

# No Evidence That Sound–Shape Associations Influence Temporal Resolution in Humans: Five Nonreplications of Parise and Spence (2009) and Meta-Analyses

Suddha Sourav<sup>1</sup>, Brigitte Röder<sup>1, 2</sup>, Franka Ambsdorf<sup>1</sup>, Andromachi Melissari<sup>3</sup>,  
Miketa Arvaniti<sup>3</sup>, and Argiro Vatakis<sup>4</sup>

<sup>1</sup> Biological Psychology and Neuropsychology, Institute of Psychology, University of Hamburg

<sup>2</sup> Jasti V Ramanamma Children's Eye Care Center, Child Sight Institute, L V Prasad Eye Institute, Hyderabad, Telangana, India

<sup>3</sup> Department of Philosophy and History of Science, University of Athens

<sup>4</sup> Department of Psychology, School of Social Science, Panteion University of Social and Political Sciences

Sound–shape associations (e.g., preferentially matching angular shapes with high-pitched sounds and smooth shapes with low-pitched ones) have been almost universally observed in humans. If cross-modally congruent sounds and shapes are more robustly integrated in humans, distinguishing them in time might be hypothetically more challenging compared to incongruent sound–shape pairings. Supporting this premise, a highly cited work by Parise and Spence (2009;  $n = 12$ ) reported worse temporal order judgement performance for audiovisual stimuli with congruent compared to incongruent sound–shape associations. Here, we report the results of five experiments across two laboratories, including a preregistered replication attempt, all ( $\sum n = 102$ ) failing to replicate the original results. Additionally, frequentist and Bayesian meta-analyses found no evidence against the null hypothesis, revealing a negligible effect size. The combined results indicate that multisensory temporal resolution in humans is unaffected by sound–shape associations, which might arise at a later (or parallel) processing stage compared to cross-modal temporal order judgements.

## Public Significance Statement

Humans generally associate arbitrary shapes and sounds in a similar way—for example, most people preferentially match the meaningless names “Bouba” and “Kiki” with smooth and spiky shapes, respectively. An influential study by Parise and Spence (2009) reported that sounds and shapes that go well together were more difficult to tell apart in time compared to sounds and shapes that did not match. Does sound–shape association really bind auditory and visual events stronger in time? In five replication experiments across two labs with more than a hundred participants (totaling more than eight times the sample size of the original study), we surprisingly found no evidence for this claim. These results argue that the timekeeping mechanisms that help us distinguish events across the auditory and the visual senses are not influenced by sound–shape associations.

**Keywords:** sound–shape association, multisensory processing, time perception, Bouba/Kiki effect, cross-modal correspondence


**Supplemental materials:** <https://doi.org/10.1037/xge0001641.supp>

Possessing multiple sensory modalities incurs substantial metabolic, biophysical, and developmental costs, offset by crucial advantages conferred to an organism by owning a set of different senses. One immediate benefit is the ability to perceive qualitatively distinct phenomena through unique signal transduction pathways.

A predator's vision, for example, may reveal the location of a prey even if it is silent and downwind, whereas the prey animal might be able to perceive a hidden, windward threat by its smell (Munoz & Blumstein, 2020). Even for events that are perceivable through more than one sensory modality, multisensory processing

This article was published Online First September 19, 2024.

Joseph Toscano served as action editor.

Suddha Sourav  <https://orcid.org/0000-0002-4172-7711>

Brigitte Röder  <https://orcid.org/0000-0003-3088-8023>

Argiro Vatakis  <https://orcid.org/0000-0002-6545-376X>

The results of the study were presented by Suddha Sourav at the 20th International Multisensory Research Forum (2022), Ulm, Germany, and at the 49th Psychology and Brain Conference (2024) in Hamburg, Germany. Simple audiovisual stimuli were used (fully described in the Method

section). The preregistration document (Experiment 3) is available at <https://aspredicted.org/ip5ht.pdf>. The open data and code can be accessed at <https://doi.org/10.25592/uhhfdm.12241>. The authors declare that there are no conflicts of interest related to this work.

The study was funded by Grant DFG Ro 2625/10-1 from the German Research Foundation awarded to Brigitte Röder (Experiments 1, 2, and 3). The authors acknowledge financial support from the Open Access Publication Fund of the University of Hamburg (1704315885-UHH-OAF) awarded to Suddha Sourav. The authors thank Lisa Stockleben,

*continued*

can increase estimation precisions (Murai & Yotsumoto, 2018), accelerate reactions (Gondan et al., 2005; Miller, 1986), and facilitate learning (Shams & Seitz, 2008). In fact, for humans with a typical developmental history, most of the perceived events might be cross-modal in nature: The laugh of a baby can be seen as well as heard, and objects that feel soft to touch usually also look the part.

Given that the same event often informs multiple sensory modalities, it is unsurprising that considerable interactions between the senses have been reported. For example, humans react faster to an audiovisual stimulus combination compared to the auditory or the visual component presented in isolation (Gondan et al., 2005; Miller, 1986). Furthermore, strong expectations about the co-occurrence of certain cross-modal stimuli, likely driven by evolutionary mechanisms as well as environmental statistics during development (Ernst, 2007), can create illusory perceptions of a stimulus in one modality based on the information received from another (Shams et al., 2000). A striking example is the *illusory audiovisual rabbit*, in which three identical beeps and two spatially distinct flashes are presented with their commencements and ends aligned. In this setting, observers often perceive an illusory flash, located spatially and temporally between the two veridical flashes (Stiles et al., 2018).

In addition to these perceptual reports that are often considered lower level effects, multisensory interactions have been reported to exert systematic influence on higher level cognitive processes as well, including linguistic processing. Sound–shape association (SSA), also known as the *Bouba/Kiki effect* (Figure 1a), provides a particularly vivid example. In SSA, humans preferentially associate smoother shapes with pseudowords like *Bouba* or *Maluma*, and in contrast, tend to match angular shapes with pseudowords such as *Kiki* or *Takete*. Multiple auditory features have been identified as factors that influence SSA, including tone frequency, temporal continuity, and spectral components of the sound (Ćwiek et al., 2021; Fort & Schwartz, 2022; O’Boyle & Tarte, 1980). SSA has been near-universally observed across ethnicities and language groups, including in ethnic groups without orthography (Bremner et al., 2013; Ćwiek et al., 2021) and has been reported for visually as well as haptically perceived shapes (Fryer et al., 2014; Hamilton-Fletcher et al., 2018; Sourav et al., 2019). To date, humans are the only great ape species in which SSA has been reported: A systematic association has not been found in a group of chimpanzees and gorillas (Margiotoudi et al., 2019) nor in a language-competent bonobo (Margiotoudi et al., 2022).

How top–down influences (e.g., higher level associations) might modify stimulus-driven bottom–up processing is a field of active research (Choi et al., 2018; Gilbert & Sigman, 2007). In the case of

SSAs, it could be argued that, from natural co-occurrence statistics, over time the human brain learns to treat congruent stimulus combinations (e.g., round shape + lower pitched tone) to be a priori more likely. As a result, the constituent stimuli for congruent SSAs might be more robustly integrated compared to stimulus combinations with incongruent SSAs. If a more robust integration would, in addition, lead to an impeded access to the statistics of the component stimuli, then one would expect to see behavioral performance differences for pairings with congruent versus incongruent SSA in tasks that require access to the component stimulus properties. For example, in temporal processing, the onsets of closely presented crossmodally congruent sound and shapes might be expected to be more difficult to distinguish compared to cross-modally incongruent stimulus combinations (Figure 1b–c).

A highly cited work by Parise and Spence (2009; Experiment 2,  $n = 12$ ) tested this hypothesis by assessing whether SSA influenced performance in a multisensory temporal order judgement (TOJ) task (Parise & Spence, 2009). To this end, auditory and visual stimuli with congruent or incongruent SSA were presented with generally asynchronous stimulus onsets, and participants had to indicate the order in which they perceived them to appear. In this study, a worse performance was observed for cross-modal stimuli with congruent SSA compared to stimuli with incongruent SSA. The authors interpreted the results to indicate that a stronger multisensory coupling, driven by congruent SSAs, might have rendered access to the temporal onsets of the individual stimuli more difficult. This study provided the first evidence that, in humans, SSAs might bias fundamental temporal processes such as the resolution for multisensory event order perception.

In a study that aimed at replicating the SSA effects reported in Parise and Spence (2009), corresponding to the second experiment of the original study, we surprisingly did not find any evidence that SSA modulated performance in a TOJ task ( $n = 34$ ). Here, with four additional experiments, including a preregistered replication attempt (<https://aspredicted.org/ip5ht.pdf>), we report the results of a set of five experiments in total, across two laboratories in Germany and Greece, all failing to replicate the original results ( $\sum n = 102$ ). Frequentist and Bayesian mini meta-analyses (Goh et al., 2016) across these five experiments, employed to more robustly estimate the effect size, likewise found no substantial evidence for SSA effects on TOJ performance and revealed a negligible effect size. To the best of our knowledge, this collection of experiments form the first known replication attempts of the original experiment in the 15 years following its publication, which we think has fundamental implications for interpreting frameworks of multisensory processing.

Hannah Burgdorf, and Florian Süßer for helping with data acquisition.

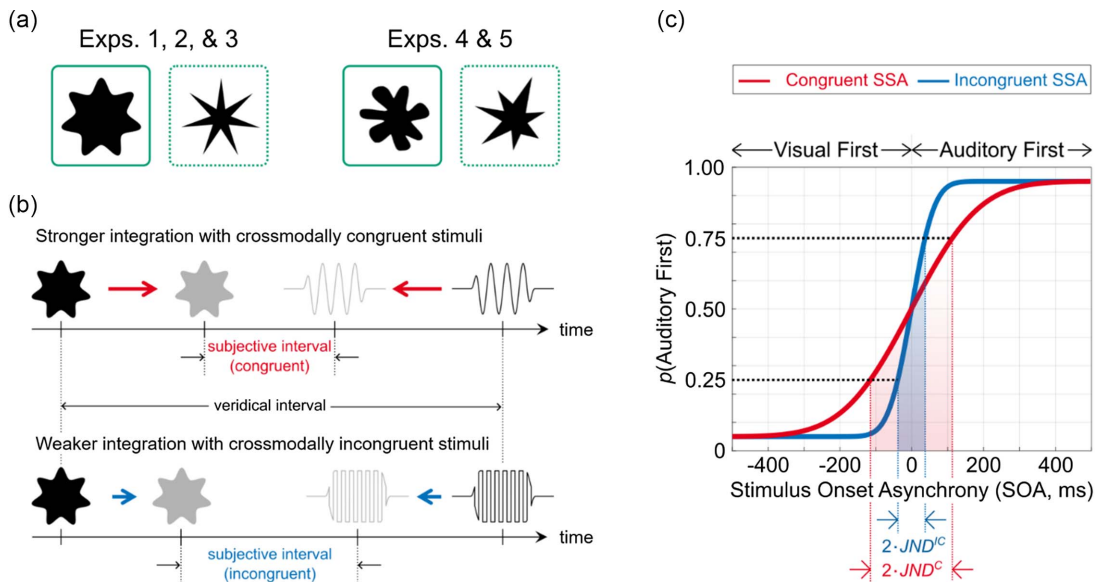
This work is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.0; <https://creativecommons.org/licenses/by-nc-nd/4.0>). This license permits copying and redistributing the work in any medium or format for noncommercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Suddha Sourav played a lead role in data curation, formal analysis, methodology, software, visualization, writing—original draft, and writing—review and editing and an equal role in conceptualization, investigation, project administration, supervision, and validation. Brigitte Röder played a lead role in funding acquisition and resources, a supporting role in formal analysis, investigation, and methodology, and an equal role in conceptualization, project

administration, supervision, and writing—review and editing. Franka Ambsdorf played a supporting role in project administration and writing—review and editing and an equal role in investigation and validation. Andromachi Melissari played a supporting role in writing—review and editing and an equal role in investigation. Miketa Arvaniti played a supporting role in writing—review and editing and an equal role in investigation. Argiro Vatakis played a supporting role in data curation, methodology, and resources and an equal role in conceptualization, investigation, project administration, supervision, and writing—review and editing.

Correspondence concerning this article should be addressed to Suddha Sourav, Biological Psychology and Neuropsychology, Institute of Psychology, University of Hamburg, Von-Melle-Park 11, 20146 Hamburg, Germany. Email: [suddha.sourav@uni-hamburg.de](mailto:suddha.sourav@uni-hamburg.de)

**Figure 1**  
SSA and Its Hypothetical Effect on Multisensory Event Ordering



*Note.* (a) Visual shapes used in the five experiments. Humans generally match the smoother shapes (in solid frames) with pseudowords like “Bouba” and the angular shapes (in dotted frames) with pseudowords like “Kiki.” (b) If asynchronously presented cross-modally congruent sounds and shapes (e.g., smooth shapes and low-frequency tones, upper row) are more strongly integrated, the estimated time of their appearances might be pulled toward each other (indicated in grey). Crossmodally incongruent sounds and shapes (e.g., smoother shapes and high-frequency tones, lower row), in contrast, would be pulled less strongly toward each other in time. (c) If this hypothesis holds, participants would be expected to show worse performance in judging the temporal order of the constituent stimuli for congruent compared to incongruent SSAs. Hypothetical psychometric curves from a temporal order judgment task, which would support the hypothesis are depicted. The *just noticeable difference* (JND), a measure of temporal confusion, is shown, defined as half of the SOA change needed for the performance to shift from the 25% to the 75% level. Exps. = experiment; SSA = sound–shape association; IC = incongruent SSA; C = congruent SSA. See the online article for the color version of this figure.

## Method

### Participants

We preprocessed the data of 127 participants who took part in one of five experiments. The data of 102 participants entered further analysis. The five experiments were conducted across two laboratories: Experiments 1, 2, and 3 were conducted at the University of Hamburg (Germany), whereas Experiments 4 and 5 were conducted at the University of Athens (Greece). In all experiments, the sex of the participants was self-reported. The data sets have heretofore not been published elsewhere.

### Experiment 1 (University of Hamburg, 2019)

Thirty-nine healthy adults took part in the first experiment. Five participants' data were discarded from the analysis as these data yielded no sets of *just noticeable difference* (JND) metric, as defined in the original study (Parise & Spence, 2009; see also the Statistical Analysis section). Of the participants whose data entered further analysis ( $n = 34$ ), 25 were female, and nine were male, with a mean age of 24.7 years (age range = 19–42 years). Thirty were right-handed.

### Experiment 2 (University of Hamburg, 2019)

Nine healthy adults took part. All participants' data entered further analysis. Six of the participants were female, and three were male. The participants had a mean age of 27.6 years (age range = 19–50 years), and all nine participants were right-handed.

### Experiment 3 (University of Hamburg, 2021)

In this preregistered experiment (<https://aspredicted.org/ip5ht.pdf>), 20 healthy adults took part. Two participants' data were discarded from further analysis because they yielded no JND, leaving 18 participants whose data were further processed. These 18 participants were on average 28.6 years old (range = 19–52 years). Four of the participants were male, and 14 were female; 16 were right-handed.

### Experiment 4 (University of Athens, 2012)

Thirty-eight healthy adults took part. We limited our analysis to 32 participants for whom raw data were available. Four participants' data were rejected due to implausible psychometric function fits, or because they yielded no JND sets, leaving 28 participants (seven female, 21 male) for further analyses.

### Experiment 5 (University of Athens, 2016)

Eighty-four participants took part, but due to an archival problem only 27 data sets were accessible. The archival problem was unrelated to the variables in the study, and the participants whose data we could not analyze can be assumed to be missing at random. Out of the available participants, the data of 14 participants were rejected because of psychometric fit nonconvergence, implausible fits, or because the data did not yield a set of JNDs. Of the 13 participants whose data were further analyzed, 12 were female, and one was male (age range = 18–19 years; no handedness data available).

### Ethics

Experiments 1, 2, and 3 were approved by the local ethical commission of the Faculty of Psychology and Movement science at the University of Hamburg. Experiments 4 and 5 were approved by the University of Athens under provisions of Greek law (4521/2018). Participants gave written informed consent and earned either course credits or a small monetary compensation for taking part in the experiments.

All experiments conformed to the ethical guidelines outlined in the [World Medical Association \(2013\)](#).

### Experimental Design

#### Stimuli

In all experiments, participants performed an audiovisual TOJ task. Each trial consisted of an auditory stimulus and a visual stimulus. In the following subsections, we describe the stimuli in detail.

**Experiments 1, 2, and 3.** Visual Stimuli: The shapes were identical to the visual stimuli employed in the second experiment of [Parise and Spence \(2009\)](#) and were drawn using the polygon tools of Inkscape, Version 0.91+ (see [Figure 1a](#), left side). In Experiment 1, the visual stimuli subtended an angle of  $10^\circ$  and in Experiment 2 an angle of  $5.2^\circ$ . In Experiment 3, the visual stimuli could be either  $5.2^\circ$  or  $10.4^\circ$  in size, used to systematically investigate a possible effect of visual stimulus size in the TOJ task (see also the preregistration document accessible at <https://aspredicted.org/ip5ht.pdf>). In Experiments 1 and 2, the stimuli were white on a black background (nominal white luminance:  $250 \text{ cdm}^{-2}$ ) and were presented for a duration of 33 ms. In Experiment 3, the stimuli were black on a white background ( $100 \text{ cdm}^{-2}$ ) and presented for 25 ms.

**Auditory Stimuli:** The low-frequency auditory stimulus was a 400-Hz sine wave tone. The high-frequency auditory stimulus, with additional overtones, was a square wave with a frequency of 1,760 Hz. These values were identical to the pertinent second experiment in [Parise and Spence \(2009\)](#). Likewise, the onsets and offsets of the stimuli had 5-ms linear ramps. The auditory stimuli were presented for 33 ms in Experiments 1 and 2 and for 25 ms in Experiment 3. For each participant, the subjective loudness levels of the auditory stimuli were equalized with the ([Watson & Pelli, 1983](#)) algorithm.

Stimulus timing was confirmed with an SFH—203 photodiode (Osram Opto Semiconductors GmbH, Regensburg, Germany), and a BOB—12758 microphone breakout board (SparkFun electronics, CO, USA), connected to an Arduino Leonardo microcontroller.

**Experiment 4.** Visual Stimuli: The stimuli closely followed those depicted in [Spence \(2011, p. 974\)](#). The shapes are shown in [Figure 1a](#) (right side). The visual stimuli were  $250 \times 250$  pixels, black on a white background, and presented on a 13-in. monitor for 200 ms.

**Auditory Stimuli:** The phonemes /o/ and /i/, articulated by a female voice for 200 ms, served as the auditory stimuli in Experiment 4. Here, /o/ was cross-modally congruent to the smooth shape (i.e., “boubalike”; fundamental frequency,  $F_0 = 194 \text{ Hz}$ , second formant frequency,  $F_2 = 892 \text{ Hz}$ , [Supplemental Material S1](#)), and /i/ was cross-modally congruent to the spiky shape (i.e., “kikilike,”  $F_0 = 221 \text{ Hz}$ ,  $F_2 = 2,841 \text{ Hz}$ ).

**Experiment 5.** Visual Stimuli: The same shapes as in Experiment 4 were utilized. The visual stimuli were  $115 \times 115$  pixels and were presented on a 17-in. screen for 33 ms. Auditory Stimuli: Sine wave stimuli with frequencies of 300 Hz and 4,500 Hz, respectively, served as the lower and higher pitched auditory stimuli. Like the visual stimuli, the auditory stimuli were presented for a duration of 33 ms.

### Procedure

In Experiments 1, 2, 3, and 4, participants sat 65 cm and, in Experiment 5, 60 cm away from the monitor. In each trial, an auditory stimulus and a visual stimulus were presented with different stimulus onset asynchronies (SOAs), and participants had to indicate the modality of the stimulus (i.e., auditory or visual) that was presented first (Experiments 1, 2, 4, and 5), or second (Experiment 3).

Throughout the present article, we follow the convention that positive SOAs indicate trials where the auditory stimulus was leading. In Experiments 1 and 2, stimuli with SOAs of 0,  $\pm 33$ ,  $\pm 67$ ,  $\pm 133$ ,  $\pm 200$ ,  $\pm 267$ ,  $\pm 333$ , and  $\pm 467$  ms were presented. As in [Parise and Spence \(2009\)](#), each unique audiovisual stimulus combination (e.g.,  $\langle V_{\text{spiky}}, A_{\text{high-pitched}} \rangle$ ) was presented 20 times for the synchronous condition (i.e.,  $\text{SOA} = 0$ ) and 10 times for each of the other SOAs.

In Experiment 3, the SOAs employed were 0,  $\pm 67$ ,  $\pm 133$ ,  $\pm 200$ ,  $\pm 267$ ,  $\pm 333$ , and  $\pm 467$  ms. As in Experiments 1 and 2, we presented each unique audiovisual stimulus pairing 20 times for the synchronous condition and 10 times for each of the other SOA values. In Experiments 1, 2, and 3, PsychoPy v. 2+ was used for stimulus presentation ([Peirce et al., 2019](#)).

In Experiment 4, the SOAs were 0,  $\pm 80$ ,  $\pm 100$ ,  $\pm 150$ , and  $\pm 250$  ms. Each unique audiovisual stimulus combination for each SOA value was presented six times in this experiment. Finally, in Experiment 5, 0,  $\pm 83$ ,  $\pm 166$ ,  $\pm 249$ , and  $\pm 332$  ms SOAs were used. Each unique audiovisual stimulus combination was presented 10 times in this experiment for each unique SOA. Experiments 4 and 5 were programmed using Presentation (Neurobehavioral Systems Inc., CA, United States).

### Statistical Analysis

#### Estimating Temporal Order Judgement Performance

Each participant's responses were pooled in one of two conditions:  $\text{SSA}_{\text{congruent}}$  (i.e.,  $\langle V_{\text{spiky}}, A_{\text{high-pitched}} \rangle$  or  $\langle V_{\text{smooth}}, A_{\text{low-pitched}} \rangle$ ) or  $\text{SSA}_{\text{incongruent}}$  (i.e.,  $\langle V_{\text{spiky}}, A_{\text{low-pitched}} \rangle$  or  $\langle V_{\text{smooth}}, A_{\text{high-pitched}} \rangle$ ).



*A<sub>high-pitched</sub>*). In Experiment 3, which featured the additional within-subject factor of *visual stimulus size* (5.2°/10.4°), we pooled the participant responses separately for each visual stimulus size. Additionally in Experiment 3, responses with reactions faster than 100 ms or slower than 3 s were rejected (see preregistered criterion at <https://aspredicted.org/ip5ht.pdf>).

To investigate whether SSA influenced TOJ performance in the experiments, first we modeled how the probability of an *auditory-first* response depended on the SOAs in each participant and condition. Individual psychometric curves were fitted to the responses with congruent and incongruent SSA using four-parameter models fits with the R package *psyphy* for Experiments 1, 2, 3, and 5 (Knoblauch, 2022). For Experiment 4, which had maximum/minimum SOAs of  $\pm 250$  ms, we fitted psychometric functions with two parameters using the same package.

Following the model fitting procedure, we extracted the JND parameter as an indicator of TOJ task performance. Identical to Parise and Spence (2009), the JND was defined as half of the change of SOA necessary for the probability of an auditory-first response to increase from .25 to .75. A lower JND indicates better performance at the TOJ task, as participants require a smaller SOA difference to reliably estimate the order of the stimuli (see Figure 1c). The JND was the dependent variable for the subsequent analysis steps.

### Hypothesis Testing and Mini Meta-Analyses

For each of the five replication experiments, the Wilcoxon signed-ranks test was used to compare the JNDs between the congruent and the incongruent SSA conditions (Parise & Spence, 2009). In Experiment 3, the preregistered replication criterion was a statistically significant JND difference for the smaller visual stimulus size (5.2°), which was equal to the size of the visual stimuli in Parise and Spence (2009). For the sake of completeness, in the Results section, we additionally report the results of the Wilcoxon signed-ranks test for the larger (10.4°) visual stimulus condition.

Furthermore, we employed parametric models to test the same hypothesis and to obtain a better estimate of the effect of SSA on TOJ performance through mini meta-analyses. This decision was driven by two reasons: First, if distributional assumptions are met, parametric tests are usually more powerful than nonparametric tests (Molloy, 2021). Second, parametric tests allowed us to synthesize evidence across the studies using established methods, which provided a tighter estimate of possible SSA effects with higher statistical power (Goh et al., 2016). This approach also emphasizes effect size estimation over significance testing.

The parametric models were lognormal models of the form  $\log_e(\text{JND}) = \alpha + \beta_{\text{SSA}} + \epsilon$  for Experiments 1, 2, 3, and 5. For Experiment 3, the model had the form  $\log_e(\text{JND}) = \alpha + \beta_{\text{SSA}} \times \beta_{V\_SIZE} + \epsilon$ , where  $\beta_{V\_SIZE}$  was the model coefficient for the additional within-subject factor visual stimulus size (5.2°/10.4°). For hypothesis testing, in Experiment 3, the small stimulus size was set as the reference level, but for the meta-analyses, we estimated  $\beta_{\text{SSA}}$  averaged over the visual stimulus sizes by employing a sum contrast for the visual stimulus size factor. Pooling the log-transformed JND values across the experiments, we found no evidence that they were distributed nonnormally, justifying the use of lognormal models: Shapiro–Wilk test, congruent condition,  $W = 0.993$ ,  $p = .837$ ; incongruent condition,  $W = 0.995$ ,  $p = .948$ . For this test of nonnormality, the JND values for the two different visual

stimulus sizes in Experiment 3 were averaged before taking their natural logarithm.

Finally, we ran an inverse-variance random-effects meta-analysis on the  $\beta_{\text{SSA}}$  parameter estimates derived from the lognormal models, using the R package *metafor* v3.8-1 (Viechtbauer, 2010). In inverse-variance meta-analysis models, an estimate with lower variance receives a higher weight for the calculation of the meta-analytic estimate. This frequentist approach was complemented by a Bayesian individual participant data (IPD) meta-analysis with subjects as a random factor, which is considered the gold standard for meta-analyses (Tierney et al., 2015). The IPD meta-analysis was performed with the *brms* package in R with the formula:  $\text{JND} \sim \alpha + \beta_{\text{SSA}} + (1 \mid \text{Study/Subject})$ . The lognormal family was chosen, with weakly informative priors for  $\alpha$  and  $\beta_{\text{SSA}}$  [ $\alpha \sim N(4.605, 2)$ ,  $\beta_{\text{SSA}} \sim N(0, 1)$ ].

### Transparency and Openness

The data and R analysis code for all experiments have been released under a permissible, open license (CC BY 4.0 International) at the research data repository of the University of Hamburg (Sourav et al., 2023; <https://doi.org/10.25592/uhhfdm.12241>). Experiments 1, 2, 4, and 5 were not preregistered. Experiment 3 was preregistered. For this experiment, the preregistration document can be found at <https://aspredicted.org/ip5ht.pdf>.

## Results

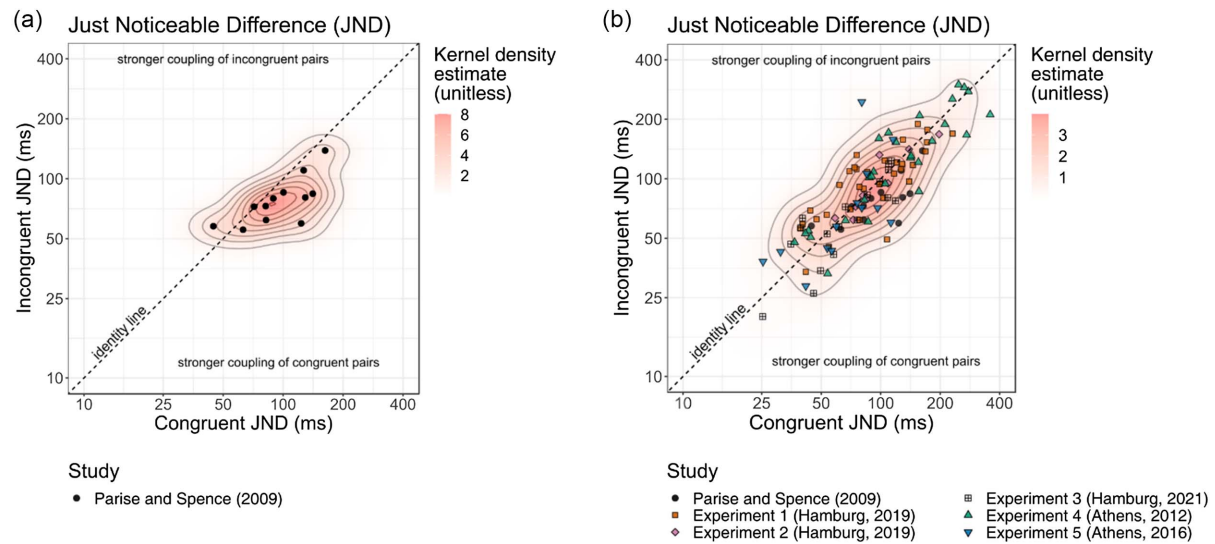
### No Replication of Parise and Spence (2009) With Identical Nonparametric Tests

First, identical to Parise and Spence (2009), using the Wilcoxon signed-ranks test, we nonparametrically examined whether TOJ performance varied between audiovisual stimuli with congruent versus incongruent SSA. The Wilcoxon signed-ranks tests found no evidence against the null hypothesis and, thus, failed to replicate the original results in each of the five experiments, Experiment 1:  $Mdn$  JND difference,  $\Delta_{\text{JND}}$  (interquartile range, IQR) = 2.92 (–13.68, 21.49) ms,  $Z = 0.676$ ,  $p = .499$ ; Experiment 2:  $Mdn\Delta_{\text{JND}}$  (IQR) = –2.27 (–10.63, 2.11) ms,  $Z = -0.681$ ,  $p = .496$ ; Experiment 3, small stimulus size (preregistered replication condition):  $Mdn\Delta_{\text{JND}}$  (IQR) = –0.59 (–16.86, 21.39) ms,  $Z = -0.170$ ,  $p = .865$ , and large stimulus size:  $Mdn\Delta_{\text{JND}}$  (IQR) = –11.78 (–18.56, 0.54) ms,  $Z = -1.835$ ,  $p = .067$ ; Experiment 4,  $Mdn\Delta_{\text{JND}}$  (IQR) = 2.50 (–21.23, 17.81) ms,  $Z = 0.056$ ,  $p = .955$ ; and Experiment 5,  $Mdn\Delta_{\text{JND}}$  (IQR) = –2.21 (–13.11, 9.83) ms,  $Z = -0.068$ ,  $p = .946$  (see also Figure 2).

### No Meta-Analytic Evidence That SSA Modulates TOJ Performance

We ran mini meta-analyses on the  $\beta_{\text{SSA}}$  parameter estimates to synthesize evidence across the experiments and to derive robust estimates of the SSA effect size. Inverse-variance random-effects meta-analysis of the five replication experiments indicated no significant effect of SSA on TOJ performance and revealed a negligible effect size,  $\beta_{\text{SSA}} = -0.004$ , 95% confidence interval (CI) [–0.056, 0.048],  $p = .880$ , corresponding to a –0.397% [–5.424%, 4.896%] change of the JND (% $\Delta_{\text{JND}}$ ) attributable to the effect of SSA (Figure 3a).

**Figure 2**  
SSA and TOJ Performance



*Note.* (a) The original article by Parise and Spence (2009) reported worse TOJ performance for stimuli with cross-modally congruent SSA, as indicated by higher JND values for crossmodally congruent pairs in the log-log scatterplot (stronger coupling of congruent pairs indicated by most of the points lying to the right of the identity line). Data digitized with Webplotdigitizer v4.5 (Rohatgi, 2022). (b) Data of the five experiments reported in the present study, overlaid on the original data of Parise and Spence (2009) for visualization. No systematic shift of JNDs based on SSA was found in any of the present five experiments. Two-dimensional kernel density estimates calculated with normal reference bandwidths are shown on the grid. SSA = sound-shape association; TOJ = temporal order judgement. Adapted from “When Birds of a Feather Flock Together”: Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes.” by C. V. Parise and C. Spence, 2009, *PLOS ONE*, 4(5), p. 4 (<https://doi.org/10.1371/journal.pone.0005664>). CC BY-NC. See the online article for the color version of this figure.

The Bayesian IPD meta-analysis likewise found no support against the null hypothesis and a minuscule effect size, lognormal model, Bayes factor,  $BF_{10} = 0.034$ ,  $\beta_{SSA} = -0.009$  ( $-0.89\% \Delta_{JND}$ ), 95% highest density credible interval (CrI)  $[-0.074, 0.056]$  (Figure 3b). Participants, nested in experiments, served as random factors for this analysis.

In addition, a region of practical equivalence (ROPE) test was performed with a null region defined as  $\pm 5\%$  change of the JND. If 95% or more of the posterior probability would fall within this region, we would claim to have “proved the null hypothesis” (Makowski et al., 2019). Crucially, this Bayesian analysis also generated the posterior probability of SSA effects contributing to a  $< 5\%$  change of JND. The ROPE test indicated that 86.13% of the CrI fell inside the ROPE region, that is, more than 85% probability that SSA modulates JND values by less than 5%.

The same mini meta-analyses were run once again including the data from the original experiment (Parise & Spence, 2009; Rohatgi, 2022). The inclusion of the Parise and Spence (2009) data did not substantially alter the results (Figure 3c–d): random-effects meta-analysis,  $\beta_{SSA} = -0.0352$ , 95% CI  $[-0.106, 0.036]$ ,  $p = .331$ ; Bayesian IPD meta-analysis,  $BF_{10} = 0.045$ ,  $\beta_{SSA} = -0.027$ , corresponding to  $-2.66\% \Delta_{JND}$ , 95% CrI  $[-0.088, 0.033]$ . The ROPE test indicated that 77.33% of the CrI fell inside the ROPE region.

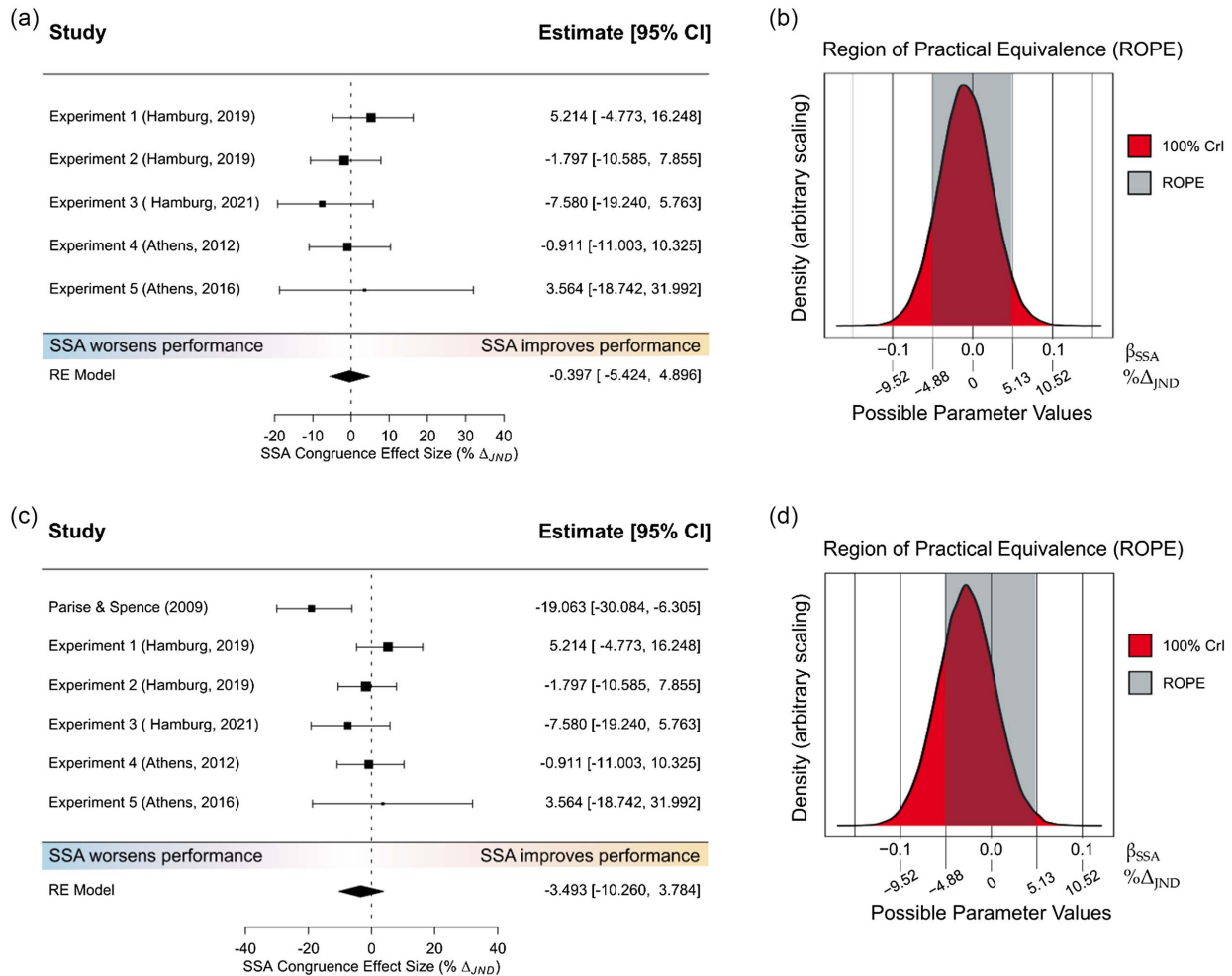
Finally, pooling across all experiments (including the original study), calculation of the Cohen’s  $d$  on the log-transformed JND values indicated a negligible effect size,  $d$  [95% CI] =  $[-0.292, 0.193]$ . A paired-samples  $t$  test would require more than 3,200 participants to detect an effect of this magnitude with

80% power (paired-samples  $t$  test power calculation,  $\alpha = .05$ ; Champely, 2020).

### Parise and Spence (2009) and the Five Replication Experiments’ Data Are Generally Compatible

Based on the observation of a considerable overlap between the JND values from the five replication experiments and the original study (see Figure 2b), in an exploratory analysis, we aimed to test whether the Parise and Spence (2009) JND data significantly differed from the replication experiments’ data. A Gaussian finite mixture model, fit to the five replication experiments’ log-transformed JND data with a variable number of components, indicated that an ellipsoidal multivariate normal model with a single component best described the replication experiments’ data based on a Bayesian information criterion,  $n = 102$ ,  $df = 5$ , Bayesian information criterion = 112.757 (mclust library v6.0.0, running on R v4.2.2). Additionally, the principal components of the replication experiments’ data, which corresponded to the major and the minor axes of this ellipsoidal multivariate normal distribution, were calculated with the prcomp function (Figure 4). This analysis revealed that the first principal component of the data, which explained the maximum JND variance, was almost coincident to the diagonal identity line, a result that is consistent with the equivalent meta-analytic evidence that the effect of SSA on JND was negligible across the five replication experiments. Incidentally, all JND data points of the original study were located inside the 95% percentile boundary derived from the five replication experiments (Figure 3, dotted

**Figure 3**  
*Meta-Analytic Estimates of the Effect of SSA on TOJ Performance*



**Note.** (a) Forest plot for meta-analytic effect of SSA on TOJ performance. An inverse-variance random-effects meta-analysis based on lognormal models was used to derive a synthesized estimate for  $\beta_{SSA}$ . The percent change of the JND metric (% $\Delta_{JND}$ ) attributed to SSA effect by the meta-analysis is shown. The diamond's center shows the meta-analytic estimate, and the edges represent 95% confidence interval (95% CI) of this estimate. (b) Bayesian individual participant data meta-analysis for estimating the effect of SSA on TOJ performance. A single-stage lognormal model with weak priors was run to estimate the effect of SSA on the JND metric. A ROPE test with  $\pm 5\%$  change of JND as the null region (Makowski et al., 2019) was additionally employed. Short of "proving the null" but indicating that SSA effects on JND are likely minuscule, 86.13% of the posterior probability fell within the null region. The full credible interval (100% CrI) is shown along with the ROPE region in grey. (c) and (d) Plots akin to (a) and (b), with the data of Parise and Spence (2009) included. Inclusion of the data of Parise and Spence (2009) did not substantially change the results. CrI = credible interval; SSA = sound–shape association; TOJ = temporal order judgement; JND = just noticeable difference; RE = random effects. Adapted from “When Birds of a Feather Flock Together”: Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes.” by C. V. Parise and C. Spence, 2009, *PLOS ONE*, 4(5), p. 4 (<https://doi.org/10.1371/journal.pone.0005664>). CC BY-NC. See the online article for the color version of this figure.

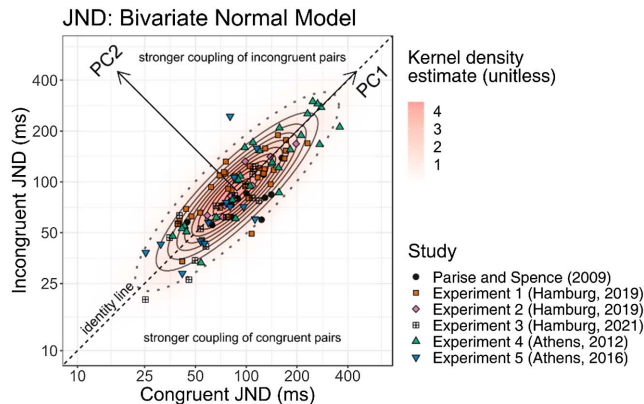
line). Finally, a two-dimensional extension of the two-sample Kolmogorov–Smirnov test found no evidence that the JND values of the original study significantly differed from the JND values of the five replication experiments, Fasano–Franceschini test,  $D = 1.189$ ,  $p = .129$  (Fasano & Franceschini, 1987).

## Discussion

The present study tested whether multisensory interactions driven by SSA influence the temporal resolution for judging stimulus

presentation order, as has been reported by Parise and Spence (Parise & Spence, 2009). In five experiments across two laboratories (Hamburg, Germany and Athens, Greece,  $\sum n = 102$ ), we tried to replicate the original results. In these experiments, participants detected the order of asynchronously presented auditory and visual stimuli that featured either a congruent or an incongruent SSA. The key hypothesis was that a stronger coupling of the auditory and visual stimuli, through congruent SSA, might lead to a less reliable access to the individual onset time estimates. In this case, we would expect to generally replicate a decrease in TOJ performance for

**Figure 4**  
*Bivariate Normal Analysis: Compatibility of Parise and Spence (2009) and the Replication Experiments*



*Note.* Data of the five replication experiments, with the original data of Parise and Spence (2009) overlaid for visualization. A bivariate normal distribution with a single cluster described the replication experiments' data best, based on Bayesian information criterion. Two-dimensional kernel density estimates of the bivariate normal distribution calculated with normal reference bandwidths are shown. The dotted ellipse indicates 95% percentile level of the bivariate normal distribution. JND = just noticeable difference; PC1 = first principal component; PC2 = second principal component. Adapted from “When Birds of a Feather Flock Together”: Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes.” by C. V. Parise and C. Spence, 2009, *PLOS ONE*, 4(5), p. 4 (<https://doi.org/10.1371/journal.pone.0005664>). CC BY-NC. See the online article for the color version of this figure.

audiovisual stimuli with congruent compared to incongruent SSA. Surprisingly, in all five experiments, we were unable to replicate the original finding: In each experiment, the participants' performance was indistinguishable between the congruent and the incongruent SSA condition, indexed by the JND metric as employed in the original study (Parise & Spence, 2009).

Replicability has been one of the key issues of the last decade in psychological science (Ioannidis, 2005; Shrout & Rodgers, 2018). It has been noted that despite low replicability, robust scientific progress is possible if effects of potential interest are swiftly followed by replication attempts (Lewandowsky & Oberauer, 2020). Since single nonreplications do not necessarily exclude the existence of an effect (e.g., several lower powered nonreplications may reveal the presence of a small effect in a meta-analysis), in the present study, we focused on synthesizing SSA effect estimates across the studies.

Using frequentist and Bayesian meta-analytic tools for evidence synthesis, we estimated the SSA effect size across experiments (“Mini meta-analysis,” Goh et al., 2016). These meta-analyses were performed first for the five replication experiments in the present study and once more including the original study of Parise and Spence (2009). The meta-analytic estimates across the five replication studies were minuscule (<1% change of JND), and we found no substantial/significant evidence against the null hypothesis. Including the original study in the meta-analyses did not shift the results substantially, with no significant/substantial evidence against the null. The Bayesian analyses additionally

indicated a posterior probability of more than 77% that SSA effects change the JND metric by less than 5%. The nonreplications along with this higher powered evidence synthesis (with more than eight times the sample size of the original study) indicate that SSA effects on temporal resolution for audiovisual order judgement tasks are likely not substantial.

Based on multivariate analysis methods, we additionally found no evidence that the JND data of Parise and Spence (2009) had a different underlying distribution compared to the five replication experiments in the present study. Gelman and Stern (2006) have pointed out that experiments whose data exhibit a minuscule and nonsignificant difference might nonetheless exhibit highly different *p* values. Thus, even large differences in significance levels while testing for the presence of an effect in two data sets cannot be interpreted as evidence of the data sets being significantly different (“The difference between ‘significant’ and ‘not significant’ is not itself statistically significant”; Gelman & Stern, 2006, p. 328). The results of the multivariate analysis in the present study indicate that, while the *p* values of the original study of Parise and Spence (2009) and the replication studies were located at different sides of the decision boundary of *p* = .05, the data sets themselves are, by and large, compatible with one another.

A related paradigm for testing cross-modal effects on temporal resolution is the *temporal ventriloquist effect*, in which participants perform a visual TOJ task. Task performance is seen to improve if a sound is presented before the first flash and after the second flash. These performance increases are typically observed above baseline conditions in which the sounds are presented simultaneously with the flashes, indicating that sound “pulls apart” the temporal estimates of the visual stimuli (Morein-Zamir et al., 2003). Previously, following up on a separate study by the authors of the original article (Parise & Spence, 2008), two studies have argued that cross-modal associations between sound pitch and visual sizes likely do not modulate the temporal ventriloquist effect (Keetels & Vroomen, 2011; Uno & Yokosawa, 2022). These studies allow a partial extrapolation of the present study's results to cross-modal associations beyond SSA and underscore the importance of multiple experiments from independent groups for ascertaining the presence of promising effects.

In the present work, we focused on investigating the hypothetical effect of SSA on multisensory temporal resolution, that is, Experiment 2 in Parise and Spence (2009). However, Parise and Spence (2009) additionally reported a similar modulation of temporal resolution based on *sound-size correspondences* in their first experiment. In this experiment, congruent sound-size pairings (e.g., a higher pitched sound combined with a smaller visual stimulus) were reported to result in worse TOJ performance. In addition to testing whether SSA modulates multisensory temporal resolution, the availability of differently sized visual stimuli in our preregistered Experiment 3 allowed us to perform a conceptual replication of Experiment 1 of Parise and Spence (2009) as well. We found no evidence that sound-size associations modulated TOJ performance either, akin to the null results for SSA (Supplemental Material S2).

The lack of evidence for SSA's influence on TOJ performance does not exclude the possibility that other temporal behavioral aspects might be modulated by SSA. Indeed, in an implicit association task paradigm, Parise and Spence (2012) have reported slower reaction times for stimulus combinations with incongruent compared to congruent SSA, a result that has subsequently been repeatedly



replicated (Peiffer-Smadja & Cohen, 2019; Silva & Bellini-Leite, 2020). We examined the reaction times for the different SSA conditions in the TOJ task in the preregistered Experiment 3 of the present study, and we found virtually identical reaction times for the different conditions (Supplemental Material S3). Thus, implicit association task and TOJ tasks might recruit distinct neural mechanisms, a hypothesis that is corroborated by reports that cortical electrophysiological markers of cross-modal associations emerge later (250 ms) than markers of temporal order processing differences (150–170 ms; Bien et al., 2012; Simon & Balla, 2020). Thus, we hypothesize that one explanation for the failure of SSA to influence multisensory temporal ordering might be that SSA-related processing occurs later than the decision-critical parts of TOJ processing, that is, after a robust assessment of the multisensory temporal order has already been estimated. Another nonexclusive account might be that TOJ-related multisensory processing predominantly occurs in a parallel (e.g., subcortical) pathway, thus bypassing the hypothetically obligatory binding effects of SSA. Both accounts indicate that at least some aspects of the unimodal stimuli, for example, onset time statistics, are not subjected to the proposed obligatory binding effects due to SSA. Taken together, the generally consistent (non-) replications of SSA effects for individual tasks, yet disparate reports across tasks, indicate that the behavioral effects of cross-modal binding might not be driven by a general mechanism but might instead be task-dependent, as has been reported in the context of Bayesian causal inference (Odegaard & Shams, 2016).

In the present study, we investigated whether SSA influences multisensory temporal resolution, and accordingly, the visual stimuli used in all experiments were simple static shapes that were either smooth or spiky. Using realistic, complex, and dynamic audiovisual events, a group of related studies have investigated how multisensory congruence in general might influence temporal resolution. For example, by combining video clips of human male versus female speakers with spoken phonemes or words, Vatakis and Spence (2007) reported that participants' temporal resolution was worse for congruent (e.g., video of a female speaker combined with a phoneme spoken by a female voice) compared to incongruent audiovisual stimulus combinations. In contrast, in a follow-up study, the authors did not find a similar temporal resolution modulation due to multisensory congruence when nonspeech stimuli (e.g., musical notes plucked on a guitar vs. played on a piano) were employed, leading them to hypothesize that multisensory congruence might selectively modulate temporal resolution for speech stimuli (Vatakis & Spence, 2008). Subsequently, to disentangle the role of speech-specific processing in modulating temporal processing while keeping lower-level stimulus statistics intact, Vroomen and Stekelenburg (2011) used a multisensory TOJ paradigm with sine-wave speech, which are highly simplified synthetic stimuli that can be either perceived as nonspeech or (irreversibly) as speech. In this paradigm, no difference emerged between the multisensory temporal resolutions of participants who perceived the sine-wave stimuli as speech and those who did not, indicating that lower level stimulus properties and not speech-specific processing might explain the previous results. In addition, a more recent study by Chuen and Schutz (2016) has reported that multisensory congruence can modulate temporal resolution for nonspeech stimuli as well, using musical instruments that were more dissimilar (marimba vs. cello) compared to guitar versus piano used in the original study by Vatakis and Spence (2008). The authors argued that the sounds' amplitude envelopes, in addition

to their frequency spectral contents, might be important for driving multisensory-congruence-based temporal processing modulations. In general, it appears that temporal correlations between certain stimulus properties across modalities might play a role in aligning continuous, complex audiovisual stimuli. From this perspective, it also seems unsurprising that the multisensory temporal resolution for audiovisual stimuli without a clearly modifiable temporal correlation between the constituent parts, as in SSA, does not seem to be not susceptible to multisensory congruence effects.

To the best of our knowledge, the experiments in the present study constitute the first replication attempts of Parise and Spence (2009). Given the 15-year gap between the publication of the original study and this replication report, it is not unlikely that other unpublished replication studies might exist ("File drawer problem," Rosenthal, 1979). We provide documented open data and code for performing the meta-analyses in the present study, in the hope that other researchers might find them useful and utilize them to derive narrower estimates for SSA effects on TOJ tasks.

In conclusion, the combined results of the five replication experiments in the present study and the meta-analytic evidence suggest that SSA likely does not modulate temporal resolution for ordering multimodal events. These results strongly suggest that TOJ performance in SSA-based tasks cannot be used as a marker of multisensory binding strength. Furthermore, the current results identify a new constraint for the influence of multisensory processing on behavioral output by demonstrating that TOJ performance is robust to a class of higher level multisensory associations.

## Constraints on Generality

In the present study, we ran five experiments to investigate whether the temporal resolution of audiovisual event ordering can be modulated by SSA in humans, as reported by Parise and Spence (2009). The finding of null results in all replication experiments, and more importantly, a negligible meta-analytic effect size synthesized across the experiments strongly indicate that the initial result is not generalizable. The five experiments in the present study included experiments that aimed to replicate the original study as close as possible (Experiment 3), as well as experiments with slightly different SOA for which a generalization could be expected but was not found (e.g., high- vs. low-pitched spoken vowel phonemes in Experiment 4, presented with an SOA range of  $\pm 250$  ms, instead of purely periodic tones presented with an SOA range of  $\pm 467$  ms). Similar to the original experiment, the replication experiments were conducted in Western European countries (United Kingdom in the original experiment vs. Greece and Germany in the replication experiments). "We have no reason to believe that the results depend on other characteristics of the participants, materials, or context" (Simons et al., 2017, pp. 1125–1126).

## References

- Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, 59(1), 663–672. <https://doi.org/10.1016/j.neuroimage.2011.06.095>
- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to

- Westerners. *Cognition*, 126(2), 165–172. <https://doi.org/10.1016/j.cognition.2012.09.007>
- Champely, S. (2020). *pwr: Basic functions for power analysis (R package)* (Version 1.3-0) [Computer software]. <https://CRAN.R-project.org/package=pwr>
- Choi, I., Lee, J.-Y., & Lee, S.-H. (2018). Bottom-up and top-down modulation of multisensory integration. *Current Opinion in Neurobiology*, 52, 115–122. <https://doi.org/10.1016/j.conb.2018.05.002>
- Chuen, L., & Schutz, M. (2016). The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude envelope cues. *Attention, Perception, & Psychophysics*, 78(5), 1512–1528. <https://doi.org/10.3758/s13414-016-1088-5>
- Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyán, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Űnal-Logacev, Ö., ... Winter, B. (2021). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1841), Article 20200390. <https://doi.org/10.1098/rstb.2020.0390>
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7(5), Article 7. <https://doi.org/10.1167/7.5.7>
- Fasano, G., & Franceschini, A. (1987). A multidimensional version of the Kolmogorov–Smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225(1), 155–170. <https://doi.org/10.1093/mnras/225.1.155>
- Fort, M., & Schwartz, J.-L. (2022). Resolving the bouba-kiki effect enigma by rooting iconic sound symbolism in physical properties of round and spiky objects. *Scientific Reports*, 12(1), Article 19172. <https://doi.org/10.1038/s41598-022-23623-w>
- Fryer, L., Freeman, J., & Pring, L. (2014). Touching words is not enough: How visual experience influences haptic-auditory associations in the “bouba-kiki” effect. *Cognition*, 132(2), 164–173. <https://doi.org/10.1016/j.cognition.2014.03.015>
- Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, 60(4), 328–331. <https://doi.org/10.1198/000313006X152649>
- Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, 54(5), 677–696. <https://doi.org/10.1016/j.neuron.2007.05.019>
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, 10(10), 535–549. <https://doi.org/10.1111/spc3.12267>
- 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. <https://doi.org/10.3758/BF03193527>
- Hamilton-Fletcher, G., Pisanski, K., Reby, D., Stefańczyk, M., Ward, J., & Sorokowska, A. (2018). The role of visual experience in the emergence of cross-modal correspondences. *Cognition*, 175, 114–121. <https://doi.org/10.1016/j.cognition.2018.02.023>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), Article e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Keetels, M., & Vroomen, J. (2011). No effect of synesthetic congruency on temporal ventriloquism. *Attention, Perception, & Psychophysics*, 73(1), 209–218. <https://doi.org/10.3758/s13414-010-0019-0>
- Knoblauch, K. (2022). *psyphy: Functions for analyzing psychophysical data in R (R package)* (Version 0.2-3) [Computer software]. <https://CRAN.R-project.org/package=psyphy>
- Lewandowsky, S., & Oberauer, K. (2020). Low replicability can support robust and efficient science. *Nature Communications*, 11(1), Article 358. <https://doi.org/10.1038/s41467-019-14203-0>
- Makowski, D., Ben-Shachar, M. S., & Lüdtke, D. (2019). BayesestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, 4(40), Article 1541. <https://doi.org/10.21105/joss.01541>
- Margiotoudi, K., Allritz, M., Bohn, M., & Pulvermüller, F. (2019). Sound symbolic congruency detection in humans but not in great apes. *Scientific Reports*, 9(1), Article 12705. <https://doi.org/10.1038/s41598-019-49101-4>
- Margiotoudi, K., Bohn, M., Schwob, N., Tagliatela, J., Pulvermüller, F., Epping, A., Schweller, K., & Allritz, M. (2022). Bo-NO-bouba-kiki: Picture-word mapping but no spontaneous sound symbolic speech-shape mapping in a language trained bonobo. *Proceedings of the Royal Society B: Biological Sciences*, 289(1968), Article 20211717. <https://doi.org/10.1098/rspb.2021.1717>
- Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, 40(5), 331–343. <https://doi.org/10.3758/BF03203025>
- Molloy, M. (2021). Biostatistics and evidence-based medicine. In K. Kleinman, L. McDaniel, & M. Molloy (Eds.), *The Harriet Lane handbook e-book* (22nd ed., pp. 653–661). Elsevier. <https://www.clinicalkey.com/#/content/book/3-s2.0-B9780323674072000292>
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, 17(1), 154–163. [https://doi.org/10.1016/S0926-6410\(03\)00089-2](https://doi.org/10.1016/S0926-6410(03)00089-2)
- Munoz, N. E., & Blumstein, D. T. (2020). Optimal multisensory integration. *Behavioral Ecology*, 31(1), 184–193. <https://doi.org/10.1093/beheco/arz175>
- Murai, Y., & Yotsumoto, Y. (2018). Optimal multisensory integration leads to optimal time estimation. *Scientific Reports*, 8(1), Article 13068. <https://doi.org/10.1038/s41598-018-31468-5>
- O’Boyle, M. W., & Tarte, R. D. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *Journal of Psycholinguistic Research*, 9(6), 535–544. <https://doi.org/10.1007/BF01068115>
- Odegaard, B., & Shams, L. (2016). The brain’s tendency to bind audiovisual signals is stable but not general. *Psychological Science*, 27(4), 583–591. <https://doi.org/10.1177/0956797616628860>
- Parise, C. V., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, 442(3), 257–261. <https://doi.org/10.1016/j.neulet.2008.07.010>
- Parise, C. V., & Spence, C. (2009). ‘When birds of a feather flock together’: Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLOS ONE*, 4(5), Article e5664. <https://doi.org/10.1371/journal.pone.0005664>
- Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, 220(3–4), 319–333. <https://doi.org/10.1007/s00221-012-3140-6>
- Peiffer-Smadja, N., & Cohen, L. (2019). The cerebral bases of the bouba-kiki effect. *NeuroImage*, 186, 679–689. <https://doi.org/10.1016/j.neuroimage.2018.11.033>
- Pearce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Rohatgi, A. (2022). *WebPlotDigitizer* (Version 4.6) [Computer software]. <https://automeris.io/WebPlotDigitizer>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, 408(6814), Article 788. <https://doi.org/10.1038/35048669>
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12(11), 411–417. <https://doi.org/10.1016/j.tics.2008.07.006>

- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, 69(1), 487–510. <https://doi.org/10.1146/annurev-psych-122216-011845>
- Silva, D. M. R., & Bellini-Leite, S. C. (2020). Cross-modal correspondences in sine wave: Speech versus non-speech modes. *Attention, Perception, & Psychophysics*, 82(3), 944–953. <https://doi.org/10.3758/s13414-019-01835-z>
- Simon, J., & Balla, V. R. (2020). Electrophysiological correlates of the spatial temporal order judgment task. *Biological Psychology*, 156, Article 107947. <https://doi.org/10.1016/j.biopsycho.2020.107947>
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. <https://doi.org/10.1177/1745691617708630>
- Sourav, S., Kekunnaya, R., Shareef, I., Banerjee, S., Bottari, D., & Röder, B. (2019). A protracted sensitive period regulates the development of cross-modal sound–shape associations in humans. *Psychological Science*, 30(10), 1473–1482. <https://doi.org/10.1177/0956797619866625>
- Sourav, S., Röder, B., Ambdsdorf, F., Melissari, A., Arvaniti, M., & Vatakis, A. (2023). *Dataset and code for ‘no evidence that sound-shape associations influence temporal resolution in humans: Five non-replications of Parise and Spence (2009) and meta-analyses.* University of Hamburg Research Data Repository. <https://doi.org/10.25592/uhhfdm.12241>
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971–995. <https://doi.org/10.3758/s13414-010-0073-7>
- Stiles, N. R. B., Li, M., Levitan, C. A., Kamitani, Y., & Shimojo, S. (2018). What you saw is what you will hear: Two new illusions with audiovisual postdictive effects. *PLOS ONE*, 13(10), Article e0204217. <https://doi.org/10.1371/journal.pone.0204217>
- Tierney, J. F., Vale, C., Riley, R., Smith, C. T., Stewart, L., Clarke, M., & Rovers, M. (2015). Individual participant data (IPD) meta-analyses of randomised controlled trials: Guidance on their use. *PLOS Medicine*, 12(7), Article e1001855. <https://doi.org/10.1371/journal.pmed.1001855>
- Uno, K., & Yokosawa, K. (2022). Pitch-elevation and pitch-size cross-modal correspondences do not affect temporal ventriloquism. *Attention, Perception, & Psychophysics*, 84(3), 1052–1063. <https://doi.org/10.3758/s13414-022-02455-w>
- Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & Psychophysics*, 69(5), 744–756. <https://doi.org/10.3758/BF03193776>
- Vatakis, A., & Spence, C. (2008). Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, 127(1), 12–23. <https://doi.org/10.1016/j.actpsy.2006.12.002>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Vroomen, J., & Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: Not that special. *Cognition*, 118(1), 75–83. <https://doi.org/10.1016/j.cognition.2010.10.002>
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics*, 33(2), 113–120. <https://doi.org/10.3758/BF03202828>
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194. <https://doi.org/10.1001/jama.2013.281053>

Received June 20, 2023

Revision received May 21, 2024

Accepted June 17, 2024 ■