibly alter the TBW width depending on the distance from which we perceive audiovisual events, then one may wonder about how the TBW width changes as a function of distance. This could, for example, be a linear decrease from near to far space, a step function representing two modes of multisensory temporal processing distinguishing between near and far space, or a more gradual decline with increases in distance (i.e. a cumulative gaussian pattern). Investigating the size of the TBW at many distances could provide insight into the mechanisms driving these changes in TBW with increases in distance.

In the same vein, adding a number of distances at which temporal binding windows are indexed, would allow questioning whether changes in multisensory temporal acuity with distance are scaffolded on a representation of peri-personal space. A recent study demonstrated modulations in evoked neural activity at the boundary of peri-personal space, but only for multisensory audio-visual conditions, and not for unisensory auditory or visual conditions (Noel, Serino et al., 2018). This modulation was seen at approximately 300 ms post-stimulus onset, which is considerably later than global field responses encoding the peri-personal space itself, which generally occur around approximately 130 ms (Bernasconi et al., 2018; Noel et al., 2019). Thus, audio-visual multisensory modulations as a function of distance may be scaffolded on earlier audio-tactile or visuo-tactile interactions, that, by definition, preferentially occur in the near space. If audio-visual temporal acuity shows a non-linear profile with respect to distance, then conceivably this representation categorically differentiates between the near and far space, as the peri-personal space representation does (Van der Stoep et al., 2015).

In future studies it will be interesting to explore putative traits that differentiate the group of subjects demonstrating altered temporal windows of integration vs. those with altered peripheral processing times for different distances. One factor that could be considered are the individual differences in unisensory peripheral processing times or the accuracy in determining the distance of perceived audiovisual events.

Acknowledgments

NVDS is supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO, project number 451-17-014). AD was supported by Grant DI 506/15-1 (German Science Foundation/DFG). HC was supported by a grant from German Science Foundation/DFG (SFB/TRR31) and Cluster of Excellence (DFG) “Hearing4all” of Oldenburg University.

Appendix A

See Diederich and Colonius (2015) for more details about applying the TWIN model to response times and temporal order judgments.

ex-Gaussian version of TWIN

Let $F_A(t) = 1 - \exp[\text{-}\lambda_A\, t]$ and $F_V\,(t) = 1 - \exp[\text{-}\lambda_V\, t]$ for $t \geq 0$ and positive parameters λ_A and λ_V.

Computation of Pr(I)

We have to evaluate the two integrals:

$$\begin{aligned} p_1 &= \Pr(A + \tau < V < A + \tau + \omega) \\ &= \int_0^\infty \{F_V(a + \tau + \omega) - F_V(a + \tau)\}\, dF_A(a), \end{aligned}$$

and

$$\begin{aligned} p_2 &= \Pr(V < A + \tau < V + \omega) \\ &= \int_0^\infty \{F_A(v + \omega - \tau) - F_A(v - \tau)\}\, dF_V(v). \end{aligned}$$

For $\tau < 0$, we write

$$p_1^{(-)} = \begin{cases} \dfrac{\lambda_V}{\lambda_V + \lambda_A}\{\exp[\lambda_A(\tau + \omega)] - \exp[\lambda_A \tau]\} & \text{if } \tau + \omega < 0; \\ \dfrac{\lambda_A}{\lambda_V + \lambda_A}\{1 - \exp[-\lambda_V(\tau + \omega)]\} & \\ \quad + \dfrac{\lambda_V}{\lambda_V + \lambda_A}\{1 - \exp[-\lambda_A \tau]\} & \text{if } \tau < 0 < \tau + \omega; \end{cases}$$

and

$$p_2^{(-)} = \frac{\lambda_V}{\lambda_V + \lambda_A}\{\exp[\lambda_A \tau] - \exp[-\lambda_A(\tau - \omega)]\}.$$

For $\tau > 0$,

$$p_1^{(+)} = \frac{\lambda_A}{\lambda_V + \lambda_A}\{\exp[-\lambda_V \tau] - \exp[-\lambda_V(\tau + \omega)]\},$$

and

$$p_2^{(+)} = \begin{cases} \dfrac{\lambda_A}{\lambda_V + \lambda_A}\{\exp[-\lambda_V(\tau - \omega)] - \exp[-\lambda_V \tau]\} & \text{if } \tau > \omega, \\ \dfrac{\lambda_A}{\lambda_V + \lambda_A}\{1 - \exp[-\lambda_V \tau]\} & \\ \quad + \dfrac{\lambda_V}{\lambda_V + \lambda_A}\{1 - \exp[-\lambda_A(\omega - \tau)]\} & \text{if } \tau < \omega, \end{cases}$$

Computation of expected reaction times in the multisensory and uni-modal conditions

We have, from Eq. (2) in the main text,

$$\mathrm{E}[RT_{VA,\tau}] = \mathrm{E}[\min(V, A + \tau)] + \mu - Pr(I) \times \Delta.$$

This becomes, after inserting the exponential distributions in E[min(V, A +τ)],

$$\mathrm{E}[RT_{VA,\tau}] = \frac{1}{\lambda_V} - \exp[-\lambda_V \tau]\left(\frac{1}{\lambda_V} - \frac{1}{\lambda_V + \lambda_A}\right) + \mu - Pr(I)\Delta.$$

For the unimodal conditions, we get

$$\mathrm{E}[RT_V] = \frac{1}{\lambda_V} + \mu \quad \text{and} \quad \mathrm{E}[RT_A] = \frac{1}{\lambda_A} + \mu.$$

Computation of psychometric function $\Psi\,(\tau)$

For the computation of the SJ probabilities under the TWIN model, the cases of $\tau < 0$ and $\tau > 0$ must also be considered separately, denoted here as $\Psi^-\,(\tau)$ and $\Psi^+\,(\tau)$:

$$\begin{aligned} \Psi^-(\tau) &= \beta + (1 - \beta)\Pr(I) \\ &= \beta + (1 - \beta)\,[p_1^{(-)} + p_2^{(-)}] \end{aligned}$$

and

$$\begin{aligned} \Psi^+(\tau) &= \beta + (1 - \beta)\Pr(I) \\ &= \beta + (1 - \beta)\,[p_1^{(+)} + p_2^{(+)}] \end{aligned}$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmp.2020.102443.

References

Aimola, L., Schindler, I., Simone, A. M., & Venneri, A. (2012). Near and far space neglect: task sensitivity and anatomical substrates. *Neuropsychologia*, *50*(6), 1115–1123. http://dx.doi.org/10.1016/j.neuropsychologia.2012.01.022.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. http://dx.doi.org/10.1016/j.cub.2004.01.029.

Alais, D., & Carlile, S. (2005). Synchronizing to real events: subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(6), 2244–2247. http://dx.doi.org/10.1073/pnas.0407034102.

Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research*, *45*(10), 1275–1284. http://dx.doi.org/10.1016/j.visres.2004.11.014.

Bernasconi, F., Noel, J.-P., Park, H. D., Faivre, N., Seeck, M., Spinelli …, L., & Serino, A. (2018). Audio-tactile and peripersonal space processing around the trunk in human parietal and temporal cortex: An intracranial EEG study. *Cerebral Cortex*, *28*(9), 3385–3397. http://dx.doi.org/10.1093/cercor/bhy156.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological Methods & Research*, *33*(2), 261–304. http://dx.doi.org/10.1177/0049124104268644.

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory process*.

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. *Attention, Perception & Psychophysics*, *75*(5), 790–811. http://dx.doi.org/10.3758/s13414-013-0475-4.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*(6), 1000–1009. http://dx.doi.org/10.1162/0898929041502733.

Di Luca, M. (2014). Light source distance affects perceived audiovisual simultaneity. *Procedia - Social and Behavioral Sciences*, *126*, 151. http://dx.doi.org/10.1016/j.sbspro.2014.02.349.

Diederich, A., & Colonius, H. (2004). Modeling the time course of multisensory interaction in manual and saccadic responses. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 395–408). MIT Press, Chap. 24.

Diederich, A., & Colonius, H. (2007). Modeling spatial effects in visualtactile saccadic reaction time. *Perception & Psychophysics*, *69*(1), 56–67.

Diederich, A., & Colonius, H. (2008). Crossmodal interaction in saccadic reaction time: separating multisensory from warning effects in the time window of integration model. *Experimental Brain Research*, *186*(1), 1–22. http://dx.doi.org/10.1007/s00221-007-1197-4.

Diederich, A., & Colonius, H. (2015). The time window of multisensory integration: relating reaction times and judgments of temporal order. *Psychological Review*, *122*(2), 232–241. http://dx.doi.org/10.1037/a0038696.

Diederich, A., Colonius, H., & Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia*, *46*(10), 2556–2562. http://dx.doi.org/10.1016/j.neuropsychologia.2008.03.026.

Engel, G. R., & Dougherty, W. G. (1971). Visual-auditory distance constancy. *Nature*, *234*(5327), 308.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. http://dx.doi.org/10.1038/415429a.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*(5), 411–434. http://dx.doi.org/10.1037/0033-295X.87.5.411.

García-Pérez, M. A., & Alcalá-Quintana, R. (2015). Converging evidence that common timing processes underlie temporal-order and simultaneity judgments: a model-based analysis. *Attention, Perception & Psychophysics*, *77*(5), 1750–1766. http://dx.doi.org/10.3758/s13414-015-0869-6.

Gondan, M., Niederhaus, B., Rösler, F., & Röder, B. (2005). Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Perception & Psychophysics*, *67*(4), 713–726. http://dx.doi.org/10.3758/BF03193527.

Halligan, P. W., & Marshall, J. C. (1991). Left neglect for near but not far space in man. *Nature*, *350*(6318), 498–500. http://dx.doi.org/10.1038/350498a0.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, *63*, 289–293.

Kelley, K., & Lai, K. (2011). Accuracy in parameter estimation for the root mean square error of approximation: sample size planning for narrow confidence intervals. *Multivariate Behavioral Research*, *46*(1), 1–32. http://dx.doi.org/10.1080/00273171.2011.543027.

Knapp, J., & Loomis, J. (2003). Visual perception of egocentric distance in real and virtual environments. In L. Hettinger, & M. Haas (Eds.), *Virtual and adaptive environments: Applications, implications, and human performance issues* (pp. 21–46). CRC Press, http://dx.doi.org/10.1201/9781410608888.pt1.

Kolarik, A. J., Moore, B. C. J., Zahorik, P., Cirstea, S., & Pardhan, S. (2016). Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception & Psychophysics*, *78*(2), 373–395. http://dx.doi.org/10.3758/s13414-015-1015-1.

Kopinska, A., & Harris, L. R. (2004). Simultaneity constancy. *Perception*, *33*(9), 1049–1060. http://dx.doi.org/10.1068/p5169.

Krueger Fister, J., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P., & Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing. *Neuropsychologia*, *88*, 92–100. http://dx.doi.org/10.1016/j.neuropsychologia.2016.02.016.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, *9*(1), 112–147. http://dx.doi.org/10.1137/S1052623496303470.

Leone, L. M., & McCourt, M. E. (2013). The roles of physical and physiological simultaneity in audiovisual multisensory facilitation. *I-Perception*, *4*(4), 213–228. http://dx.doi.org/10.1068/i0532.

Lewald, J., & Guski, R. (2004). Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception. *Neuroscience Letters*, *357*(2), 119–122. http://dx.doi.org/10.1016/j.neulet.2003.12.045.

Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization.

Matlab (2019). *Version 9.7.0.1296695 (R2019b update 4)*. Natick, Massachusetts: The MathWorks Inc.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, *7*(10), 3215–3229.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662.

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, *75*(5), 1843–1857.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. http://dx.doi.org/10.1016/0010-0285(82)90010-X.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, *40*(5), 331–343. http://dx.doi.org/10.3758/BF03203025.

Noel, J.-P., Chatelle, C., Perdikis, S., Jöhr, J., Lopes Da Silva, M., Ryvlin ., P., & Serino, A. (2019). Peri-personal space encoding in patients with disorders of consciousness and cognitive-motor dissociation. *NeuroImage. Clinical*, *24*, Article 101940. http://dx.doi.org/10.1016/j.nicl.2019.101940.

Noel, J.-P., De Niear, M., Van der Burg, E., & Wallace, M. T. (2016). Audiovisual simultaneity judgment and rapid recalibration throughout the lifespan. *Plos One*, *11*(8), Article e0161698. http://dx.doi.org/10.1371/journal.pone.0161698.

Noel, J.-P., Lukowska, M., Wallace, M., & Serino, A. (2016). Multisensory simultaneity judgment and proximity to the body. *Journal of Vision*, *16*(3), 21. http://dx.doi.org/10.1167/16.3.21.

Noel, J.-P., Modi, K., Wallace, M. T., & Van der Stoep, N. (2018). Audiovisual integration in depth: multisensory binding and gain as a function of distance. *Experimental Brain Research*, *236*(7), 1939–1951. http://dx.doi.org/10.1007/s00221-018-5274-7.

Noel, J.-P., Serino, A., & Wallace, M. T. (2018). Increased neural strength and reliability to audiovisual stimuli at the boundary of peripersonal space. *Journal of Cognitive Neuroscience*, 1–18. http://dx.doi.org/10.1162/jocn_a_01334.

Noel, J.-P., & Van der Stoep, N. (2017). *Multisensory RT + SJ*. OSF, http://dx.doi.org/10.17605/OSF.IO/B6HVS, Available at: osf.io/b6hvs.

Noesselt, T., Bergmann, D., Heinze, H.-J., Münte, T., & Spence, C. (2012). Coding of multisensory temporal patterns in human superior temporal sulcus. *Frontiers in Integrative Neuroscience*, *6*(64), http://dx.doi.org/10.3389/fnint.2012.00064.

Odegaard, B., Wozny, D. R., & Shams, L. (2015). Biases in visual, auditory, and audiovisual perception of space. *PLoS Computational Biology*, *11*(12), Article e1004649. http://dx.doi.org/10.1371/journal.pcbi.1004649.

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *The Journal of Neuroscience*, *33*(17), 7463–7474. http://dx.doi.org/10.1523/JNEUROSCI.4678-12.2013.

Parise, C. V., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nature Communications*, *7*(11543), http://dx.doi.org/10.1038/ncomms11543.

Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, *29*(39), 12265–12274. http://dx.doi.org/10.1523/JNEUROSCI.3501-09.2009.

Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychological Bulletin*, *124*(2), 123–164.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590.

Rohde, M., van Dam, L. C. J., & Ernst, M. O. (2016). Statistically optimal multisensory cue integration: a practical tutorial. *Multisensory Research*, *29*(4–5), 279–317. http://dx.doi.org/10.1163/22134808-00002510.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what i am saying? exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147–1153. http://dx.doi.org/10.1093/cercor/bhl024.

Schubert, A.-L., Hagemann, D., Voss, A., & Bergmann, K. (2017). Evaluating the model fit of diffusion models with the root mean square error of approximation. *Journal of Mathematical Psychology*, *77*, 29–45. http://dx.doi.org/10.1016/j.jmp.2016.08.004.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, *1296*, 31–49. http://dx.doi.org/10.1111/nyas.12121.

Spence, C., & Driver, J. (Eds.), (2004). *Crossmodal space and crossmodal attention* (1st ed.). Oxford: Oxford University Press.

Spence, C., Lee, J., & Van der Stoep, N. (2017). Responding to sounds from unseen locations: crossmodal attentional orienting in response to sounds presented from the rear. *The European Journal of Neuroscience*, http://dx.doi.org/10.1111/ejn.13733.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology*, *13*(13), R519–21. http://dx.doi.org/10.1016/S0960-9822(03)00445-7.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience*, *9*(4), 255–266. http://dx.doi.org/10.1038/nrn2331.

Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. *Nature*, *421*(6926), 911. http://dx.doi.org/10.1038/421911a.

Van der Stoep, N., Nijboer, T. C. W., Van der Stigchel, S., & Spence, C. (2015). Multisensory interactions in the depth plane in front and rear space: a review. *Neuropsychologia*, *70*, 335–349. http://dx.doi.org/10.1016/j.neuropsychologia.2014.12.007.

Van der Stoep, N., Serino, A., Farnè, A., Di Luca, M., & Spence, C. (2016). Depth: the forgotten dimension in multisensory research. *Multisensory Research*, *29*, 493–524. http://dx.doi.org/10.1163/22134808-00002525.

Van der Stoep, N., Van der Stigchel, S., Nijboer, T. C. W., & Van der Smagt, M. J. (2016). Audiovisual integration in near and far space: effects of changes in distance and stimulus effectiveness. *Experimental Brain Research*, *234*(5), 1175–1188. http://dx.doi.org/10.1007/s00221-015-4248-2.

Van der Stoep, Nathan, Visser-Meily, J. M. A., Kappelle, L. J., de Kort, P. L. M., Huisman, K. D., Eijsackers ., A. L. H., & Nijboer, T. C. W. (2013). Exploring near and far regions of space: distance-specific visuospatial neglect after stroke. *Journal of Clinical and Experimental Neuropsychology*, *35*(8), 799–811. http://dx.doi.org/10.1080/13803395.2013.824555.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception & Psychophysics*, *72*(4), 871–884. http://dx.doi.org/10.3758/APP.72.4.871.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Research. Cognitive Brain Research*, *22*(1), 32–35. http://dx.doi.org/10.1016/j.cogbrainres.2004.07.003.

Wallace, M. T., & Stein, B. E. (2007). Early experience determines how the senses will interact. *Journal of Neurophysiology*, *97*(1), 921–926. http://dx.doi.org/10.1152/jn.00497.2006.

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present. *Acta Acustica United with Acustica*, *91*(3), 409–420.