Multimodal time perception:

The role of temporal ventriloquism in the integration of multisensory intervals of conflicting duration

Dissertation

Der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat)

vorgelegt von
María Dolores de la Rosa Gámiz
aus Castro del Río/Spanien

Tübingen
2019

Tag der mündlichen Qualifikation: 10.05.2019

Dekan: Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter: Dr. Karin Bausenhart

2. Berichterstatter: Prof. Dr. Rolf Ulrich

Table of contents

1. Abstract

Time perception can be subject to severe perceptual distortions when different sensory modalities provide incongruent temporal information. Specifically, perceived duration of visual intervals is typically impaired by concurrently presented auditory intervals with conflicting durations. However, the literature on multisensory duration perception is so far quite scarce, and thus our knowledge of how multisensory intervals with different durations are processed and integrated is limited. The main aim of the present thesis was therefore to investigate the underlying mechanisms and processes as well as the principles that determine the multisensory integration effects on perceived duration in order to reach a better understanding of multisensory duration integration. Importantly, the major results of the present thesis inform about four relevant aspects of multimodal integration of perceived duration. First, combining audiovisual intervals produced distorted perceived duration within a task-relevant modality when accompanied by task-irrelevant intervals with conflicting durations. Nonetheless, this multimodal conflict did not produce costs on discrimination sensitivity but rather led to more stable duration percepts. Moreover, this integration effect was asymmetric, but bidirectional: auditory task-irrelevant interval duration influenced perceived visual duration more than vice versa. Yet, auditory dominance on the integrated percept is not complete, that is, visual duration information also affected auditory perceived duration to a certain extent. Second, the magnitude of this multimodal integration effect remained constant with increasing interval duration, which can be interpreted in terms of a temporal ventriloquist effect determining the closing/opening of an internal switch mechanism. Third, regarding the locus of the underlying integration process, a simple decisional or strategic bias could be ruled out by demonstrating that perceived visual duration was still affected even when the concurrent auditory stimulation did not provide sufficient information about interval duration. However, there was no evidence that auditory duration directly affects a non-temporal aspect of visual perception, namely, identification of masked letters. Finally, the examination of potential sustained effects showed that multisensory duration integration is produced in only a transient, immediate manner. These latter null results, however, cannot rule out a perceptual account of the multisensory duration integration effects. Further investigation is still needed to provide more direct and compelling evidence on the locus of the processes underlying multisensory integration effects on perceived duration.

2. Zusammenfassung

Unsere Wahrnehmung von zeitlicher Information kann stark verzerrt werden, wenn verschiedene Sinne uns inkongruente zeitliche Information über das selbe Objekt liefern. Insbesondere die visuelle Dauerdiskrimination wird typischerweise durch gleichzeitig dargebotene auditive Intervalle mit inkongruenter Dauer verschlechtert. Obwohl in den letzten Jahren zunehmend Forschung zur multisensorischen Wahrnehmung im Allgemeinen stattgefunden hat, gibt es insbesondere zur Integration von multisensorischen Intervallen mit inkongruenten Dauern noch relativ wenig Evidenz. Deshalb ist das grundlegende Ziel dieser Dissertation eine Untersuchung der Prozesse, Mechanismen und zugrundeliegenden Prinzipien der multisensorischen Integration von Reizdauern, um ein besseres Verständnis dieses Aspektes der Zeitwahrnehmung zu erreichen. Die wichtigsten Ergebnisse der vorliegenden Arbeit informieren über vier Hauptaspekte der multisensorischen Integration von wahrgenommener Dauer. Erstens, die Integration audiovisueller Intervalle resultierte in einer verzerrten wahrgenommenen Dauer in einer aufgabenrelevanten Modalität, wenn in einer aufgabenirrelevanten Modalität gleichzeitig ein Intervall mit einer inkongruenten Dauer dargeboten wurde. Trotzdem verursachte dieser multisensorische Konflikt keine Kosten für die Genauigkeit der Diskrimination, sondern resultierte in einer stabilen Diskriminationsleistung. Außerdem war dieser Integrationseffekt zwar asymmetrisch, aber bidirektional: die Dauer eines auditiven aufgabeirrelevanten Intervalls beeinflusste die visuelle wahrgenommene Dauer stärker als umgekehrt; allerdings war diese auditive Dominanz auf das integrierte Perzept nicht komplett. Das heißt, visuelle inkongruente Dauerinformation beeinflusste auch die auditive Dauerwahrnehmung zu einem bestimmten Anteil. Zweitens, das Ausmaß des multisensorischen Integrationseffektes blieb über ansteigende Intervalldauern hinweg konstant. Dies kann gemäß eines zeitlichen Bauchredeneffektes interpretiert werden, der das Schließen und Öffnen eines internen Switch-Mechanismus eines Schrittmacher-Akkumulator-Modelles reguliert. Drittens, in Bezug auf den Wirkungsort des Integrationsprozesses kann eine relativ späte, entscheidungsbasierte oder strategische Verzerrung ausgeschlossen werden, da die visuelle wahrgenommene Dauer auch dann noch beeinflusst wird, wenn die gleichzeitige auditive Stimulation keine ausreichende Information über die Intervalldauer liefert. Es zeigten sich andererseits aber keinerlei Hinweise darauf, dass die auditive Dauer direkt nicht-zeitliche Aspekte der visuellen Wahrnehmung, nämlich die Identifikation von maskierten Buchstaben, beeinflusst. Zuletzt zeigte die Untersuchung von potenziellen langfristigen Integrationseffekten, dass die multisensorische Dauerintegration nur transient und unverzüglich auftritt und somit also nicht zu anhaltenden Veränderungen der wahrgenommenen Dauer führt. Weitere Forschung ist notwendig, um direktere Evidenz für die Lokalisation der multisensorischen Integration von Zeitdauern in der Verarbeitungskette zu erhalten.

3. List of publications

1. Bausenhart, K. M., De la Rosa, M. D., & Ulrich, R. (2014). Multimodal integration of time: Visual and auditory contributions to perceived duration and sensitivity. *Experimental Psychology*, *61*, 310–322. https://doi.org/10.1027/1618-3169/a000249

2. De la Rosa, M. D., & Bausenhart, K. M. (2013). Multimodal integration of interval duration: Temporal ventriloquism or changes in pacemaker rate? *Timing and Time Perception*, *1*, 189–215. https://doi.org/10.1163/22134468-00002015

3. De la Rosa, M. D., & Bausenhart, K. M. (2018). Enhancement of letter identification by concurrent auditory stimuli of varying duration. *Acta Psychologica*, *190*, 38–52. https://doi.org/10.1016/j.actpsy.2018.07.001

4. De la Rosa, M. D., & Bausenhart, K. M. (2018). Still no evidence for sustained effects of multisensory integration of duration. *Multisensory Research*, *31*, 601–622. https://doi.org/10.1163/22134808-18001296

4. Personal contribution

Nr.	Accepted publication yes/no	List of authors	Candidate position	Scientific ideas by the candidate (%)	Data generation by the candidate (%)	Analysis and Interpretation by the candidate (%)	Paper writing done by the candidate (%)
1	Yes	Bausenhart, K. M. **De la Rosa, M. D.** Ulrich, R.	2	20%	20%	25%	25%
2	Yes	**De la Rosa, M. D.** Bausenhart, K. M.	1	60%	66%	70%	75%
3	Yes	**De la Rosa, M. D.** Bausenhart, K. M.	1	50%	80%	80%	80%
4	Yes	**De la Rosa, M. D.** Bausenhart, K. M.	1	60%	90%	90%	70%

5. Introduction

Time is possibly one of the most intriguing phenomena in the scientific world and thus subject to investigation from different areas of knowledge. Physicists, psychologists, and philosophers have long tried to understand time from their different perspectives. From physics, for example, it is claimed that there is no mathematical support to the existence of an arrow of time as we perceive it, that is, with one-way direction from past to future. Furthermore, the fundamental physical laws formulated from Newton, going through Einstein, to more modern physicists show that past and future are completely symmetrical (see, e.g., Greene, 2004). Indeed, Albert Einstein even came to claim that *"the distinction between past, present and future is only a stubbornly persistent illusion"*(extract from a letter to Michele Besso´s family in 1955, cited from Einstein, Speziali, & Besso, 1979, p. 538). An interesting psychological fact about time is that even without any specific sensory organ exclusively dedicated to the perception of time, human beings still have the fundamental ability to perceive the passage of time, to experience simultaneity and succession, and to estimate durations (Le Poidevin, 2000). Therefore, taking all these particularities about time into account, since many decades investigating how we perceive time constitutes a fundamental issue for psychologists, cognitive linguists, and neuroscientists.

In the physical world, there is no adequate stimulus for time, that is, temporal information is not present in isolation. Rather it emerges from stimulation surrounding us in a variety of modalities, which are perceived via the corresponding sensory systems. Often, temporal information is provided by stimulation in two or more modalities simultaneously, and thus processed in initially distinct sensory pathways. Therefore, temporal information provided by different sensory modalities has to be combined to achieve a coherent and precise temporal estimation of the events surrounding us. This emphasizes the importance of understanding the mechanisms and principles of how multisensory temporal information is integrated. In general, the main aim of the present thesis was to investigate how the perceived duration of concurrently presented auditory and visual intervals is integrated when these interval modalities provide congruent or incongruent duration information. Interestingly, empirical evidence regarding the multisensory perception of duration is rather scarce, as compared to the vast number of studies dedicated to investigate multisensory integration within the domains of spatial processing (e.g., localization and perception of spatial correspondence) and temporal processing of time points (e.g., temporal order or simultaneity). Therefore, the following section starts with a general overview of existing multisensory integration research in these domains by presenting well-known multisensory illusions and their determining cognitive principles. Furthermore, I describe the most important psychological models to explain multisensory information integration and existing evidence regarding these potential processes and mechanisms underlying the described multisensory illusions. In the second part of this Introduction, I will relate these general findings to the domain of duration perception. I will review existing empirical evidence regarding multisensory integration of duration information, and on this background I will outline the research gaps which form the core of the present thesis.

5.1. Multisensory integration and multisensory illusions

In everyday life, information emerging from different sensory modalities needs to be constantly integrated within our nervous system, in order to create a coherent perceptual impression of reality (Calvert, Spence, & Stein, 2004). However, under certain conditions of stimulation such multisensory integration may lead to erroneous binding of non-corresponding information, and thus produce a distorted multimodal percept. Such perceptual distortions, also termed *multisensory illusions*, can be employed as an important tool to investigate the perceptual processes and cognitive mechanisms underlying human multisensory perception. In fact, such illusions provide crucial evidence regarding how different modalities with slightly conflicting information interact in order to reduce perceptual discrepancies in the resulting percept (Calvert et al., 2004; L. Chen & Vroomen, 2013). Different multisensory illusions have been observed, depending on the stimulus properties affected by the conflicting multimodal information, for example within the domain of speech perception, the perception of spatial location, or temporal perception. A classic example of a multisensory illusion is the McGurk effect in audiovisual speech perception (McGurk & Macdonald, 1976). According to this illusory effect, when a muted video shows a person repeatedly pronouncing the syllable /ga-ga/ and, simultaneously, the syllable /ba-ba/ is repeatedly played through a loudspeaker, visual and auditory conflicting information interact and affect the resulting speech percept. Specifically, participants typically report to perceive a third, qualitatively different syllable, as for example, /da-da/. A prominent example within the spatial dimension it is well-known *spatial ventriloquist* effect on perceived location (Alais & Burr, 2004; Bertelson, 1999; Bertelson & Aschersleben, 1998; Bertelson & Radeau, 1981; Bertelson, Vroomen, de Gelder, & Driver, 2000; Radeau & Bertelson, 1987). According to this effect, when a visual and an auditory stimulus are presented simultaneously but at slightly different locations, the auditory stimulus is typically perceived as emerging from the visual stimulus location (Figure 1A). An analogous phenomenon can be also observed regarding the temporal characteristics of a stimulus: the *temporal ventriloquist* effect (Bertelson & Aschersleben, 2003; Keetels, Stekelenburg, & Vroomen, 2007; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Vroomen & de Gelder, 2004). Accordingly, when the onsets of a visual and an auditory stimulus are slightly asynchronous, the perceived visual onset is typically shifted in time towards the perceived auditory onset (Figure 1B). These illusions demonstrate multisensory interactions within the perceptual system which may produce coherent perception of a combined multisensory event despite the unimodal signals provide slightly inconsistent information.

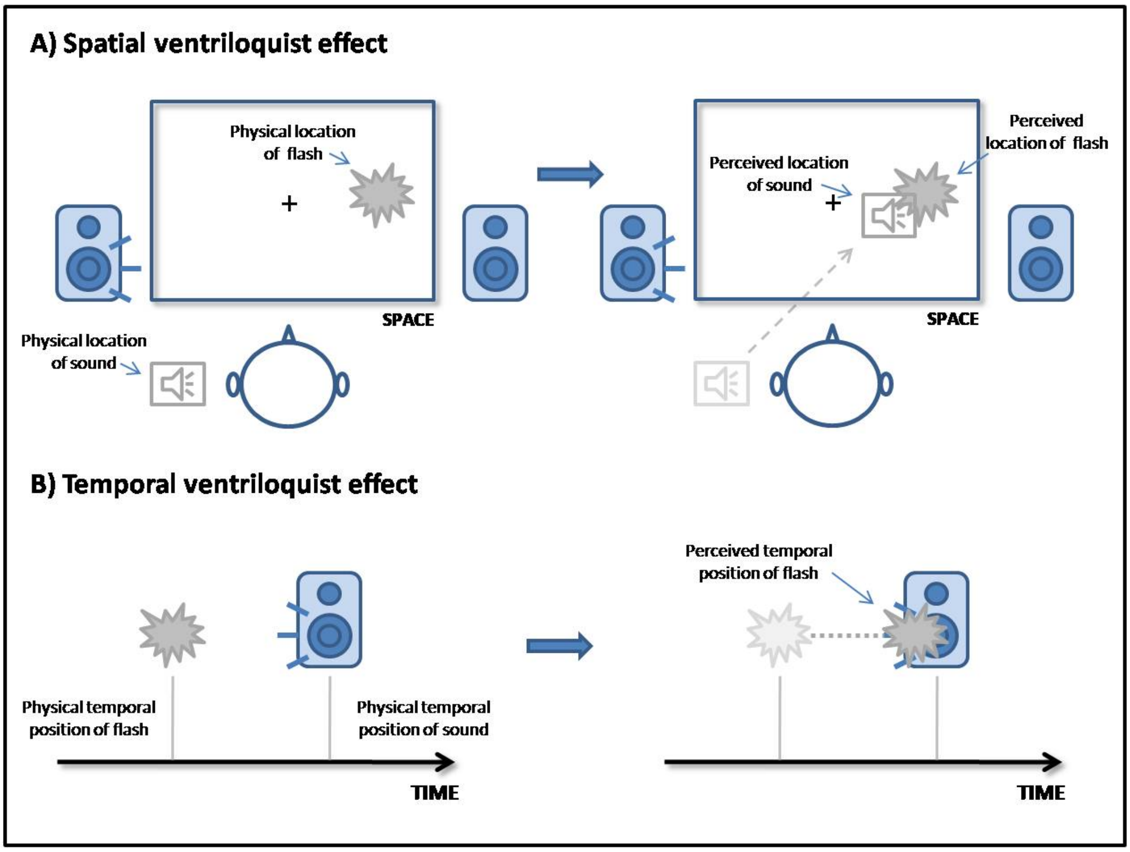

Figure 1. A) Spatial ventriloquist effect: in the spatial domain, if a visual stimulus is presented concurrently with an auditory stimulus, but at a different spatial location, the visual stimulus attracts the auditory stimulus leading to a distorted auditory location perception: the auditory stimulus is perceived as emerging from same location as the visual stimulus. B) Temporal ventriloquist effect: the temporal onset of a visual stimulus is attracted towards the temporal onset of an auditory stimulus.

5.1.1. Prerequisites for multisensory integration: Temporal correspondence, spatial correspondence, and the assumption of unity

The mere co-presence of conflicting information alone, however, is not sufficient to evoke such illusory effects. Certain perceptual as well as even more cognitive factors are also necessary to elicit multisensory integration (Calvert et al., 2004; L. Chen & Vroomen, 2013; Y. C. Chen & Spence, 2017; Spence, 2007). Specifically, perceptual (or “low-level”) factors for multisensory integration refer to the spatial and temporal criteria under which multisensory integration is effective. That is, multisensory integration typically requires that the modalities’ inputs occur at approximately the same location (cf. the ‘spatial rule’, see below) and approximately at the same time (cf. the “temporal rule”, Stein & Meredith, 1993). Specifically, multisensory illusions typically permit a certain range of temporal or spatial discrepancies between modalities. However, when these discrepancies get too large, multisensory integration is no longer observed. This led to the notion of a “temporal window of integration” within which multisensory illusions are likely to be evoked (Asaoka & Gyoba, 2016; Klink, Montijn, & van Wezel, 2011; van Wassenhove, Grant, & Poeppel, 2007). Likewise, evidence also points to the existence of an analogous integration window in the spatial dimension. This spatial rule has been argued to be an important prerequisite for multisensory integration in spatial tasks, whereas its role for multisensory integration in temporal tasks has been recently questioned (Spence, 2013). For example, the spatial ventriloquist effect is typically observed when the different modalities emerge within a temporal window between -100 (sound presented first) and +300 (light presented first) ms, while the spatial conflict between modalities should not exceed ~15° (Lewald, Ehrenstein, & Guski, 2001; Slutsky & Recanzone, 2001). Emergence of the temporal ventriloquist effect is constrained to a temporal window of ~200 ms around simultaneity of the modalities’ inputs (Asaoka & Gyoba, 2016; Klink et al., 2011; Morein-Zamir et al., 2003), however, no strict spatial limitations seem to be in play (Spence, 2013; Vroomen & Keetels, 2006).

Importantly, these spatial and temporal constraints are also considered to contribute to a higher-level factor for multisensory integration known as “assumption of unity” (Y. C. Chen & Spence, 2017; Spence, 2007; Vatakis & Spence, 2007; Welch & Warren, 1980). This assumption supposes the explicit cognitive notion that the information from different sensory sources belongs to a single object. For example, it is more likely to suppose that a visual and an auditory stimulus emerge from same object, if the conflict between these two modalities is constrained to a limited temporal and/or spatial window. Importantly, the cognitive assumption of unity is commonly assumed to be a crucial prerequisite for multisensory integration (Y. C. Chen & Spence, 2017; Spence, 2007).

5.1.2. How are multiple sensory inputs combined?

When the conditions outlined above are met, the modalities’ inputs are combined during stimulus processing. Different models have been proposed to account for how the contribution of each modality to the combined multimodal percept is determined. One classical model of multimodal integration is known as *modality appropriateness* (Welch & Warren, 1980). According to this model, certain modalities provide the most precise information about certain attributes or features of a stimulus, and therefore, the information conveyed within these modalities will dominate over information from other modalities. Consequently, the multisensory percept of a stimulus is determined by the most ‘appropriate’ sensory input. For the case of spatial ventriloquism, for example, it has been suggested that the visual modality provides the most accurate information about spatial localization and thus dominates over spatial information from other modalities as, for example, audition (Welch, 1999; Welch & Warren, 1986). Similarly, for the case of temporal ventriloquism, it has been suggested that audition typically conveys temporal information more precisely than vision, and thus, auditory temporal information consistently dominates over visual temporal information (Bertelson & Aschersleben, 2003; Getzmann, 2007; Morein-Zamir et al., 2003).

A more flexible model has been elaborated more recently based on Bayesian inference laws to account for multisensory integration. According to this approach, multisensory information is integrated following a statistically optimal cue combination process. Specifically, rather than one modality completely dominating over the other one, each modalities’ contribution is determined by the sensory reliability of its current input (Alais & Burr, 2004; Ernst & Banks, 2002; Hartcher-O’Brien, Di Luca, & Ernst, 2014). The sensory signal of each modality is weighted by its inverse variance, and thus, the more reliable a given modality (the less variant its signal), the more it contributes to the combined multisensory percept. This cue combination is considered as statistically optimal since it minimizes the perceptual variance of the combined percept. Evidence for this cue combination rule has been provided in different domains such as audiovisual location perception (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003), visuo-haptic perception (Ernst & Banks, 2002) as well as time perception (Hartcher-O’Brien et al., 2014; Shi, Chen, & Müller, 2010).

5.1.3. At which level of processing are multiple sensory inputs combined?

Based on the latter, statistically optimal cue-combination account, it seems unsurprising that several studies report modulations of perceptual processing for multisensory compared to unimodal stimuli (Y. C. Chen & Spence, 2011; Frassinetti, Bolognini, & Làdavas, 2002; Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008; Shams, Kamitani, & Shimojo, 2002; Stein, London, Wilkinson, & Price, 1996; Vroomen & de Gelder, 2000). At this point, however, the question arises at which level of processing the audiovisual interactions underlying the aforementioned multisensory illusions take place: do these illusory effects induce ‘true’ perceptual changes within early, perceptual stages of processing or do they result from decisional biases and thus reflect information combination within later processing stages?

Based on previous empirical data, however, these two alternatives are often hard to distinguish because in many studies the perceptual and response bias are coupled. This can be easily illustrated considering the following exemplary multisensory illusions. For example, Stein et al. (1996) reported that lights are perceived as being brighter when accompanied by a pulse of white noise than when presented alone. Another prominent example is the double flash illusion (Shams, Kamitani, & Shimojo, 2000; Shams et al., 2002) according to which participants report to perceive two consecutive flashes when two sounds were concurrently presented even though the flash was always continuous. Both illusions seem to suggest either quantitative (enhanced brightness) or qualitative (number of flashes) changes in visual perception caused by accompanying auditory stimulation. But one might critically note that these effects may also reflect that, rather than perceiving the stimuli differently, participants might just base their decisions on the information available in the auditory modality. This latter conclusion was reached by Odgaard, Arieh, and Marks (2003) regarding the sound-induced enhancement reported by Stein et al. (1996). Specifically, Odgaard et al. observed that this crossmodal effect disappeared with a low proportion of sound presentation. On the contrary, evidence for a perceptual bias was provided by Shams et al. (2002) regarding the aforementioned double-flash illusion. Interestingly, when only a single sound was concurrently presented with multiple flashes, participants still reported seeing various flashes, instead of only one flash. This means that responses were not deliberately based on the auditory information, which might readily rules out a response bias account of the double-flash illusion.

Even more decisive evidence for perceptual multisensory interactions can be obtained when the required responses are completely disentangled from multisensory biasing information, that is, when the information provided in one modality is completely uninformative regarding the experimental task in another, task-relevant modality. Such an approach was taken by studies which have shown increased perceptual sensitivity for visual stimuli when accompanied by uninformative sounds. A prominent example of this experimental approach is the study of Y. C. Chen and Spence (2011). These authors observed that identification of a masked target letter was clearly improved when a sound was concurrently presented compared to when the target letter was presented alone. Since the employed sounds did not provide any information about the letter identity, the observed sound-induced enhancement cannot be attributed to a decisional, later stage of processing.

While the study cited above demonstrate that the mere presence (vs. absence) of a sound may modulate visual perception, another line of evidence shows that specific *temporal auditory* information can modulate perception of certain *non-temporal visual* target features (de Haas, Cecere, Cullen, Driver, & Romei, 2013; Morein-Zamir et al., 2003; Shams et al., 2000, 2002; Vroomen & Keetels, 2009). Morein-Zamir et al. (2003), for example, conducted a temporal order judgment (TOJ) task in which participants had to judge which of two brief visual stimuli, one presented above and one below fixation, was presented first. These two visual stimuli were separated by a random SOA. In addition, two brief sounds were presented from a central position, either simultaneously with the two visual stimuli, or separated by somewhat longer SOAs than the visual stimuli, such that sound preceded the first visual stimulus, and trailed the second visual stimulus. Importantly, the results showed enhanced performance in discriminating temporal order of the visual stimuli when these were preceded and trailed by sounds compared to performance in the simultaneous condition. Moreover, presenting two sounds intervening between the two visual stimuli led to a decreased temporal order judgment performance. These findings were interpreted in terms of a temporal ventriloquist effect by which the perceived occurrence of the lights was pulled in time towards the sound occurrence, which therefore facilitated temporal order discrimination. Crucially, the sounds, which were always presented centrally, did not provide any spatial information about the location of the visual stimuli, which was necessary for performing the temporal order task. This led the authors to suggest that the audiovisual interactions underlying the observed temporal ventriloquist effect are unlikely to emerge in post-perceptual processing, as for example, in form of a response bias.

This behavioural evidence in favour of a perceptual account of multimodal integration is well in line with neurophysiological evidence regarding how sensory-specific brain areas are sensitive to information from other modalities at presumably rather early stages of processing (Colin, Radeau, Soquet, Dachy, & Deltenre, 2002; Colin, Radeau, Soquet, Demolin, et al., 2002; Getzmann & Lewald, 2014; Kayser, Philiastides, & Kayser, 2017; Shams, Kamitani, Thompson, & Shimojo, 2001; Starke, Ball, Heinze, & Noesselt, 2017; Stekelenburg, Vroomen, & de Gelder, 2004). Interestingly, Stekelenburg et al. (2004), for example, showed that the illusory shift in sound location induced by a physical location shift of a simultaneously presented light (that is, a spatial ventriloquist effect) evoked an auditory mismatch negativity (MMN; Näätänen & Alho, 1995). This component, which is related to underlying automatic and preattentive processes, was similar to an MMN evoked by a physical shift in sound location. ERPs measures were also employed by Shams et al. (2001) to investigate the double-flash illusion describe above. These authors observed that in illusion trials, the presentation of two sounds concurrently with a single flash produced visual evoked potentials similar to those emerging after presentation of two physical flashes. In sum, these studies provide evidence on early crossmodal influences in areas and components traditionally considered as unimodal, which thus indicates that different modalities are already integrated at relatively early processing stages. This argues against the notion that multisensory integration effects are exclusively based on decisional or strategic biases originating in later processing stages

5.1.4. Sustained effects of multisensory integration

So far, I have described existing evidence on immediate effects of combining multisensory information (i.e., how is a visual stimulus perceived when a sound is concurrently presented?). Yet, integration effects also have been shown to persist over time. Specifically, prolonged exposure to a constant multisensory temporal conflict can result in aftereffects on subsequent judgments of temporal order and/or simultaneity (L. Chen & Vroomen, 2013; Vroomen & Keetels, 2010). Fujisaki, Shimojo, Kashino, and Nishida (2004) and Vroomen, Keetels, de Gelder, and Bertelson (2004) showed first evidence on such sustained multisensory integration effects regarding the temporal ventriloquist effect. For example, Vroomen et al. (2004) employed a recalibration task, in which participants were repeatedly exposed to a brief audiovisual stimulus pair with a consistent temporal conflict between the modalities (e.g., auditory stimuli leading visual stimuli by 100 ms). After this exposure phase, participants had to perform a temporal order judgment task in which they indicated which of two subsequent unimodal stimuli, a sound or a light, was presented first. The results showed a shift of the point of subjective simultaneity towards the direction of the conflict presented during the exposure phase. For example, following exposure to auditory leading stimuli, a subsequently presented unimodal auditory test stimulus had to be presented earlier than a visual one in order to be perceived as simultaneous. Following publication of Vroomen et al. (2004)´s results, a large number of studies have consistently demonstrated such flexibility of perception to recalibrate to persistent multisensory conflicting information in the temporal domain (Di Luca, Machulla, & Ernst, 2009; Fujisaki et al., 2004; Hanson, Heron, & Whitaker, 2008; Harrar & Harris, 2008; Heron, Roach, Whitaker, & Hanson, 2010; Heron, Whitaker, McGraw, & Horoshenkov, 2007; Keetels & Vroomen, 2007; Machulla, Di Luca, Froehlich, & Ernst, 2012; Vatakis, Navarra, Soto-Faraco, & Spence, 2007) as well as the spatial domain (Bertelson, Frissen, Vroomen, & de Gelder, 2006; Frissen, Vroomen, de Gelder, & Bertelson, 2003, 2005; Lewald, 2002; Recanzone, 1998). Again, such recalibration effects cannot be easily reconciled with the notion that multisensory illusions merely reflect responses being biased towards the information from a certain modality. Instead, these prolonged consequences of exposure to multimodal conflict point towards genuine interactions in the processing of multimodal information.

5.2. Perceived duration of uni- and multimodal stimuli

Up to here, I have outlined ample evidence regarding multisensory integration in the spatial as well as the temporal dimension. Yet, within the temporal dimension, most studies have focused on perceived simultaneity or temporal order, that is, on the perception of succession. The main aim of the present thesis is thus the study of multisensory integration in the domain of *duration* perception, on which the literature is much scarcer. In the next section, I will briefly describe a prominent psychological model of duration perception, the pacemaker-accumulator model. Afterwards, I describe existing evidence on the perception of intra-, inter- and multimodal temporal intervals, and how these findings relate to the aforementioned theoretical account of duration perception. By identifying research gaps within this experimental context, the experimental goals of the present thesis will be developed and outlined.

5.2.1. The pacemaker-accumulator model

The dominant psychological model to explain duration perception is the *pacemaker-accumulator model* of the scalar timing theory (Gibbon, 1977; Gibbon, Church, & Meck, 1984). Specifically, this model proposes the existence of three mechanisms that are involved in the processing of temporal information: an internal clock, a memory store, and a decision stage. The internal clock consists of a pacemaker, a switch component, and an accumulator. According to this model, the pacemaker component is continuously emitting pulses. Whenever an interval to be judged is presented, the switch closes and permits the pulses to be stored into the accumulator. When the interval ceases, the switch opens again and therefore, the pulses are no longer transmitted to the accumulator. The number of accumulated pulses then represents the perceived duration of the interval, which, for example, can be compared with a number of pulses stored in reference memory (e.g., the representation of previously stored interval duration). Therefore, the more pulses are accumulated, the longer is the perceived duration. In principle, changes in perceived duration may result from an increased pacemaker rate or from a change in the latencies of the switch component, (i.e., earlier or later closing and/or opening of the switch). One way to empirically disentangle the effects of these two mechanisms is examining the magnitude of the perceptual duration distortion across a wide range of interval durations. Specifically, if a certain manipulation affects the switch mode, its effect on perceived duration should not depend on the physical duration of the interval. For example, a certain manipulation may cause the switch to close earlier and/or open later. This change in switch latency would cause a corresponding increase in perceived duration, and the magnitude of this effect should be independent of the duration of the interval duration enclosed between the switch operations. Accordingly, any effect mediated by the latency of the switch should remain constant (i.e., additive) with increasing interval duration (Burle & Casini, 2001). In contrast, accelerating the pacemaker pulse rate should produce an overadditive effect as interval duration increases. For example, if a certain manipulation elicits a faster pulse rate, more pulses will be registered per time unit, and consequently perceived duration will be prolonged. Since pulse rate is associated multiplicatively with interval duration, the magnitude of this effect should be proportional to interval duration, getting more pronounced as interval duration increases (Penton-Voak, Edwards, Percival, & Wearden, 1996).

Further models have extended the original internal-clock conception by including, for example, the influence of cognitive process on perceived duration. For example, the *attentional-gate model*, proposed by Zakay and Block (1997) considers the influence of attention on duration perception. This model additionally includes an attentional gate between the pacemaker and the switch component. This attentional gate opens up whenever time is attended, thus permitting pulses to be transmitted from the pacemaker to the subsequent accumulator, and thus is a co-determinant of perceived duration. Therefore, if attentional resources are drawn to non-temporal factors (e.g., by including a secondary task), shorter duration will be perceived.

5.2.2. Unimodal intervals: the role of stimulus modality in duration perception

The human ability to accurately perceive and estimate duration is crucial for interacting with the large variety of constantly changing events surrounding us. Accurate temporal processing is necessary to estimate and predict occurrence and duration of these events and thus, enables us to prepare and control our actions accordingly. However, despite the importance of accurate duration perception, it has been demonstrated that the perceived duration of a given event typically does not depend exclusively on its physical duration. Instead, it can be strongly affected by contextual factors of the stimulus presentation (for an overview, see Bausenhart, Bratzke, & Ulrich, 2016), for example, as the modality of the target event (Wearden, Edwards, Fakhri, & Percival, 1998), the modality of accompanying events (K. M. Chen & Yeh, 2009; Klink et al., 2011; Romei, De Haas, Mok, & Driver, 2011; Walker & Scott, 1981) and the presence and nature of preceding stimulation (Dyjas, Bausenhart, & Ulrich, 2014; Heron et al., 2012; Penton-Voak et al., 1996).

A basic and well-replicated finding regarding the role of interval modality is that auditory intervals are typically perceived as being longer (Goldstone & Lhamon, 1972; Ulrich, Nitschke, & Rammsayer, 2006; Walker & Scott, 1981; Wearden et al., 1998), and with a greater discrimination accuracy than visual intervals of the same physical duration (Gamache & Grondin, 2010; Ulrich et al., 2006). This result has been attributed to changes of the pacemaker pulse rate depending on the interval modality (Wearden et al., 1998). Specifically, it is assumed that auditory intervals increase the state of arousal, which in turn increases the pulse rate of the pacemaker component of the internal clock mechanism. This would result in a higher number of pulses submitted to the accumulator, and thus, longer perceived duration of auditory than of visual intervals. Consistent with this notion and the predictions of the pacemaker-accumulator model outlined above, the difference in perceived duration between unimodal visual and unimodal auditory intervals increases with increasing interval duration (Penton-Voak et al., 1996; Wearden et al., 1998), thereby supporting the notion that modality-based effects on perceived duration are based on modulations of the pacemaker rate.

5.2.3. Multimodal duration integration: decrease of discrimination sensitivity or changes in perceived duration?

When temporal intervals are presented bimodally rather than unimodally, it is typically observed that perceived duration of visual intervals is strongly distorted towards the duration of concurrently presented auditory intervals (Asaoka & Gyoba, 2016; Heron, Hotchkiss, Aaen-stockdale, Roach, & Whitaker, 2013; Klink et al., 2011; Romei et al., 2011). Importantly, two lines of evidence should be distinguished depending on the relative timing between the modalities composing the multimodal interval. First, some studies have investigated the effect of combining auditory and visual intervals with *identical* durations (K. M. Chen & Yeh, 2009; Walker & Scott, 1981). For example, Walker and Scott (1981) employed a time reproduction task in which participants had to reproduce the duration of either unimodal auditory intervals, unimodal visual intervals, or concurrently presented auditory and visual intervals. As outlined above, unimodal auditory intervals were reproduced longer than unimodal visual ones of the same duration. Interestingly, the reproduced duration of the combined audiovisual intervals was very similar to the reproduced duration of the unimodal auditory intervals, and again, the difference to unimodal visual reproductions increased with increasing interval duration (K. M. Chen & Yeh, 2009). Accordingly, a possible interpretation is that the auditory modality completely dominates the pacemaker's pulse rate, and thus perceived duration, also for bimodal stimuli.

Second, some other studies have also investigated the effects of combining auditory and visual intervals with *conflicting* durations (Klink et al., 2011; Romei et al., 2011). For example, Klink et al. (2011) employed a discrimination task in which participants had to indicate which duration of a pair of filled intervals (standard and comparison interval duration) presented to a task-relevant modality (either visual or auditory) was longer. In order to investigate the multimodal bias effect, intervals were also presented within the task-irrelevant modality. These accompanying irrelevant intervals could be of the same duration or longer and shorter than the relevant modality interval (i.e., an irrelevant long interval was always presented with a relevant short interval, and vice versa). The results of this experiment were consistent with the evidence for auditory dominance over visual duration described above: discrimination accuracy for task-relevant visual intervals was strongly impaired by the presence of irrelevant auditory intervals with incongruent durations, but not vice versa. That is, task-irrelevant visual intervals did not lead to impaired discrimination accuracy of relevant auditory interval durations.

Similarly, other studies have also shown that accuracy of visual duration discrimination decreases when visual intervals are accompanied by task-irrelevant auditory intervals of incongruent durations, while accuracy increased with the presentation of congruent auditory duration (Romei et al., 2011; Sarmiento, Shore, Milliken, & Sanabria, 2012). These results, therefore, provide important evidence of multimodal integration of perceived duration. However, these previous studies were predominantly based on accuracy measures of duration discrimination, and therefore, it remains unclear whether the impairment of visual duration discrimination in the presence of incongruent auditory intervals is due to a decrease in sensitivity, to a biased perception of the visual duration towards the auditory one, or both. This issue is addressed in Experiment 1[1] of this thesis.

5.2.4. Visual and auditory contribution to the multimodal perceived duration

Importantly, the studies outlined above have almost unequivocally shown unidirectional effects, that is, audition affected visual duration discrimination and reproduction, but visual duration did not affect auditory duration discrimination or reproduction. In accordance with the modality appropriateness account, these findings suggest that the auditory temporal information, presumably due to the typically higher temporal resolution of the auditory system compared to the visual one (Bertelson & Aschersleben, 2003; Getzmann, 2007; Morein-Zamir et al., 2003), plays a dominant role in duration perception. Thereby, it may completely overrule visual duration information and thus determine the multimodal percept (Asaoka & Gyoba, 2016; K. M. Chen & Yeh, 2009; Klink et al., 2011; Romei et al., 2011; Walker & Scott, 1981). According to the Bayesian cue combination account outlined above, however, complete auditory dominance is not mandatory – rather, both modalities may contribute to the combined multimodal duration percept. In fact, it has been recently demonstrated that under conditions where the reliability of the auditory signal is degraded, the typical auditory dominance on multimodal perceived duration declines, and in turn, the influence of the visual temporal information on the combined

---

[1] For more details see Appendix A (Bausenhart, K. M., De La Rosa, M. D., & Ulrich, R. (2014). Multimodal integration of time: Visual and auditory contributions to perceived duration and sensitivity. *Experimental Psychology*, *61*(4), 310–322. https://doi.org/10.1027/1618-3169/a000249).

percept increases (Hartcher-O'Brien et al., 2014; Walker & Scott, 1981). Therefore, the present Experiment 1 also addressed whether perceived multimodal duration is strictly dominated by the auditory duration or, otherwise, visual temporal information can also contribute to the multisensory perceived duration by biasing perceived auditory duration.

5.2.5. Potential mechanisms underlying multimodal integration of perceived duration

Importantly, Klink et al. (2011) proposed two potential mechanisms involved in the integration of audiovisual conflicting durations: First, they suggested that the integration of multisensory intervals might be determined by a temporal ventriloquist effect on the mode of the switch mechanism of the internal clock. Accordingly, the onset and offset of the visual internal are drawn towards the onset and offset of the concurrently presented auditory interval (see Figure 2). Therefore, when the visual interval is accompanied by a longer auditory interval, the perceived multisensory onset and offset cause the switch to close earlier and open later, respectively. This yields a larger number of pulses forwarded into the accumulator, thus resulting in longer perceived duration. Likewise, when the accompanying auditory interval is shorter, then the multisensory perceived onset and offset cause the switch to close later and open earlier, respectively, resulting in less accumulated pulses, and thus, in shorter perceived duration. In addition to the switch mode, these authors proposed a second mechanism based on changes of the pulse rate of the internal pacemaker component, which would be predominantly determined by the auditory information (K. M. Chen & Yeh, 2009; Penton-Voak et al., 1996; Wearden et al., 1998). As already outlined above, previous studies (K. M. Chen & Yeh, 2009; Walker & Scott, 1981) have shown direct evidence for this latter notion by employing audiovisual congruent intervals with varying durations. However, no direct evidence on the effect of a ventriloquist effect on the switch mode underlying multimodal integration of conflicting durations has been provided so far. This issue was addressed in Experiments 2-4[2].

---

[2] For more details see Appendix B (De la Rosa, M. D., & Bausenhart, K. M. (2013). Multimodal integration of interval duration: Temporal ventriloquism or changes in pacemaker rate? *Timing and Time Perception*, *1*(2), 189–215. https://doi.org/10.1163/22134468-00002015)

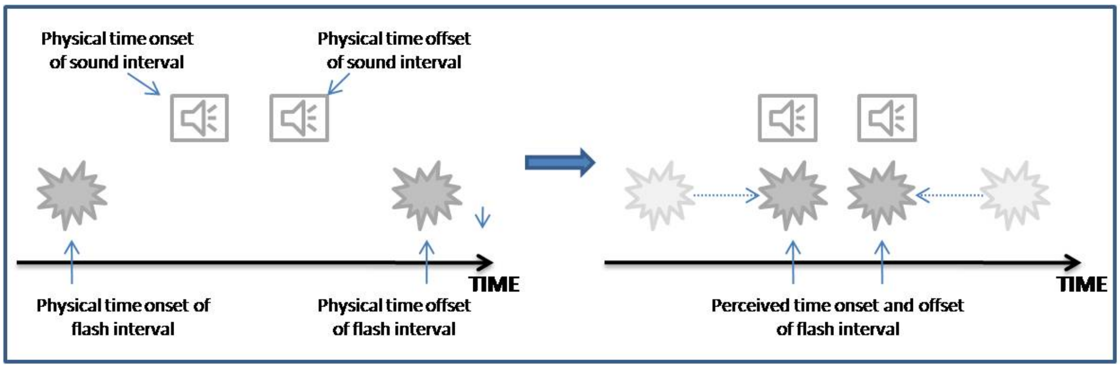

Figure 2. Temporal ventriloquist effect on multisensory duration integration: the perceived onset and offset of the visual interval is pulled towards the onset and offset of a concurrently presented auditory interval.

5.2.6. At which level of processing does multisensory duration integration take place?

Taking into account the studies outlined above, as well as the present Experiments 1 to 4, there is ample evidence for multisensory influences on duration discrimination (Bausenhart, De la Rosa, & Ulrich, 2014; De la Rosa & Bausenhart, 2013; Klink et al., 2011; Romei et al., 2011; Sarmiento et al., 2012). However, also for the domain of duration perception, it remains unclear whether these effects depend on integration within perceptual processing stages, that is, at a rather low level of processing, or otherwise, depend on later, decisional processes, that is, at a higher level of processing. A strict decisional account of auditory dominance, for example, would imply that the participants, deliberately or incidentally, respond based on the auditory temporal information instead of the visual one. As an extreme example, auditory dominance may result if participants simply close their eyes during stimulus presentation and attend only to the auditory stimulation. Such conditions would of course rule out a genuine perceptual integration account of any observed multimodal biasing effect. Yet, some previous evidence argues against a purely decisional bias as basis of the multisensory integration effects on perceived duration. For example, Klink et al. (2011) investigated whether a task-irrelevant auditory interval would still bias perceived duration of a simultaneous task-relevant visual interval, if the pair of audiovisual intervals is preceded by a regular stream of irrelevant auditory stimulation. According to the authors, perceptual multisensory integration should be impaired under these conditions, since intramodal grouping (that is, perception of a rhythmic stream within the auditory modality) would prevent intermodal grouping (that is, integration of the single visual interval with the auditory accompanying interval). On the other hand, participants should still be able to respond on basis of the auditory information – that is, the multimodal bias effect should remain if it was due to a decisional bias instead of a genuine integration of the perceptual information. The results showed that the multisensory bias effect diminished under these conditions, therefore indicating that it is contingent on intramodal grouping – that is, on perceptual multimodal integration rather than a mere decisional bias. Still, it remains possible that the multisensory integration effects on perceived duration observed in Experiments 1-4 of this thesis may be due to a decisional bias by which participants attend and respond to auditory temporal information instead of the (task-relevant) visual one. Therefore, Experiment 5[3] was created in order to rule out this possibility by employing uninformative, but still potentially biasing task-irrelevant sounds.

Even more compelling evidence for a genuine perceptual integration may be provided, as outlined above, if the duration information from the biasing domain is disentangled from the response domain. In fact, several studies have taken this approach and provided evidence for perceptual integration of temporal information, by assessing, for example, temporal order judgments based on spatial location (Morein-Zamir et al., 2003), Ternus motion perception (Shi et al., 2010) and discrimination performance in a four-dot masking paradigm (Vroomen & Keetels, 2009). Most important for the present thesis is an astonishing finding from de Haas et al. (2013), who investigated the effect of sound duration on the detection of concurrently presented Gabor patches. Specifically, the participants´ task was to detect a Gabor patch that could be presented in one of two intervals of dynamic white noise. The task was divided into two different conditions. In the unimodal condition, the Gabor patch was presented alone for one of eight different durations between ~24 and ~192 ms. In the audiovisual condition, the Gabor patch lasted for a fixed duration of ~24 ms and was always accompanied by a continuous sound whose duration varied between the same eight durations as the visual target in the unimodal condition. In the unimodal condition, unsurprisingly, target detection sensitivity increased with increasing duration of the visual target. Crucially, in the audiovisual condition, detection sensitivity for the 24-ms Gabor patch increased with increasing sound duration from 24 ms up to 60 and 96 ms, even though the sound itself was uninformative with regard to which of the two intervals contained the visual target. The authors concluded that early audiovisual interactions might be responsible for this effect. Specifically, the increasing sound duration might have produced sustained visual activation, resulting in a prolonged representation of the visual target, which therefore would facilitate target detection. Similar to the studies outlined above, this proposed mechanism is basically consistent with a temporal ventriloquist effect by which the onset and offset of the visual target were pulled towards the sound onset and offset respectively. In the present thesis, we investigated whether this intriguing phenomenon could be transferred to the masked-letter identification task employed by Y. C. Chen and Spence (2011). As already described in the previous section, Y. C. Chen & Spence observed that, relative to a

---

[3] For more details see also Appendix B.

unimodal target presentation, masked letter identification performance was improved by the concurrent presentation of sounds. However, they employed a fixed sound duration and therefore, it is unclear whether an additional effect of varying sound duration on such a higher-level task as letter discrimination may be observed. Experiments 6-8[4] were thus conducted to investigate whether varying sound duration can modulate masked letter identification performance, which, in turn, would provide strong evidence in favour of genuine multisensory interactions in the integration of duration.

5.2.7. Can multimodal integration effects on perceived duration persist over time?

As outlined in the previous section, ample evidence on sustained multisensory integration effects has been provided by studies employing temporal order judgments and subjective simultaneity perception. However, comparable evidence in the domain of duration perception is much scarcer. So far, only one published study has addressed this issue (Heron et al., 2013). Specifically, Heron et al. (2013) performed two main experiments (and a series of additional control experiments). First, they investigated the immediate effects of multimodal duration integration. Specifically, they employed a two-alternative forced-choice task in which participants had to discriminate between a 320-ms test stimulus and a reference stimulus with variable duration. These task-relevant intervals could be presented in the auditory or the visual modality. Moreover, the test stimulus was accompanied by an irrelevant stimulus in the opposite modality and with a duration of either 200 ms or 510 ms. The results showed similar immediate multisensory integration effects on perceived duration as shown by previous studies (Klink et al., 2011; Romei et al., 2011) and in Experiments 1-4 of the present thesis: perceived visual duration was strongly biased towards the duration of concurrently presented irrelevant auditory intervals. However, the reverse effect was not observed. In the second main experiment, the authors tested for sustained multimodal effects. In an adaptation phase, participants were exposed to a repeated presentation of audiovisual incongruent intervals. In a following test phase, participants were asked to reproduce the duration of a unimodally presented visual interval. The authors suggested two different hypotheses regarding the hierarchical order of duration adaptation and multimodal integration. In a previous study (Heron et al., 2012) they showed that prolonged exposure to a relatively long or short interval produces repulsive aftereffects on perceived duration of a subsequently presented interval, that is, duration

---

[4] For more details, see Appendix C (de la Rosa, M. D., & Bausenhart, K. M. (2018). Enhancement of letter identification by concurrent auditory stimuli of varying duration. *Acta Psychologica*, *190*, 38–52. https://doi.org/10.1016/j.actpsy.2018.07.001)

adaptation. Accordingly, if multisensory duration integration precedes duration adaptation, then during the adaptation phase, perceived duration of the visual adaptation intervals would be biased towards the duration of the concurrent auditory intervals. Then, participants would adapt to this lengthened or shortened multimodal perceived duration, which should consequently result in a repulsive adaptation aftereffect on reproduced duration in the test phase (e.g., a visual test interval should be perceived as being shorter when participants previously adapted to a prolonged multisensory duration). On the other hand, however, if duration adaptation precedes multimodal duration integration, in the adaptation phase, participants would adapt to the perceived visual duration (since it is not yet integrated with the auditory duration), and therefore, in the test phase, no repulsive adaptation effect would be observed based on the auditory biasing information presented during the test phase. The results were in line with this second possibility, that is, perceived duration of unimodal visual intervals in the test phase was not affected by the previous exposure to audiovisual incongruent intervals. These findings lead to the conclusion that multisensory integration is produced only in a transient, immediate manner, and is preceded by duration adaptation in the hierarchy of processing. Experiments 9 and 10[5] of the present thesis were conducted with the aim to provide independent evidence on the sustained multisensory integration effects on perceived duration, with a set of experimental conditions presumably even more effective to produce immediate and potential sustained multisensory integration effects, and a crucial control condition to test for the effectiveness of this manipulation.

6. Experiments

6.1. Multimodal duration integration: the role of interval modality in perceived duration and sensitivity

Previous studies have shown that visual duration discrimination accuracy is strongly impaired when concurrent auditory intervals are presented with incongruent duration but not vice versa (Klink et al., 2011; Romei et al., 2011). Importantly, these previous results are based on accuracy measures of duration discrimination (as, e.g., percent correct identification). Therefore, it is not clear whether these impairments of visual duration discrimination are due to changes in sensitivity or to changes in perceived duration. Experiment 1 was conducted in order to address this issue by assessing psychometric functions for the discrimination of audiovisual intervals

---

[5] More details of Experiments 9 and 10 in Appendix D (De la Rosa, M. D., & Bausenhart, K. M. (2018). Still no evidence for sustained effects of multisensory integration of duration. *Multisensory Research*, *31*(7), 601–622. https://doi.org/10.1163/22134808-18001296)

with congruent and incongruent durations. This procedure allows for assessing separately integration effects on perceived duration, quantified by the Point of Subjective Equality (*PSE*) and on discrimination sensitivity, quantified by the Just Noticeable Difference (*JND*). Additionally, this experiment also allowed for testing whether the multimodal integration effect is strictly determined by the auditory temporal information as suggested by previous studies (K. M. Chen & Yeh, 2009; Walker & Scott, 1981), or in turn, whether visual duration can also contribute to the multisensory percept, by biasing perceived duration of concurrent auditory intervals (cf. also Hartcher-O'Brien et al., 2014).

Specifically, in Experiment 1 participants had to perform a duration discrimination task in which they had to compare the duration of empty standard (fixed duration of 500 ms) and comparison intervals (varying duration) presented in a task-relevant modality. In two separate experimental sessions either vision or audition was designated as task-relevant (see Figure 3). In order to test the multimodal integration effects, the onset and offset markers of the relevant standard and comparison intervals were accompanied by interval markers in the task-irrelevant modality. For example, in the visual session, the onset and offset markers of the standard and comparison intervals were accompanied by task-irrelevant auditory markers (and vice versa in the auditory session). In each trial, task-relevant and task-irrelevant markers of the standard interval were always presented simultaneously. However, for the comparison interval, there were either no irrelevant interval markers (*unimodal* condition), or they defined an interval either 100 ms *longer*, 100 ms *shorter*, or of the *same* duration as the task-relevant comparison interval. Task-relevant and –irrelevant intervals were center-aligned, that is, their temporal midpoints were simultaneously aligned (e. g., the irrelevant interval in the longer condition started 50 ms before and ended 50 ms after the relevant one). Participants´ task was to indicate which of the intervals presented in the task-relevant modality was longer. Moreover, they were also instructed to ignore the intervals presented in the task-irrelevant modality. In order to ensure the participants´ adherence to this instruction, a relatively high proportion of unimodal trials (50% of all trials) was employed and all four bias conditions (unimodal, shorter, same, and longer) were randomly intermixed. Since the unimodal trials provided no temporal information about the duration of the comparison interval in the task-irrelevant modality, in 50% of all trials, it would have been impossible for participants to perform the task by relying on the task-irrelevant information alone. This high proportion of unimodal trials and their unpredictable occurrence should cause participants to attend to and respond on base of the task-relevant modality, and thus reduce the possibility of a decisional bias, as outlined above, underlying any potential multisensory integration effects.

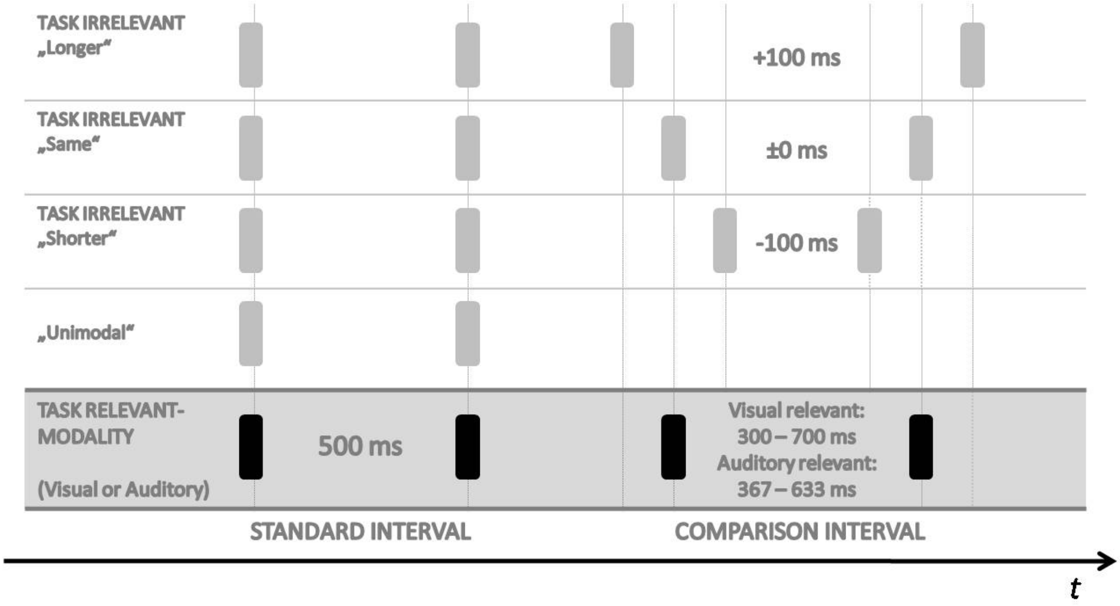

Figure 3. Temporal positions of relevant and irrelevant markers for the standard and the comparison interval in the four bias conditions of Experiment 1.

Logistic psychometric functions were fitted, for each participant and condition, to the proportion of “comparison longer” responses for all comparison durations. From these functions, *PSE* and *JND* were computed. *PSE* corresponds to the comparison duration at which the participant perceives comparison and standard to be equal in duration and thus, reflects the perceived duration of the comparison interval (accordingly, the larger *PSE*, the shorter perceived duration). *JND* amounts to half the interquartile range of the psychometric function and provides a measure of sensitivity of duration discrimination between the standard and comparison intervals (thus, the lower the value of *JND*, the higher discrimination sensitivity).

In summary, the results showed a strong influence of the duration of irrelevant auditory intervals on perceived visual duration. Specifically, in comparison to the “same” condition, task-relevant visually marked comparison intervals were judged as being longer or shorter when accompanied by irrelevant auditorily marked intervals with longer or shorter duration, respectively (see Figure 4, left panel). Interestingly, the multimodal bias was bidirectional, such that irrelevant visual intervals also influenced auditory perceived duration, even though to a lesser extent. Regarding discrimination sensitivity (Figure 4, right panel), the discrepancy between the “longer” and “shorter” conditions, which produced pronounced changes in perceived duration, did not lead to a corresponding decline in discrimination sensitivity. Indeed, discrimination sensitivity was similar between the “longer”, “shorter”, and “same” bias conditions for both modalities. Only in the unimodal visual task condition, discrimination sensitivity was lower than in all other conditions, which is in line with previous studies (Gamache & Grondin, 2010; Ulrich et al., 2006) showing lower duration discrimination sensitivity for visual intervals than for auditory intervals. Therefore, even though the multisensory integration of incongruent interval durations may evoke pronounced biases in perceived duration, it does not go along with a decline in sensitivity for duration.

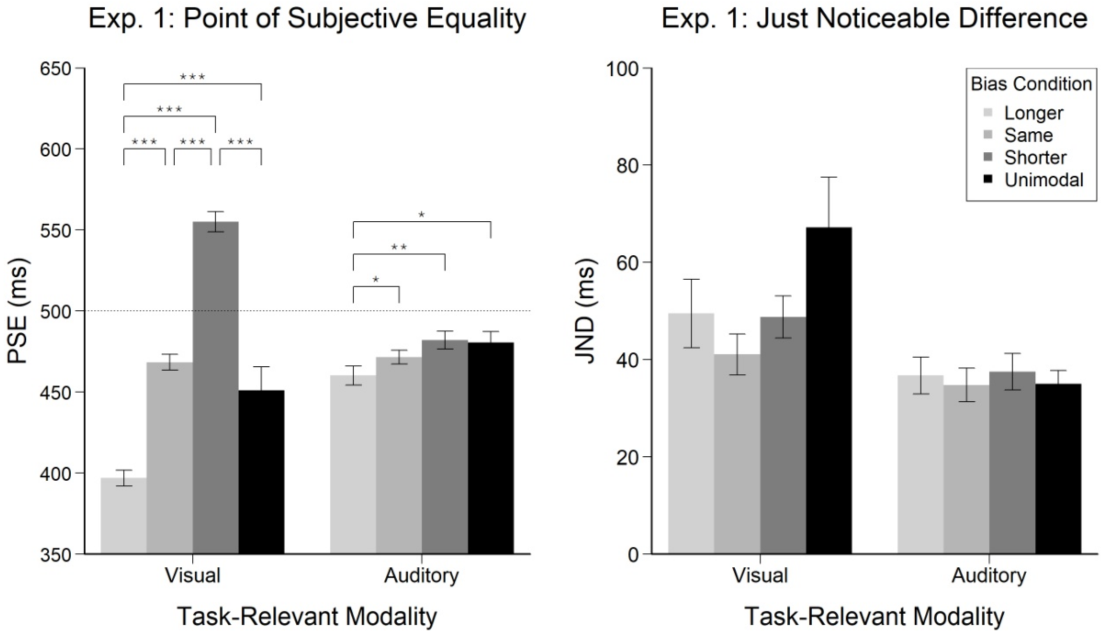

Figure 4. Mean Point of Subjective Equality (*PSE*; left panel) and mean Just Noticeable Difference (*JND*; right panel) in ms, depending on task modality and bias condition in Experiment 1.

6.2. Mechanisms underlying multimodal integration of perceived duration

Experiment 2 was conducted in order to investigate which mechanisms underlie these multimodal duration integration effects. As outlined above, within the pacemaker-accumulator model of temporal perception, two basic components have been considered as potentially prone to be affected by multisensory integration. First, the latencies of the internal switch component may be altered by means of temporal ventriloquist effects (i.e., the closing and opening times of the switch component would be “pulled” towards the interval markers in the irrelevant modality). Second, the pacemaker´s pulse rate may be differentially affected by inputs from different modalities (i.e., an “early” auditory pulse in the case of a “longer” auditory interval might increase the pulse rate and thus lead to especially long estimates of the subsequent visual interval). In order to dissociate these two mechanisms, we compared the magnitude of the effect of combining audiovisual congruent and incongruent durations across a large range of interval durations. To this end, we employed a time reproduction task, in which participants had to reproduce the duration of visually marked empty intervals, similar to those employed in the visual duration discrimination task of Experiment 1. These visual intervals could last for either 500, 753, 1000, or 1253 ms. As in Experiment 1, visual intervals could be also presented alone (*unimodal* condition), or accompanied by auditory marked intervals with slightly *longer* (+100 ms), *same* (± 0 ms), or *shorter* (-100 ms) duration. Visual and auditory intervals were center-aligned, and participants had to reproduce the visually marked intervals while ignoring the auditory ones. Based on the results of Experiment 1, we expected that the reproduced visual durations would be biased towards the duration of the task-irrelevant auditory intervals. Crucially, if this multimodal integration effect is mediated by a temporal ventriloquist effect on the switch latencies, the auditory biasing effect on reproduced visual duration should be independent of interval duration and thus have a similar magnitude across the whole range of employed interval durations (i.e., it should be additive with interval duration). Otherwise, if multimodal integration causes changes of the pacemaker´s pulse rate, the auditory biasing effect on reproduced visual duration should increase with increasing interval duration (i.e., it should be overadditive with interval duration, see also 3.2.1. and 3.2.3.).

The results of Experiment 2 replicated the results of Experiment 1 by showing that perceived visual duration was strongly biased towards the duration of concurrently presented auditory intervals. Specifically, the longest durations were reproduced when visual intervals were accompanied by longer auditory intervals, while the shortest durations were reproduced when visual intervals were accompanied by shorter auditory intervals or presented unimodally. In addition, visual intervals accompanied by auditory intervals of the same duration were reproduced as slightly longer than unimodal visual intervals. Unexpectedly, however, this observed multimodal bias in perceived duration decreased with increasing interval duration (see Figure 5, left panel). This pattern of results contrasts with both of the hypotheses outlined above. First, this observed underadditive biasing effect is especially inconsistent with the assumption of an increased pacemaker rate which would be predict an increase of the multimodal bias with increasing interval duration. Second, the result pattern also contrasts with the additivity expected in case of a temporal ventriloquist effect on the switch latencies. Interestingly, as can be seen in Figure 5, left panel, another effect might have complicated interpretation of the observed results: it is evident that overall, the shortest intervals within the duration range were overestimated while the longest intervals were underestimated relative to their physical duration. This effect is an instance of Vierordt´s law (Vierordt, 1868) which has been repeatedly shown to affect the results of reproduction methods (for an overview, see Lejeune & Wearden, 2009). It is possible that this effect also confounds the results regarding the multimodal bias effect. Accordingly, the general overestimation of short intervals might also enlarge a "true" multimodal biasing effect at short intervals, while the general underestimation of long intervals might result in an observed decrease of the biasing effect at long intervals.

Regarding the comparison between the "same" and the "unimodal" condition, the results support previous studies (K. M. Chen & Yeh, 2009; Walker & Scott, 1981) by showing that bimodal intervals with congruent duration were reproduced longer than unimodal visual intervals. Interestingly, however, the difference in perceived duration between both conditions was additive across interval duration. This is in contrast to these previous studies, which, as outlined in the Introduction, showed overadditivity across interval duration. Two possible explanations may hold for the observed pattern of results. On the one hand, and as discussed above, a Vierordt effect might have counteracted a possible overadditivity in our experiment. On the other hand, in the present experiment we employed empty intervals, while previous studies employed filled intervals. Indeed, filled auditory intervals have been specifically associated with enhanced arousal (Wearden, Norton, Martin, & Montford-Bebb, 2007). Accordingly, processing of filled and empty auditory intervals seems to be qualitatively different, with only the former interval type producing enhanced arousal. Thus, overadditivity of the difference between unimodal visual and audiovisual congruent intervals might be more likely observed when the accompanying auditory intervals are filled rather than empty (see Walker & Scott, 1981, Experiment 4, for a related result). Therefore, in Experiment 3, we replicated Experiment 2, however with filled rather than empty auditory and visual intervals. Very similar to the results of Experiment 2, visual duration reproductions were again strongly biased towards the auditory irrelevant duration, that is, compared to the "same" condition, reproductions were prolonged in the "longer" condition and shortened in the "shorter" condition. Importantly, one crucial difference emerged with respect to the results of Experiment 2: reproductions in all bimodal conditions were longer than those in the "unimodal" condition, and as expected, these differences now increased clearly with increasing interval duration (see Figure 5, right panel). Therefore, these results replicate the previously observed overadditive interaction for the multimodal integration of audiovisual filled intervals, and moreover, demonstrate that the employed paradigm is basically suited to observe potential overadditive effects. Despite this successful "sanity check", the biasing effect of the different multimodal conditions (i.e., the difference between the "longer" and the "shorter" condition), again decreased with increasing interval duration. As discussed previously, this pattern of underadditivity is basically inconsistent with the two hypothetical mechanisms outlined above, but again might be slightly confounded by a Vierordt effect, which might have counteracted an underlying additive (or even overadditive) biasing effect. In any case, it is noteworthy that filled and empty intervals of incongruent (shorter or longer) duration (i.e., the effect of sound duration) produced very similar effects on perceived duration of visual intervals, while the difference between uni- and bimodal intervals (i.e., the effect of sound presence) seems to depend severely on the interval type.

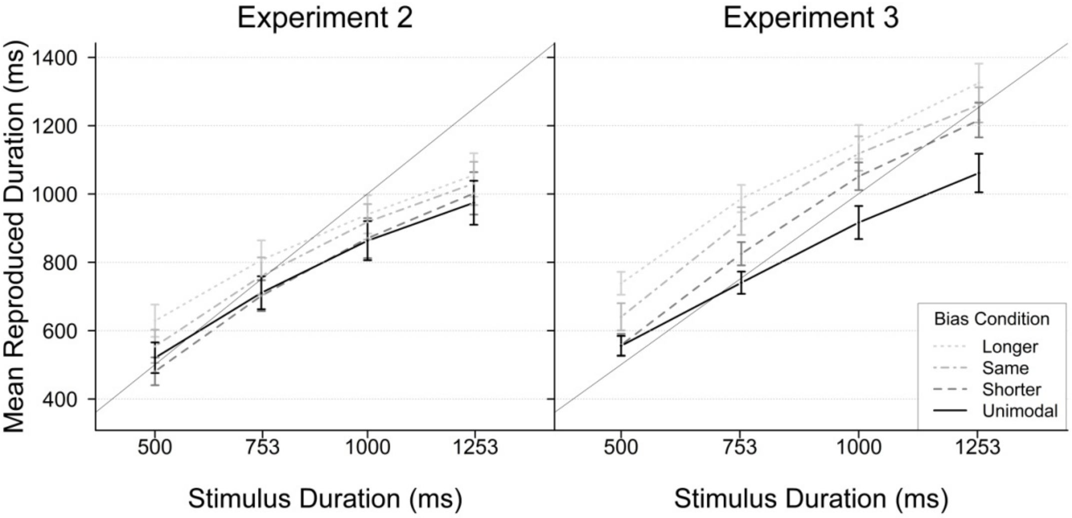

Figure 5. Mean reproduced duration in Experiment 2 (empty intervals) and Experiment 3 (filled intervals) as function of Bias Condition (longer, same, shorter, and unimodal) and Stimulus Duration. The continuous gray line depicts the veridical line.

In order to overcome potential methodological limitations of the reproduction task, Experiment 4 involved again empty intervals and a paired-comparison method similar to the one employed in Experiment 1, in which a standard interval was always presented before a variable comparison interval. Since in this task participants have to directly compare the two presented intervals, these judgments cannot be subject to perceptual distortions based on the whole range of employed durations, such as the Vierordt effect. Unlike Experiment 2 and 3, three visual standard durations, ranging from 500 to 2000 ms, were tested in separate experimental sessions. The “same” condition was omitted in order to shorten the total duration of the experiment, since this method generally requires a much larger number of trials. As in Experiment 1, *PSE* and *JND* estimates were computed separately for each participant, standard duration and bias condition. Overall and consistent with the results of Experiments 1-3, larger (smaller) *PSEs* and thus, shorter (longer) perceived duration was observed when the task-relevant visual intervals were accompanied by “shorter” (“longer”) auditory intervals. Importantly, in this experiment, the comparison between the “longer” and the “shorter” conditions did reveal a constant difference across the wide range of employed interval durations, that is, additivity across the employed interval durations (see Figure 6). Thus, the results of this experiment are well in line with the notion that the perceptual conflict between incongruent audiovisual durations is resolved by means of a temporal ventriloquist effect on the perceived onset and offset of the intervals, which in turn affects the switch component of the internal clock mechanism.

Moreover, even though conflicting auditory durations led to a strong distortion of perceived visual duration, measures of *JND* revealed that the temporal discrepancy between the interval modalities again did not produce a decline in discrimination sensitivity. Overall, the unimodal visual condition yielded even relatively larger *JNDs*, and thus, reduced sensitivity of duration discrimination, compared to the bimodal incongruent conditions.

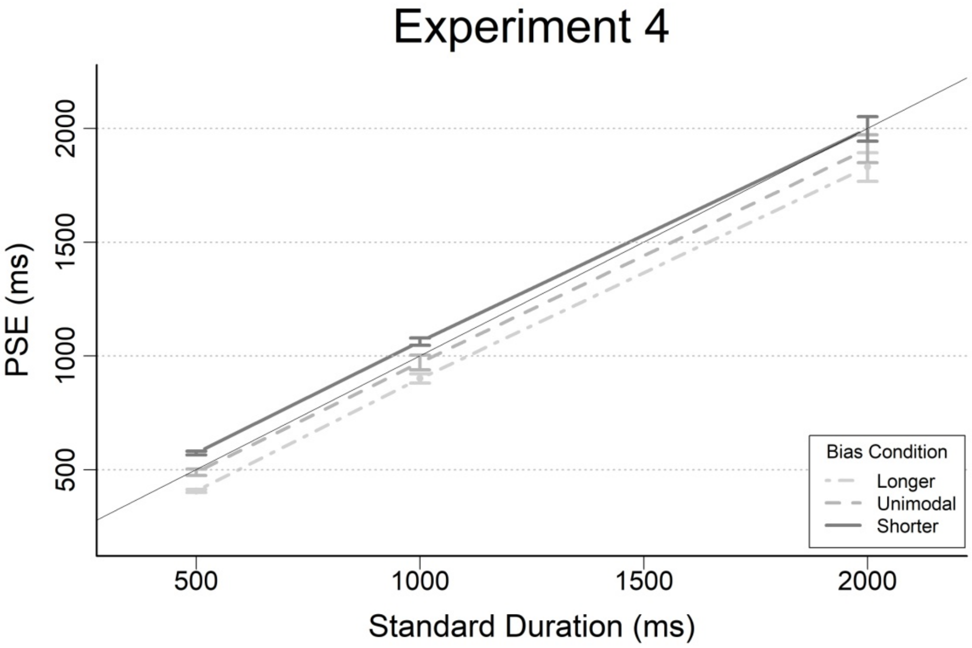

Figure 6. Mean PSE in Experiment 4 as a function of Bias Condition and Standard Duration. Error bars represent the standard error of the mean. The continuous black line depicts the veridical line.

6.3. On the locus of multisensory duration integration

In order to draw valid conclusions about the multisensory mechanisms involved in the results of Experiments 1-4, the possibility should be considered that participants responded to the irrelevant (e.g., auditory) temporal information instead of the relevant (e.g., visual) one. As outlined above, in this case, the observed effects should be interpreted as due to decisional or attentional bias rather than to a genuine perceptual integration. With the aim to prevent such harmful processing strategies, different precautions had been taken, as careful instruction and a relatively high proportion of randomly interspersed unimodal trials (e.g., 50% in Experiment 1), such that the task could not be adequately performed by consistently attending and responding to the task-irrelevant modality.

Still, more convincing evidence of an underlying perceptual integration process rather than decisional bias would require a multimodal integration effect on perceived duration when potentially biasing stimulation is presented in the irrelevant modality but without providing sufficient information about interval duration – such that participants could not perform the duration discrimination task based on the task-irrelevant modality alone. In Experiment 5, such an experimental situation was created. Specifically, the same visual task as in Experiment 1 was employed, but in this case the auditory modality did not provide information about the duration of the comparison interval. Specifically, as in Experiment 1, the visual markers of the standard intervals were accompanied by simultaneous auditory pulses. However, the visual markers of the comparison intervals were accompanied by only a single auditory pulse, presented according to three different bias conditions: before, simultaneous with, or after one of the visual markers. Thus, by means of a temporal ventriloquist effect, it might either produce a "shorter" (e.g., when presented after the visual onset, or before the visual offset), "same" (when simultaneous) or "longer" (e.g., when presented before the visual onset, or after the visual offset) multimodal percept. Again, the results are consistent with the results observed in Experiments 1-4. Even though to a reduced extent, a clear auditory biasing effect on perceived duration of the visual intervals was observed even when only a single irrelevant auditory pulse was presented. This was independent of whether the auditory pulse accompanied the onset or the offset of the visual intervals, and thus, cannot be attributed to preparatory processing before the interval onset. The analysis on *JND* estimates showed a higher sensitivity when the auditory pulse accompanied the onset rather than the offset of the visual intervals, but, as in previous experiments, the presence of incongruent temporal information (e.g., shorter and longer vs. same condition) did not lead to sensitivity impairments. In sum, multimodal duration integration effects can also be observed, when the information from the irrelevant modality alone is insufficient for performing the task, and thus, a biased decision strategy based on the information from the irrelevant modality can be excluded.

As outlined in the Introduction, another possibility to demonstrate genuine multisensory integration effects within perceptual processing would be to assess whether auditory duration can modulate non-temporal aspects of visual perception. We addressed this issue in Experiment 6 by assessing the effect of manipulating the duration of concurrently presented sounds on the identification of masked letters. To this end, we employed a letter identification task combined with a backwards-masking paradigm (Kinsbourne & Warrington, 1962; Rolke, 2008; Turvey, 1973). In this task, participants had to identify letters (A, S, D, F, G, H, J, or K), presented centrally for 20 ms and followed by a visual random-noise pattern mask after a variable interstimulus interval (ISI) ranging from 0 and 33 ms. Following the experimental methodology of Experiments 1-5, in Experiment 6, the target letter could be presented alone (*unimodal* condition) or accompanied by a sound with *longer* (33 ms), *same* (20 ms), or *shorter* (13 ms) duration. Again, sounds and target letters were temporally center-aligned. According to the results of Y. C. Chen and Spence (2011) from another masking task, we expected that the presentation of any sound should improve letter identification performance specifically at intermediate ISI durations (see Introduction for details). Additionally, if the duration of the concurrent sound can also modulate letter identification by temporally shortening or lengthening the corresponding visual object representation (de Haas et al., 2013), letter identification performance should increase with the duration of the concurrent sounds.

The results of Experiment 6 (cf. Figure 7, left panel), however, showed no reliable effects of the accompanying sounds. The presence of sound in the “same” condition only tended to improve letter discrimination compared to the “unimodal” condition, and the manipulation of the sound duration did not modulate letter identification at all. Further analyses indicated that the tendential sound-induced enhancement was constricted to the 7-ms ISI, which yielded an intermediate performance level of around 60% of correct responses. Interestingly, comparing performance levels between our and Y. C. Chen and Spence´s study, these authors observed sound-induced facilitation within a specific range of ISIs between 27 and 40 ms, which also yielded around 60-70 % correct responses. For all ISIs above 7 ms in our Experiment 6, accuracy already reached relatively high levels above 80%. Presumably some aspects of our specific employed masking paradigm lead to a low task difficulty and thus a steeper initial slope of the masking-ISI function, compared to the one of Y. C. Chen and Spence (2011). One notable difference in this regard is the relatively higher target-to-background contrast compared to the study of Y. C. Chen and Spence (2011), which might have produced a strong visual activation that enabled high accuracy by itself. Given existing evidence that target contrast may modulate sound-induced benefits (e.g., Noesselt et al., 2010), in Experiment 7 letters were presented with a lower target-to-background contrast. The results of this second experiment indeed showed overall lower performance levels, and importantly, a reliable sound-induced enhancement of letter identification compared to the unimodal condition (cf. Figure 7, right panel). Again, however, varying sound duration did not produce any modulation of letter identification.

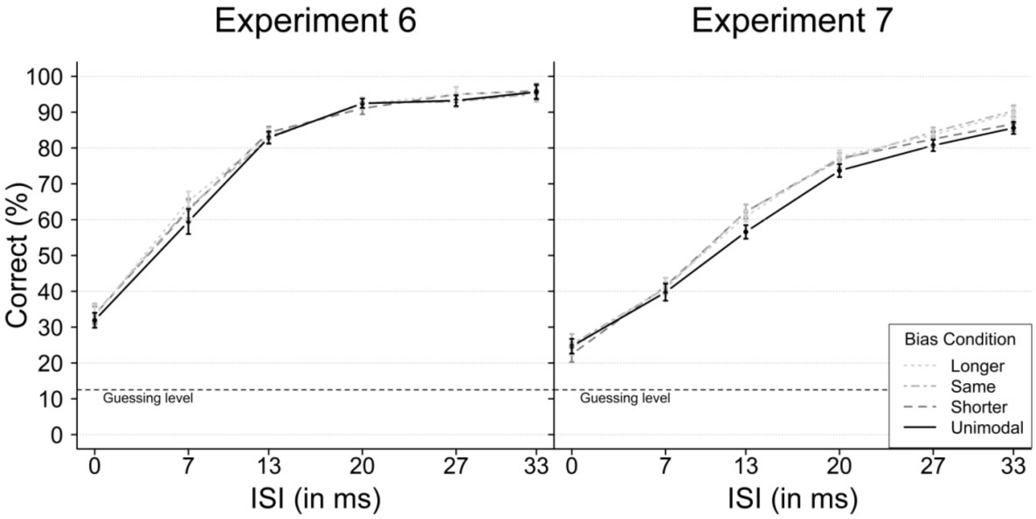

Figure 7. Mean percentage of correct responses in Experiment 6 and in Experiment 7 as a function of Sound Condition and ISI.

Taking Experiment 6 and 7 together, one may speculate that the sound-induced enhancement may be modulated by letter contrast, such that sound presence enhances letter identification when letters are presented at a relatively low, but not at high contrast. This seems consistent with the principle of "inverse effectiveness" (Meredith & Stein, 1983, 1986; Stein & Meredith, 1993), which has also received behavioural support from studies employing lower-level tasks as contrast detection (Noesselt et al., 2010; Senkowski, Saint-Amour, Höfle, & Foxe, 2011). This principle of multisensory integration implies maximal crossmodal enhancement of a multisensory stimulus representation when the corresponding unisensory signals evoke a weak response in isolation. This suggests early multisensory interactions by which a sound might enhance the perceptual representation of a relatively weak visual signal. Two further experiments (see Appendix C) were conducted to test this assumption more directly, by manipulating target contrast as a within-subject variable. In these experiments, only two sound conditions (same and unimodal) were tested. Both experiments replicated the general sound–induced benefit on letter identification observed in Experiment 7, but it was invariant across target contrasts and ISI durations Additional analyses taking interindividual performance levels into account suggested that the observed sound-induced enhancement might instead be most pronounced at intermediate performance levels.

Returning to the main purpose of Experiments 6 and 7, the lack of a modulation of letter identification by sound duration can lead to two different suggestions: first, it might be possible that the manipulation of sound duration was too subtle to affect the duration of the multimodal percept, and consequently, also no effect on letter identification can be expected. Second, the sound duration manipulation might have been basically effective in altering perceived multimodal duration, but without producing any corresponding benefit effect on the identification of the target letter. Therefore, Experiment 8 was conducted as a manipulation check in order to investigate whether the duration manipulation was basically effective in eliciting multisensory integration effects. In this experiment, the target letter could be presented alone (no-sound condition) or accompanied by an auditory interval of different durations (7, 20, or 33 ms). Unlike in the previous experiments, letter duration also varied between 7, 20, and 33 ms. In two sub-experiments, the same participants received two different task instructions. In Experiment 8a, as in the previous experiments, participants were asked to identify the visual letter identity. In Experiment 8b, they were asked to identify the duration of the visual letters (“short”, “medium”, or “long”), again while ignoring the task-irrelevant sounds.

If the manipulation of sound duration was too feeble to produce a modulation of perceived duration, then none of the experiments should show an effect of varying sound duration. However, Experiment 8b showed that varying the duration of the sounds clearly affected perceived duration of the masked target letters (Figure 8, right panel), such that participants’ responses were clearly biased towards the duration of the accompanying sounds (e.g., a 7-ms visual target accompanied by sounds of longer duration were significantly often misreported as being of longer duration). Therefore, the manipulation of sound duration is basically effective in producing an effect on the identification of visual letter duration. In addition, the results of Experiment 8a showed that, as expected, letter identification accuracy increased significantly with increasing visual target duration (see Figure 8, left panel), but not with increasing auditory duration. Therefore, even if the manipulation of sound duration alters reports of perceived duration, it does not alter the letter identification performance. In addition, Experiment 8a replicated the benefit of letter identification caused by the presence of concurrent sounds, however, unexpectedly only for the longest letter duration - which may indicate that this general multisensory integration effect is somewhat sensitive to the experimental context.

In summary, however, the results of Experiments 7 and 8a showed that the mere presence of sound yielded an enhanced masked letter identification performance. In contrast to the results of Y. C. Chen and Spence (2011), this sound benefit did not vary across ISI durations, and the theoretical implications of this different result pattern will be discussed below. Despite the general sound-induced benefit, however, varying the duration of the concurrent sound did not differentially affect letter identification performance, even though it produced a pronounced effect on perceived visual duration.

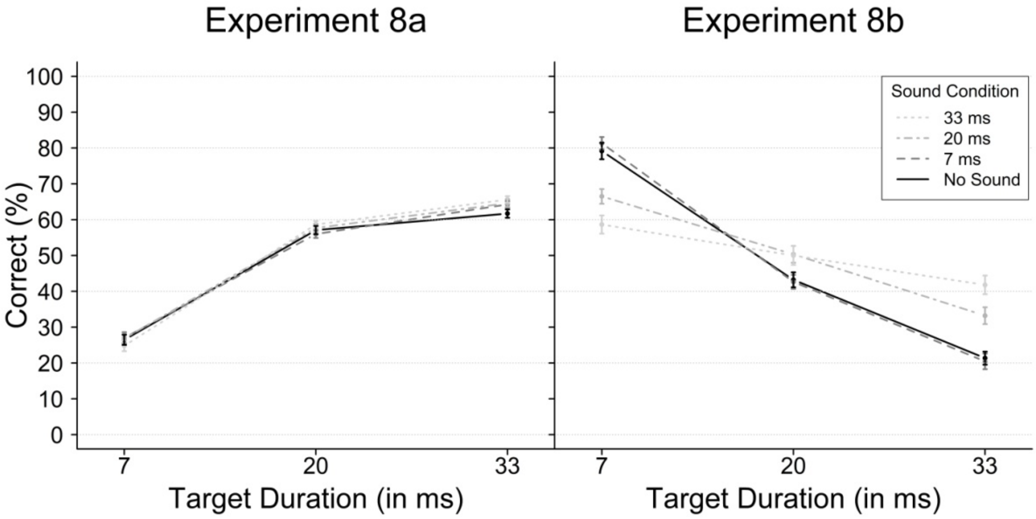

Figure 8. Mean percentage of correct responses in Experiment 8a (letter identification task) and Experiment 8b (interval identification task) as a function of Target Duration and Sound Condition.

6.4. Exploring the sustained effects of multisensory duration integration

The last main core of this thesis was to investigate whether or not the observed multisensory integration effects on perceived duration would produce sustained effects over time. To this end, Experiment 9 consisted of an adaptation task similar to that employed by Heron et al. (2013). Specifically, this task was divided into two different phases: an adaptation phase and a test phase. In the adaptation phase, participants were, for around three minutes, exposed to a stream of audiovisual filled intervals (similar to those employed in Experiment 3) with a constant temporal discrepancy between the modalities. Specifically, visual intervals (753 ms) were accompanied by auditory intervals of either *longer* (+100 ms), *same*, or *shorter* (-100 ms) duration. Each of these multimodal bias conditions was presented in a separate adaptation phase. In the test phase following each adaptation phase, three top-up presentations of these audiovisual intervals were followed by the unimodal presentation of a visual interval with a variable duration of either 500, 753, or 1000 ms. The participants´ task was to reproduce the duration of these unimodally presented visual intervals. If the conflicting durations presented during the adaptation phase would lead to conflict adaptation (e.g., if the audiovisual conflict was solved by a perceptual “expansion” or “shrinking” of the perceived visual interval duration), this should thus result in sustained changes in duration perception for the subsequently presented unimodal intervals. As Heron et al. (2013) hypothesized, if multisensory integration precedes duration adaptation, repulsive duration aftereffects should be observed in the reproductions of the test phase. Accordingly, adapting to “longer” or “shorter” perceived visual duration induces a shortening or lengthening of perceived test duration, respectively

(Heron et al., 2012). However, if multisensory integration operates after duration adaptation, no repulsive aftereffects would be observed in the test phase. Finally, another possibility might be that perceived duration is recalibrated by reducing the multisensory conflict, but in the absence of duration adaptation. Consequently, sustained multisensory effects should be observed but in the form of perceptual attraction (e.g., adapting to “longer” perceived visual durations induces a lengthening of perceived visual unimodal test duration).

As described previously, Heron et al. (2013) did not observe any evidence for sustained effects of multisensory presentation. In the present Experiment 9, we employed a set of experimental conditions which might presumably be even more effective to produce multisensory biasing effects than Heron et al.´s. Specifically, we employed a proportionally smaller audiovisual conflict, that is the auditory bias in relation to visual duration (e.g., (853 ms – 753 ms)/753 ms = 0.13). In Experiment 3, these values had resulted in a relatively large multisensory bias effect (~70 % of the physically presented conflict). In comparison, in Heron et al.´s study the proportional conflict was much larger (0.59), which, in the light of the evidence of a “window of integration”, may in turn have reduced the effectiveness of multimodal integration. In fact, the observed multimodal bias in Heron et al.’s immediate condition was rather small (~32%). According to this comparison between both studies, our set of experimental conditions should lead to even more effective multisensory integration, and thus, maximize the potential to also observe sustained effects of such integration.

Despite these methodological differences, the results of Experiment 9 are in line with those of Heron et al. (2013). The audiovisual conflict presented during the adaptation phase did not evoke any change in reproductions of the unimodally presented visual intervals during the test phase (see Figure 9, left panel). Therefore, no sustained effect of multisensory integration was observed, neither in the form of repulsion nor attraction. In order to interpret this observed null result as being due to the lack of sustained effects of multisensory integration, it is important, however, to ascertain immediate effects of multisensory integration within the same paradigm. Specifically, the repeated presentation of conflicting intervals during the adaptation phase might be actually detrimental to multisensory integration per se. For example, the prolonged exposure to consistent audiovisual conflict might have led participants to change their multisensory processing strategy from combined to separate processing of the modalities during adaptation (Mahani, Sheybani, Bausenhart, Ulrich, & Ahmadabadi, 2017). That is, participants might have started to perceive the temporally incongruent intervals as separate, independent stimuli rather than as a combined audiovisual interval percept. Second, it has been shown that the repeated perceptual exposure to an audiovisual conflicting stimulus pair narrows the temporal window of multisensory integration in a simultaneity judgment task and accordingly, might have also affected multisensory integration in the present experiment (Powers, Hillock, & Wallace, 2009).

Finally, it has been demonstrated that the proportion of congruent vs. incongruent audiovisual intervals affects multisensory integration. Specifically, a smaller multisensory integration effect is observed as the proportion of incongruent trials increases (Sarmiento et al., 2012). Therefore, it is possible that the prolonged exposure of conflicting audiovisual intervals might have hampered multisensory integration and consequently, neither immediate nor sustained effects of integration would be observed after the duration adaptation phase.

In order to address this issue, Experiment 10 was conducted as a control experiment. It was identical to Experiment 9 except for one crucial difference: in the test phase, each visual interval was also accompanied by an auditory interval with longer, same, or shorter duration (with the same conflict as in the adaptation phase). If the long exposure to multisensory conflict during the adaptation phase prevents multisensory integration, then no auditory biasing effect on perceived visual duration should be observed during the test phase. Otherwise, if such an auditory biasing effect is observed, then this would imply that multisensory integration was also preserved in Experiment 9, but without evoking any sustained effect on the perceived test duration. The results were clear-cut: Conflicting auditory intervals presented in the test phase along with the visual intervals led to clear biasing effects on perceived visual duration (see Figure 9, right panel), of comparable magnitude as those observed in Experiment 3. This result shows that the repeated exposure to a multimodal conflict did not hamper multisensory integration of audiovisual intervals in the present experimental setup. Therefore, this strengthens the interpretation according to which an immediate multisensory integration presumably emerged during the adaptation phase, but without evoking any sustained effects of this integration in the unimodal test phase of Experiment 9.

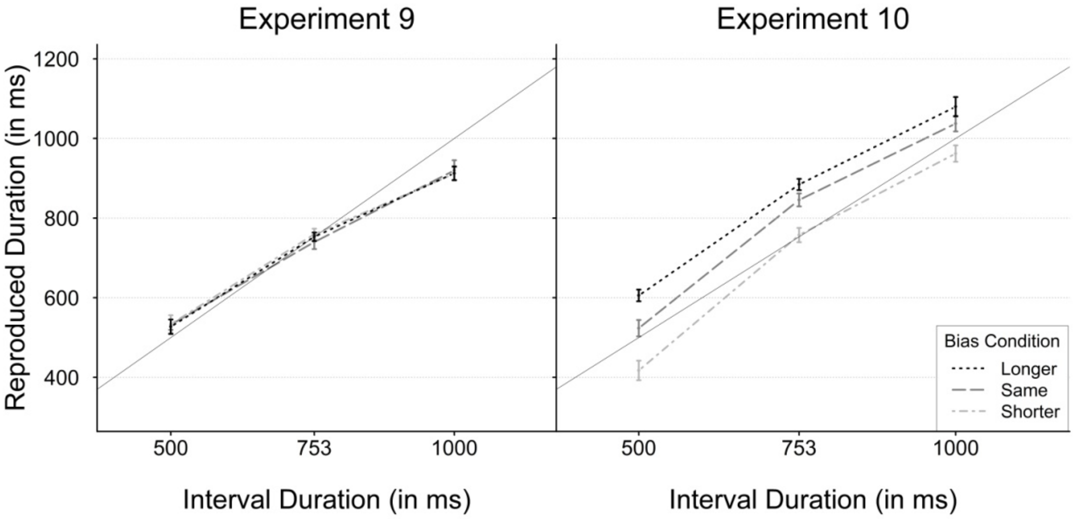

Figure 9. Mean reproduced duration as a function of Bias Condition and Interval Duration in Experiment 9 (left panel) and Experiment 10 (right panel). The continuous black line depicts the veridical line.

7. General discussion

The present thesis was conducted with the aim to extend previous evidence on different aspects regarding the underlying mechanisms and principles of multisensory integration of perceived duration. Specifically, a series of 10 experiments were conducted according to four different experimental core questions.

First, I aimed to investigate whether the previously observed impairment of visual duration discrimination, when visual intervals were accompanied by auditory conflicting intervals, was due to a “true” biasing effect on perceived visual duration towards the auditory one, or to a decrease of discrimination sensitivity caused by the incongruent information from the different modalities. The results presented in this thesis replicated and extended previous studies (Klink et al., 2011; Romei et al., 2011) by showing that relevant interval durations were clearly biased towards the duration of concurrently presented irrelevant conflicting intervals. Specifically, the perceived duration of relevant visual intervals was strongly biased towards the incongruent duration of concurrent irrelevant auditory intervals: Visual intervals were perceived as longer or shorter, when accompanied by longer or shorter auditory intervals, respectively. More recently, independent evidence for multimodal integration effects on perceived duration were provided by Sarmiento et al. (2012) as well as Asaoka and Gyoba (2016), thus supporting the respective results presented in the present thesis.

Interestingly, in contrast to previous studies showing that multimodal perception is predominantly determined by the auditory temporal information (K. M. Chen & Yeh, 2009;

Klink et al., 2011; Ortega, Guzman-Martinez, Grabowecky, & Suzuki, 2014; Walker & Scott, 1981), Experiment 1 also showed a rather bidirectional, although non-symmetrical, multisensory integration effect. Even though auditory conflicting durations led to a more pronounced biasing effect on perceived visual duration, irrelevant visual interval duration also influenced perceived auditory duration, although with a smaller bias magnitude than the auditory one. Therefore, contrary to previous suggestions about the auditory dominance in the time domain (K. M. Chen & Yeh, 2009; Walker & Scott, 1981) conforming to the modality appropriateness model of multimodal integration, the present results might be consistent with a cue combination model, for example based on statistically optimal integration of the modalities inputs according to their reliability. Recently, a study by Hartcher-O'Brian et al. (2014) provided independent and more direct evidence for this notion by explicitly varying signal reliability in the auditory modality.

On the other hand, our results also extend previous evidence regarding the effects of multisensory integration on discrimination sensitivity (e.g., Experiments 1 and 4). Assessing complete psychometric functions in a duration discrimination task allowed for dissociating between the effects of multisensory integration on perceived duration and sensitivity. The corresponding results showed that the observed biasing effect caused by incongruent temporal information did not lead to a decrease of sensitivity for duration discrimination. In fact, the presence of conflicting auditory intervals even led to an improvement in sensitivity compared to unimodal visual presentation. Thus, the presentation of auditory intervals with conflicting temporal information did not cause additional costs, but rather benefits regarding duration discrimination sensitivity. Superficially, this finding contrasts with the results of previous studies such as, for example, Romei et al. (2011). These authors showed an increase in perceptual sensitivity for bimodal *congruent* intervals in comparison to unimodal visual ones, while bimodal *incongruent* intervals led to a decrease in perceptual sensitivity in comparison to unimodal visual ones. However, in Romei et al.'s study (2011), the correct response always corresponded to the condition where irrelevant auditory intervals involved congruent duration (e.g., the short visual interval was always accompanied by the short auditory interval), while the wrong responses always corresponded to the condition where the irrelevant auditory intervals involved incongruent durations (e.g., the short visual interval was always accompanied by the long auditory interval and vice versa). Thus, auditory congruent intervals increased the probability for a correct response, and therefore, led to enhanced performance. Likewise, auditory incongruent intervals decreased the probability for a correct response, and therefore, led to impaired performance, as compared to the unimodal visual condition. The observed results of the present thesis, however, overcame this methodological limitation by showing dissociable effects of multisensory integration on perceived duration and discrimination sensitivity.

The second main core was to investigate the potential mechanisms underlying these multisensory integration effects, with regard to a prominent model of duration perception, that is, the internal clock model of the scalar expectancy theory (Gibbon, 1977; Gibbon et al., 1984). To this end, in Experiments 2, 3, and 4, the magnitude of the auditory biasing effect on perceived visual duration was assessed across a wide range of interval durations, in order to distinguish whether the multimodal bias can be attributed to changes in pacemaker rate or to changes in the timing of the switch mechanism. The results of Experiment 2 and 3 failed to provide clear evidence on this question, since the magnitude of the multimodal bias unexpectedly decreased across interval duration. Presumably this result was due to limitations of the employed duration reproduction method. However, Experiment 4 overcame these methodological limitations by the use of a paired comparison method. It showed that the difference in perceived duration of visual intervals, depending on whether they were accompanied by longer or shorter auditory intervals, remained constant across a large range of interval durations. This finding is highly consistent with the assumption that multimodal integration of conflicting durations affects the switch component of a pacemaker-accumulator mechanism (Burle & Casini, 2001), which might be mediated by a temporal ventriloquist effect (Bertelson & Aschersleben, 2003; Klink et al., 2011; Morein-Zamir et al., 2003). According to this mechanism, a temporal ventriloquist effect would cause the perceived onset and offset of the visual interval to be drawn towards the onset and offset of the auditory interval. Consequently, the audiovisually integrated onset and offset, and not the unimodal visual onset and offset, determine the closing and opening of the switch component, and thus, the number of accumulated pulses that determine duration perception.

In addition to this mechanism proposed for the integration of temporally incongruent intervals from different modalities, it should be noted that Experiments 2 and 3 provide also evidence for an additional multisensory effect on the pacemaker rate. This effect, however, is independent of interval congruency, but rather is mediated by interval type. Specifically, filled auditory intervals presumably evoke immediate arousal und thus may increase the pacemaker rate (Wearden et al., 2007). This increased pacemaker rate then also determines the perceived duration of multimodal filled intervals, again suggesting relative auditory dominance on the resulting audiovisual duration percept. In sum, multisensory integration of interval duration may affect both the switch mechanism (i.e., resolving conflicting information from different modalities) as well as the pacemaker rate (i.e., immediate arousal from the auditory modality dominates the audiovisual duration percept). Thus, this study provides important evidence on the potential mechanisms underlying the multisensory integration of audiovisual duration.

Moreover, these results are relevant for the investigation of time perception per se. Specifically, there is ample evidence regarding the effects of manipulating the pacemaker´s pulse rate on time perception (K. M. Chen & Yeh, 2009; Penton-Voak et al., 1996; Wearden et al., 1998), however research on the role of the switch mechanism on time perception remains scarcer. The closing/opening of the switch component is typically assumed to be elicited directly by the perceived onset/offset of the stimulus. Therefore, the timing of stimulus onset/offset cannot be manipulated independently of the physical interval duration (e.g., if the offset is presented later, physical stimulus duration will consequently also be longer). This issue inherent to the manipulation of the switch component might have actually prevented so far a more direct investigation of the role of the switch mechanism for time perception (Bendixen, Grimm, & Schröger, 2006). Therefore, establishing the temporal ventriloquism-like effect for conflicting multimodal intervals is relevant not only for multimodal integration but eventually may also provide a novel experimental method to manipulate and study the switch component of the internal clock.

The next core of this thesis consisted of exploring the perceptual vs. decisional processes that might underlie the multimodal integration of perceived duration. Several experimental results are pertinent to this question. First, a clear multisensory duration integration effect was observed even when a high proportion of unimodal trial was randomly interspersed within the bimodal trials. Under such conditions, responding based only on the auditory temporal information might have been especially complicated given that it would have required frequent (and unpredictable) intermodal attentional shifts – which have been repeatedly shown to impair duration discrimination performance (Grondin, Ivry, Franz, Perreault, & Metthé, 1996; Rousseau, Poirier, & Lemyre, 1983; Ulrich et al., 2006). Therefore, it is quite unlikely that participants in the present experiments based their responses only on the task-irrelevant modality. Second, an auditory biasing effect was observed even when the concurrently presented irrelevant sounds did not provide any information about interval duration. Specifically, longer (shorter) durations were perceived when a single sound pulse either preceded (followed) the onset or followed (preceded) the offset of the task-relevant visual interval compared to a simultaneous sound presentation. Given that auditory stimulation alone did not provide duration information, it is not possible that participants could have relied exclusively on the auditory modality in order to estimate visual durations. Therefore, this result provides important evidence against a pure decisional-bias account of the multisensory integration effects on perceived duration. However, it should be noted that visual intervals were in fact judged longer when auditory pulses accompanied the onset rather than the offset of the visual intervals, independent of their relative temporal asynchrony with the respective visual marker. This suggests that in addition to a genuine multimodal biasing effect, an attentional or decisional effect may contribute to the observed results. For example, the auditory stimulus accompanying the visual onset might have acted as a warning or accessory signal (cf. Hackley & Valle-Inclan, 1998; Niemi & Näätänen, 1981), which might in turn open an attentional gate as specified by the attentional-gate model of time perception (Zakay & Block, 1997). Accordingly, the presentation of the auditory stimulus together with the visual onset would produce a shortening of the perceptual latency of interval onset, and therefore, increase its perceived duration (Bausenhart, Rolke, Seibold, & Ulrich, 2010; Grondin & Rammsayer, 2003; Seifried, Ulrich, Bausenhart, Rolke, & Osman, 2010). However, it is unlikely that such an attentional mechanism also caused the multisensory bias evoked by the different audiovisual delays, since this bias was observed independently of whether the auditory pulse accompanied the onset or the offset of the visual intervals. Importantly, perceived visual duration was prolonged even when the auditory pulse followed the offset of the visual interval. This cannot be explained on basis of an attentional process but rather points to a perceptual multisensory integration effect mediated by temporal ventriloquism.

The conclusion of perceptual changes evoked by multisensory integration also receives independent support from the observation of statistically optimal cue combination in duration perception (Hartcher-O’Brien et al., 2014). Nonetheless, a more direct investigation of the potential perceptual processes underlying the multisensory duration integration was conducted in the presented work by exploring the effect of manipulating the duration of concurrent sounds on masked letter identification. First, the results showed that the mere presentation of sounds enhanced letter identification performance in comparison to the unimodal presentation of the target letter. This sound-induced enhancement not only replicates the basic finding of Y. C. Chen and Spence´s (2011) study but also extended this result to the identification of centrally presented letters followed by random-dot noise masks, as opposed to lateral presentation of the letters followed by a structurally similar masking letter as employed in Y. C. Chen and Spence´s study. Importantly, this result points to early audiovisual interactions at a perceptual processing stage, since letter identification is enhanced by the concurrently presentation of uninformative sounds which did not provide any information about the identity of the target letters. Consequently, a simple bias towards responding to the information provided in the auditory modality cannot explain the sound-induced facilitation of letter identification. In contrast to the effect elicited by the sound presentation, however, varying the duration of the sounds did not differentially modulate letter identification performance, even when two important prerequisites for such an effect could be demonstrated in Experiment 8. Specifically, directly manipulating the duration of the letters strongly affected identification performance (Exp. 8a). Moreover, manipulating the duration of the concurrent sounds affected judgments of letter duration (Exp. 8b). This indicates that the physical multimodal discrepancy was large enough to produce multimodally biased perceived duration, and that changes in duration of the visual representation should, in principle, affect letter identification. Still, and contrary to the results of de Haas et al. (2013) regarding contrast detection, no evidence was observed supporting potential changes in the persistence of the letter representation (and thus, letter identification performance) produced by sound duration.

One feasible interpretation of this result, which stands in contrast to the results outlined above, is that the multisensory integration effect on perceived duration is not perception-based but rather due to a decisional bias, that is, participants respond on basis of the auditory duration information when they have to perform visual duration judgments. Since auditory duration information is uninformative regarding letter identity, participants cannot utilize this information for letter identification. Yet, other accounts should not be refuted, based on such a null result. First, it has to be acknowledged that a sound-duration-induced benefit might have been undetected due to lack of experimental power. Specifically, the possibility remains that an effect of sound duration on letter identification may emerge when even larger audiovisual discrepancies are employed. For example, de Haas et al. (2013), observed significantly enhanced performance for a sound 48 ms longer than the 24 ms visual target. Yet, such large discrepancies might also hamper multisensory integration effects through a violation of the assumption of unity (Y. C. Chen & Spence, 2017; Welch & Warren, 1980). Furthermore, in the present paradigm, presenting longer sounds would have led to a temporal overlap with the target and the mask, which might have hampered perceptual segregation of the two stimuli and thus hindered any benefit by the sound presentation (Y. C. Chen & Spence, 2011). A more direct replication of the paradigm employed by Y. C. Chen and Spence, which involved much longer ISI durations, might enable the testing of such larger audiovisual asynchronies, and thus shed light on this issue.

Speculatively, it may also be possible that the mechanisms underlying low-level visual perception as contrast detection and those involved in higher-level perceptual tasks such as letter identification are differentially modulated by sound duration. Nevertheless, given the dissociable effects of sound duration on identification of letter duration and letter identity within the present experiments, it can be readily suggested that at least for the present conditions of stimulation, both types of judgments are not based on a common, unified object representation. Accordingly, the duration and identity of a visual perceptual object seem to be separable object features that can be differentially modulated by the duration of concurrently presented sounds.

At this point, other experimental methods might be helpful to explore the potential perceptual multisensory integration of perceived duration more directly. For example, electrophysiological measures as perception-related ERP components or difference potentials, as the MMN, may inform about the locus of the integration of audiovisual conflicting intervals durations. For example, employing the MMN component to the study of the spatial ventriloquist effect (Colin, Radeau, Soquet, Dachy, et al., 2002; Stekelenburg et al., 2004) has provided crucial evidence in favour of potential underlying audiovisual interactions at relatively early, pre-decisional stages of processing. Specifically, it has been shown that an illusory auditory location shift, produced by a physical visual location shift, evoked an auditory MMN in frontal areas typically implicated in auditory deviance processing. Similarly, one might investigate whether, for example, an illusory shift of perceived duration caused by manipulating the duration of a certain modality can produce an MMN at electrode sites corresponding to deviance processing in another modality, and compare this “illusory” MMN with a physical duration deviant within this modality.

Finally, the last aim of the present thesis was to explore whether the multisensory duration integration effect on perceived duration can persist over time, after prolonged exposure to audiovisual conflict. The results of this investigation showed that adaptation to audiovisual conflicting durations did not affect perceived duration of subsequently presented unimodal visual intervals. Moreover, adaptation to audiovisual conflicting intervals did not prevent multisensory integration per se, since a strong multimodal biasing effect was observed on reproduced duration of subsequently presented audiovisual conflicting intervals. These findings do not only replicate the results of Heron et al. (2013), but also extend these previous results by showing multisensory integration effects on perceived duration even after prolonged exposure to the audiovisual conflict. Regarding the locus of multisensory duration integration, observing any sustained effects might have provided evidence in favour of multimodal integration taking place during perceptual (low-level) processing, since such sustained multisensory effects on unisensory processing could of course not be attributed to a biased response towards (nonexistent) information from a task-irrelevant modality. Yet, even though there is no direct evidence for a perceptual locus of multisensory duration integration, this possibility cannot also not be ruled out on the basis of the negative (null) results of the present Exp. 9 and Heron et al.´s recalibration experiments. Nonetheless, one might wonder why sustained effects for multimodal conflicting durations were not observed, because such sustained effects have been very consistently demonstrated for the processing of other temporal aspects, as for example, audiovisual asynchrony probed in temporal order judgment tasks (e.g., Di Luca et al., 2009; Fujisaki et al., 2004; Hanson et al., 2008; Harrar & Harris, 2008; Heron et al., 2010, 2007; Keetels & Vroomen, 2007; Machulla et al., 2012; Vatakis et al., 2007; Vroomen et al., 2004). Presumably, an explanation can be based on the type of conflicting stimulation employed for assessing multisensory duration integration. For example, in previous studies of temporal order recalibration, participants are adapted to a consistent conflict between two brief stimuli (e.g., an auditory pulse always slightly preceding a visual pulse). The modalities' perceptual latencies then are recalibrated to resolve this persistent conflict and thus, subsequently, sustained changes in the perceptual latencies can be assessed. However, for the case of interval duration, there is also a consistent audiovisual conflict, however, between the durations of the interval modalities. That is, the onset and offset of the intervals are both consistently biased, but in opposing directions (e.g., for longer auditory intervals, the auditory onset precedes the visual one and the auditory offset follows the visual one). This opposing temporal asynchrony between onset and offset signals might be resolved, as outlined above, via temporal ventriloquism and thus cause a biased immediate perceived duration. However, at the same time, the opposing temporal conflicts for on- and offset might prevent effective recalibration of the perceptual latencies and consequently any sustained effects as temporal order recalibration. In sum, this explanation fits reasonably well with the results of Experiments 1 and 4, which suggest that multimodal integration of duration is indeed mediated by temporal ventriloquism operating on the interval on- and offsets. In other words, what is integrated are the perceived time points, which demarcate the start and end of temporal intervals, but not the representation of the intervals per se.

In summary, the present thesis extends previous evidence on different aspects involved in multisensory integration of the perceived duration of audiovisual conflicting intervals. First, multimodal integration of conflicting audiovisual duration leads to strong biasing effect on perceived duration. Moreover, it has been demonstrated that the multisensory duration integration is rather bidirectional, although asymmetrical: perceived duration of auditory intervals was also biased towards concurrent conflicting visual durations although to a lesser extent than vice versa. Second, crucial evidence has been provided regarding the mechanisms underlying this multimodal integration effect of perceived conflicting durations. Specifically, our results are consistent with the notion that incongruent multimodal intervals are integrated through temporal ventriloquism, which mediates the closing/opening of the switch component of an internal clock. Regarding the locus of the processes underlying these multisensory integration effects, direct evidence of perceptual audiovisual interaction could not be provided in the presented thesis. On the one hand, multisensory duration integration occurred even when a high proportion of unimodal intervals were employed. Moreover, clear multisensory bias effects on visual perceived duration have been observed even when the concurrent irrelevant sounds did not provide information about the interval duration. These results argue against a mere strategic or decisional bias (i.e., attending or responding to the task-irrelevant interval modality) underlying the observed multisensory duration integration effects. On the other hand, neither sustained effects of multisensory integration of conflicting information nor a modulation of letter identification performance through the duration of accompanying sounds were observed. Such effects, if present, would clearly suggest a perceptual locus of the underlying multisensory integration effect. Yet, the absence of such effects cannot be likewise interpreted as evidence for a post-perceptual effect. Therefore, the research question on the locus of multisensory duration integration still remains open and should be further investigated in future research.

8. References

Alais, D., & Burr, D. (2004). Ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262. https://doi.org/10.1016/S0960-9822(04)00043-0

Asaoka, R., & Gyoba, J. (2016). Sounds modulate the perceived duration of visual stimuli via crossmodal integration. *Multisensory Research*, *29*, 319–335. https://doi.org/10.1163/22134808-00002518

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*, 1391–1397. https://doi.org/10.1364/JOSAA.20.001391

Bausenhart, K. M., Bratzke, D., & Ulrich, R. (2016). Formation and representation of temporal reference information. *Current Opinion in Behavioral Sciences*, *8*, 46–52. https://doi.org/10.1016/j.cobeha.2016.01.007

Bausenhart, K. M., De la Rosa, M. D., & Ulrich, R. (2014). Multimodal integration of time: Visual and auditory contributions to perceived duration and sensitivity. *Experimental Psychology*, *61*, 310–322. https://doi.org/10.1027/1618-3169/a000249

Bausenhart, K. M., Rolke, B., Seibold, V. C., & Ulrich, R. (2010). Temporal preparation influences the dynamics of information processing: Evidence for early onset of information accumulation. *Vision Research*, *50*, 1025–1034. https://doi.org/10.1016/j.visres.2010.03.011

Bendixen, A., Grimm, S., & Schröger, E. (2006). The relation between onset, offset, and duration perception as examined by psychophysical data and event-related brain potentials. *Journal of Psychophysiology*, *20*, 40–51. https://doi.org/10.1027/0269-8803.20.1.40

Bertelson, P. (1999). Ventriloquism: A case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (Vol. 129, pp. 347–362). Amsterdam: Elsevier B.V. https://doi.org/10.1016/S0166-4115(99)80034-X

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin and Review*, *5*, 482–489. https://doi.org/10.3758/BF03208826

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension---1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, *50*, 147–155. https://doi.org/10.1016/S0167-8760(03)00130-2

Bertelson, P., Frissen, I., Vroomen, J., & de Gelder, B. (2006). The aftereffects of ventriloquism: Patterns of spatial generalization. *Perception and Psychophysics*, *68*, 428–436. https://doi.org/10.3758/BF03193687

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*, 578–584. https://doi.org/10.3758/BF03207374

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception and Psychophysics*, *62*, 321–332. https://doi.org/10.3758/BF03205552

Burle, B., & Casini, L. (2001). Dissociation between activation and attention effects in time estimation: Implications for internal clock models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 195. https://doi.org/10.1037/0096-1523.27.1.195

Calvert, G., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Chen, K. M., & Yeh, S. L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychologica*, *130*, 225–234. https://doi.org/10.1016/j.actpsy.2008.12.008

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

Chen, Y. C., & Spence, C. (2011). The crossmodal facilitation of visual object representations by sound: Evidence from the backward masking paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1784–1802. https://doi.org/10.1037/a0025638

Chen, Y. C., & Spence, C. (2017). Assessing the role of the "unity assumption" on multisensory integration: A review. *Frontiers in Psychology*, *8*, 445. https://doi.org/10.3389/fpsyg.2017.00445

Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002). Electrophysiology of spatial scene analysis: The mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clinical Neurophysiology*, *113*, 507–518. https://doi.org/10.1016/S1388-2457(02)00028-7

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*, 495–506. https://doi.org/10.1016/S1388-2457(02)00024-X

de Haas, B., Cecere, R., Cullen, H., Driver, J., & Romei, V. (2013). The duration of a co-occurring sound modulates visual detection performance in humans. *PLoS ONE*, *8*. https://doi.org/10.1371/journal.pone.0054789

De la Rosa, M. D., & Bausenhart, K. M. (2013). Multimodal integration of interval duration: Temporal ventriloquism or changes in pacemaker rate? *Timing and Time Perception*, *1*, 189–215. https://doi.org/10.1163/22134468-00002015

Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, *9*, 1–16. https://doi.org/10.1167/9.12.7

Dyjas, O., Bausenhart, K. M., & Ulrich, R. (2014). Effects of stimulus order on duration discrimination sensitivity are under attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 292–307. https://doi.org/10.1037/a0033611

Einstein, A., Speziali, P., & Besso, M. (1979). *Correspondance, 1903-1955*. Paris: Hermann.

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

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*, 332–343. https://doi.org/10.1007/s00221-002-1262-y

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2003). The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychologica*, *113*, 315–327. https://doi.org/10.1016/S0001-6918(03)00043-X

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2005). The aftereffects of ventriloquism: Generalization across sound-frequencies. *Acta Psychologica*, *118*, 93–100. https://doi.org/10.1016/j.actpsy.2004.10.004

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*, 773–778. https://doi.org/10.1038/nn1268

Gamache, P. L., & Grondin, S. (2010). Sensory-specific clock components and memory mechanisms: Investigation with parallel timing. *European Journal of Neuroscience*, *31*, 1908–1914. https://doi.org/10.1111/j.1460-9568.2010.07197.x

Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception*, *36*, 1089–1103. https://doi.org/10.1068/P5741

Getzmann, S., & Lewald, J. (2014). Modulation of auditory motion processing by visual motion early cross-modal interactions in human auditory cortices. *Journal of Psychophysiology*, *28*, 82–100. https://doi.org/10.1027/0269-8803/a000113

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*, 279–325. https://doi.org/10.1037/0033-295X.84.3.279

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, *423*, 52–77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x

Goldstone, S., & Lhamon, W. T. (1972). Auditory-visual differences in human temporal judgment. *Perceptual and Motor Skills*, *34*, 623–633. https://doi.org/10.2466/pms.1972.34.2.623

Greene, B. (2004). *The fabric of the cosmos: Space, time, and the texture of reality*. New York: Alfred A. Knopf.

Grondin, S., Ivry, R. B., Franz, E., Perreault, L., & Metthé, L. (1996). Markers' influence on the duration discrimination of intermodal intervals. *Perception & Psychophysics*, *58*, 424–433. https://doi.org/10.3758/BF03206818

Grondin, S., & Rammsayer, T. (2003). Variable foreperiods and temporal discrimination. *Quarterly Journal of Experimental Psychology*, *56A*, 731–765. https://doi.org/10.1080/02724980244000611

Hackley, S. A., & Valle-Inclan, F. (1998). Automatic alerting does not speed late motoric processes in a reaction-time task. *Nature*, *391*, 786–788. https://doi.org/10.1038/35849

Hanson, J. V. M., Heron, J., & Whitaker, D. (2008). Recalibration of perceived time across sensory modalities. *Experimental Brain Research*, *185*, 347–352. https://doi.org/10.1007/s00221-008-1282-3

Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research*, *186*, 517–524. https://doi.org/10.1007/s00221-007-1253-0

Hartcher-O'Brien, J., Di Luca, M., & Ernst, M. O. (2014). The duration of uncertain times: Audiovisual information about intervals is integrated in a statistically optimal fashion. *PLoS ONE*, *9*, e89339. https://doi.org/10.1371/journal.pone.0089339

Heron, J., Aaen-Stockdale, C., Hotchkiss, J., Roach, N. W., McGraw, P. V., & Whitaker, D. (2012). Duration channels mediate human time perception. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 690–698. https://doi.org/10.1098/rspb.2011.1131

Heron, J., Hotchkiss, J., Aaen-stockdale, C., Roach, N. W., & Whitaker, D. (2013). A neural hierarchy for illusions of time: Duration adaptation precedes multisensory integration. *Journal of Vision*, *13*, 1–12. https://doi.org/10.1167/13.14.4.doi

Heron, J., Roach, N. W., Whitaker, D., & Hanson, J. V. M. (2010). Attention regulates the plasticity of multisensory timing. *European Journal of Neuroscience*, *31*, 1755–1762. https://doi.org/10.1111/j.1460-9568.2010.07194.x

Heron, J., Whitaker, D., McGraw, P. V., & Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision*, *7*, 1–8. https://doi.org/10.1167/7.13.5

Kayser, S. J., Philiastides, M. G., & Kayser, C. (2017). Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. *NeuroImage*, *148*, 31–41. https://doi.org/10.1016/j.neuroimage.2017.01.010

Keetels, M., Stekelenburg, J., & Vroomen, J. (2007). Auditory grouping occurs prior to intersensory pairing: Evidence from temporal ventriloquism. *Experimental Brain Research*, *180*, 449–456. https://doi.org/10.1007/s00221-007-0881-8

Keetels, M., & Vroomen, J. (2007). No effect of auditory-visual spatial disparity on temporal recalibration. *Experimental Brain Research*, *182*, 559–565. https://doi.org/10.1007/s00221-007-1012-2

Kinsbourne, M., & Warrington, E. K. (1962). The effect of an after-coming random pattern on the perception of brief visual stimuli. *Quarterly Journal of Experimental Psychology*, *14*, 223–234. https://doi.org/10.1080/17470216208416540

Klink, P. C., Montijn, J. S., & van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention, Perception & Psychophysics*, *73*, 219–236. https://doi.org/10.3758/s13414-010-0010-9

Le Poidevin, R. (2000). The experience and perception of time. In *The Standford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Lejeune, H., & Wearden, J. H. (2009). Vierordt's the experimental study of the time sense (1868) and its legacy. *European Journal of Cognitive Psychology*, *21*, 941–960. https://doi.org/10.1080/09541440802453006

Lewald, J. (2002). Rapid adaptation to auditory-visual spatial disparity. *Learning and Memory*, *9*, 268–278. https://doi.org/10.1101/lm.51402

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioural Brain Research*, *121*, 69–79. https://doi.org/10.1016/S0166-4328(00)00386-7

Machulla, T. K., Di Luca, M., Froehlich, E., & Ernst, M. O. (2012). Multisensory simultaneity recalibration: Storage of the aftereffect in the absence of counterevidence. *Experimental Brain Research*, *217*, 89–97. https://doi.org/10.1007/s00221-011-2976-5

Mahani, M. A. N., Sheybani, S., Bausenhart, K. M., Ulrich, R., & Ahmadabadi, M. N. (2017). Multisensory perception of contradictory information in an environment of varying reliability: Evidence for conscious perception and optimal causal inference. *Scientific Reports*, *7*, 3167. https://doi.org/10.1038/s41598-017-03521-2

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. https://doi.org/10.1038/264746a0

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*, 389–391. https://doi.org/10.1126/science.6867718

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*, 640–662. https://doi.org/10.1152/jn.1986.56.3.640

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, *17*, 154–163. https://doi.org/10.1016/S0926-6410(03)00089-2

Näätänen, R., & Alho, K. (1995). Mismatch negativity-a unique measure of sensory processing in audition. *International Journal of Neuroscience*, *80*, 317–337. https://doi.org/10.3109/00207459508986107

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, *89*, 133–162. https://doi.org/10.1037/0033-2909.89.1.133

Noesselt, T., Bergmann, D., Hake, M., Heinze, H. J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, *1220*, 157–163. https://doi.org/10.1016/j.brainres.2007.12.060

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H. J., & Driver, J. (2010). Sound-induced enhancement of low-intensity vision: Multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *Journal of Neuroscience*, *30*, 13609–13623. https://doi.org/10.1523/JNEUROSCI.4524-09.2010

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, *65*, 123–132. https://doi.org/10.3758/BF03194789

Ortega, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2014). Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Attention, Perception & Psychophysics*, *76*, 1485–1502. https://doi.org/10.3758/s13414-014-0663-x

Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*, 307–320. https://doi.org/10.1037/0097-7403.22.3.307

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

Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs - Thomas (1941) revisited. *Psychological Research*, *49*, 17–22. https://doi.org/10.1007/BF00309198

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences*, *95*, 869–875. https://doi.org/10.1073/pnas.95.3.869

Rolke, B. (2008). Temporal preparation facilitates perceptual identification of letters.

Perception & Psychophysics, *70*, 1305–1313. https://doi.org/10.3758/PP.70.7.1305

Romei, V., De Haas, B., Mok, R. M., & Driver, J. (2011). Auditory stimulus timing influences perceived duration of co-occurring visual stimuli. *Frontiers in Psychology*, *2*, 1–8. https://doi.org/10.3389/fpsyg.2011.00215

Rousseau, R., Poirier, J., & Lemyre, L. (1983). Duration discrimination of empty time intervals marked by intermodal pulses. *Perception & Psychophysics*, *34*, 541–548. https://doi.org/10.3758/BF03205907

Sarmiento, B. R., Shore, D. I., Milliken, B., & Sanabria, D. (2012). Audiovisual interactions depend on context of congruency. *Attention, Perception, & Psychophysics*, *74*, 563–574. https://doi.org/10.3758/s13414-011-0249-9

Seifried, T., Ulrich, R., Bausenhart, K. M., Rolke, B., & Osman, A. (2010). Temporal preparation decreases perceptual latency: Evidence from a clock paradigm. *The Quarterly Journal of Experimental Psychology*, *63*, 2432–2451. https://doi.org/10.1080/17470218.2010.485354

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*, 2200–2208. https://doi.org/10.1016/j.neuroimage.2011.03.075

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 2000. https://doi.org/10.1038/35048669

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*, 147–152. https://doi.org/10.1016/S0926-6410(02)00069-1

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *NeuroReport*, *12*, 3849–3852. https://doi.org/10.1097/00001756-200112040-00049

Shi, Z., Chen, L., & Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: The influence of time interval. *Experimental Brain Research*, *203*, 723–735. https://doi.org/10.1007/s00221-010-2286-3

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*, 7–10. https://doi.org/10.1097/00001756-200101220-00009

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*, 61–70. https://doi.org/10.1250/ast.28.61

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. https://doi.org/10.1111/nyas.12121

Starke, J., Ball, F., Heinze, H. J., & Noesselt, T. (2017). The spatio-temporal profile of multisensory integration. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13753

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*, 497–506. https://doi.org/10.1162/jocn.1996.8.6.497

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. *Cognitive Neuroscience*. Cambridge, MA, US: MIT Press.

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, *357*, 163–166. https://doi.org/10.1016/j.neulet.2003.12.085

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, *80*, 1. https://doi.org/10.1037/h0033872

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Crossmodal temporal discrimination: Assessing the predictions of a general pacemaker-counter model. *Perception & Psychophysics*, *68*, 1140–1152. https://doi.org/10.3758/BF03193716

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*, 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*, *181*, 173–181. https://doi.org/10.1007/s00221-007-0918-z

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception and Psychophysics*, *69*, 744–756. https://doi.org/10.3758/BF03193776

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen [The experimental study of the time sense]*. Tübingen, Germany: Laupp.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1583–1590. https://doi.org/10.1037//0096-1523.26.5.1583

Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 513–518. https://doi.org/10.1037/0096-1523.30.3.513

Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: no role in temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1063–1071. https://doi.org/10.1037/0096-1523.32.4.1063

Vroomen, J., & Keetels, M. (2009). Sounds change four-dot masking. *Acta Psychologica*, *130*, 58–63. https://doi.org/10.1016/j.actpsy.2008.10.001

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, and Psychophysics*, *72*, 871–884. https://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. *Cognitive Brain Research*, *22*, 32–35. https://doi.org/10.1016/j.cogbrainres.2004.07.003

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1327–1339. https://doi.org/10.1037/0096-1523.7.6.1327

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why “sounds are judged longer than lights”: Application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology*, *51B*, 97–120. https://doi.org/10.1080/713932672

Wearden, J. H., Norton, R., Martin, S., & Montford-Bebb, O. (2007). Internal clock processes and the filled-duration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 716–729. https://doi.org/10.1037/0096-1523.33.3.716

Welch, R. B. (1999). Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (Vol. 129, pp. 371–387). Amsterdam: Elsevier B.V. https://doi.org/10.1016/S0166-4115(99)80036-3

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*, 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Sensory processes and perception* (Vol. I, p. 25.1-25.36). New York: Wiley.

Zakay, D., & Block, R. A. (1997). Temporal Cognition. *Current Directions in Psychological Science*, *6*, 12–16. https://doi.org/10.1111/1467-8721.ep11512604

9. Appendices

APPENDIX A – Study 1

Bausenhart, K. M., De la Rosa, M. D., & Ulrich, R. (2014). Multimodal integration of time: Visual and auditory contributions to perceived duration and sensitivity. *Experimental Psychology*, *61*, 310–322. https://doi.org/10.1027/1618-3169/a000249

Research Article

Multimodal Integration of Time

Visual and Auditory Contributions to Perceived Duration and Sensitivity

Karin M. Bausenhart, Maria Dolores de la Rosa, and Rolf Ulrich

University of Tübingen, Germany

Abstract. Recent studies suggest that the accuracy of duration discrimination for visually presented intervals is strongly impaired by concurrently presented auditory intervals of different duration, but not vice versa. Because these studies rely mostly on accuracy measures, it remains unclear whether this impairment results from changes in perceived duration or rather from a decrease in perceptual sensitivity. We therefore assessed complete psychometric functions in a duration discrimination task to disentangle effects on perceived duration and sensitivity. Specifically, participants compared two empty intervals marked by either visual or auditory pulses. These pulses were either presented unimodally, or accompanied by task-irrelevant pulses in the respective other modality, which defined conflicting intervals of identical, shorter, or longer duration. Participants were instructed to base their temporal judgments solely on the task-relevant modality. Despite this instruction, perceived duration was clearly biased toward the duration of the intervals marked in the task-irrelevant modality. This was not only found for the discrimination of visual intervals, but also, to a lesser extent, for the discrimination of auditory intervals. Discrimination sensitivity, however, was similar between all multimodal conditions, and only improved compared to the presentation of unimodal visual intervals. In a second experiment, evidence for multisensory integration was even found when the task-irrelevant modality did not contain any duration information, thus excluding noncompliant attention allocation as a basis of our results. Our results thus suggest that audiovisual integration of temporally discrepant signals does not impair discrimination sensitivity but rather alters perceived duration, presumably by means of a temporal ventriloquism effect.

Keywords: multisensory integration, perceived duration, discrimination sensitivity, temporal ventriloquism

The mechanisms underlying the perception of time have fascinated and puzzled researchers since the advent of psychology. Maybe one of the most intriguing properties of time perception is that there is no specific sense organ dedicated to the perception of the passage of time. Nonetheless, intervals presented to the auditory modality are generally judged as being longer and are discriminated with a higher accuracy than intervals presented to the visual modality (Goldstone & Lhamon, 1972; Ulrich, Nitschke, & Rammsayer, 2006; Walker & Scott, 1981; Wearden, Edwards, Fakhri, & Percival, 1998). This and related results have sometimes led to the suggestion that the cognitive representation of time is closely linked to auditory processing (Bratzke, Seifried, & Ulrich, 2012; Guttman, Gilroy, & Blake, 2005; Kanai, Lloyd, Bueti, & Walsh, 2011).

If such a close relationship between the auditory modality and internal representations of time exists, it should also become evident when participants have to judge the duration of multimodal time intervals. Interestingly, even though the perceived durations of unimodal auditory and visual intervals differ, we typically do not experience any conflict in perceived duration when we are exposed to a natural event which combines auditory and visual features. Accordingly, there seems to be a process of multimodal integration for the perception of duration, just as has been demonstrated for other perceptual domains, as the processing of spatial origin (e.g., Bertelson & Radeau, 1981; Slutsky & Recanzone, 2001) or simultaneity and successiveness (e.g., Burr, Banks, & Morrone, 2009; Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). In order to investigate this phenomenon, Walker and Scott (1981) conducted a series of duration reproduction experiments. They could not only establish the typical difference in perceived duration of auditory and visual intervals, with unimodal auditory intervals being reproduced longer than visual intervals of physically identical duration, but also, they showed that the reproduced duration for combined audiovisual intervals of identical duration closely resembled the reproduced duration for unimodally presented auditory intervals. Accordingly, these results suggest an auditory dominance over vision for time reproduction. Similar results have also been demonstrated more recently by Chen and Yeh (2009) within an oddball experiment.

A potential complication in the interpretation of the experiments outlined above is that they all employed auditory and visual stimuli of physically identical duration. It may be more promising for the study of audiovisual dominance to present compound stimuli with slightly conflicting durations, and to investigate how and in favor of

Experimental Psychology 2014; Vol. 61(4):310–322
DOI: 10.1027/1618-3169/a000249

which modality the resulting perceptual conflict is resolved. Actually, Klink, Montijn, and van Wezel (2011) have recently demonstrated a deterioration of duration discrimination accuracy for visually presented filled intervals, when concurrently presented filled auditory intervals provided incongruent duration information. Specifically, participants had to identify the longer interval of a pair of filled intervals in a certain (relevant) modality (e.g., visual or auditory). The difficulty of this discrimination task was chosen to be close to each participant's individual duration discrimination threshold, as measured in a pretest. In addition, intervals were also presented in the irrelevant modality. The duration of these irrelevant intervals was in conflict to the relevant ones, that is, the long relevant interval was paired with a shorter irrelevant one, and vice versa. The results show that discrimination performance (i.e., percentage of correct identification of the longer interval) for visual relevant intervals was strongly impaired by the information from the irrelevant auditory modality. When audition was the relevant modality, irrelevant visual stimuli did not cause such a performance decrement. Thus, this result again demonstrates an auditory dominance in the time domain, that is, an (irrelevant) auditory interval can affect perception of a relevant visual interval, but not vice versa.

More experimental support for these results stems from demonstrations that discrimination performance for visual intervals decreased when task-irrelevant auditory intervals provided conflicting information, but increased when the auditory intervals provided congruent information (Romei, De Haas, Mok, & Driver, 2011; Sarmiento, Shore, Milliken, & Sanabria, 2012). Taken together, on one hand, these studies provide evidence for multimodal integration of interval duration by demonstrating performance decrements within the visual modality, when biasing information is provided in the auditory modality. Clearly, such performance decrements are a feasible indicator of whether multisensory integration of interval duration occurred or not. On the other hand, however, such results seem to be at odds with recent comprehensive accounts of multimodal integration in other task domains (Alais & Burr, 2004; Ernst & Banks, 2002; Heron, Whitaker, & McGraw, 2004), which proceed from the notion that combining information of two separate modality channels leads to a *more stable* perceptual representation, and thus should enable performance *benefits* in discrimination tasks. In other words, the integrated representation of multimodally presented, incongruent time intervals may be biased toward the interval duration of one of the modalities (and consequently, be perceived shorter or longer than the actual interval duration in the other modality), but this representation should also be rather stable, and thus result in a high discrimination sensitivity (see also Gamache & Grondin, 2010). Importantly, the dependent measures employed by previous studies in order to indicate performance decrements (i.e., percentage of correct responses in Klink et al., 2011; *d′* in Romei et al., 2011)[1] might have been affected by changes in perceptual bias, sensitivity, or both (see also General Discussion). Therefore, previous studies cannot disentangle multisensory integration effects on perceived duration and on discrimination sensitivity, and in addition, they do not provide any quantitative measure of the magnitude of the multisensory bias effect.

Therefore, the goal of the present study is to extend previous evidence on multimodal integration of *incongruent versus congruent* multimodal time intervals by assessing psychometric functions for the discrimination of such multimodally presented intervals. First, with this procedure, the amount of perceptual bias and changes in sensitivity of perception can be assessed separately as the location (as quantified by the Point of Subjective Equality, PSE) and the spread (as quantified by the Just Noticeable Difference, JND) of the psychometric function, respectively. Based on the results of previous studies (Klink et al., 2011; Romei et al., 2011; Walker & Scott, 1981), a strong influence of auditory interval duration on the discrimination of visually presented time intervals is expected. Second, since the assessment of full psychometric functions allows for a fine-grained analysis of even small effects on either of these measures, this experiment will also provide a sensitive test of whether there is complete auditory dominance of audition over vision, as suggested by previous results, or whether there is also a multisensory integration effect in the reverse direction, that is, whether visually presented time intervals can also affect perceived auditory duration.

Experiment 1

Method

Participants

Twelve participants took part in this experiment. One participant of the original sample had to be replaced later due to poor task performance, which resulted in an estimation of JND larger than the sample mean + 2 *SD*. The final sample included nine women, eight right-handers, and the mean age was 22.0 (*SD* = 2.1) years.

---

[1] In principle, Signal Detection Theory allows for computing *d′* as a measure of sensitivity and *c* as a measure of the response criterion (i.e., bias). In a forced choice procedure as employed by Romei et al. (2011), however, *c* will only inform about a potential position bias, that is, whether participants preferably respond with one of the response alternatives "first signal longer" or "second signal longer" (cf. Wickens, 2001). In the study of Romei et al. no such position bias was present. Most importantly, *d′* depends on the difference between the *z*-transforms of the hit rate and the false alarm rate. These values, however, might have been affected by genuine decreases in sensitivity as well as biases in perceived duration. For example, incongruent audiovisual stimulus pairs might lead to especially noisy internal stimulus representations. Alternatively, they might affect relative perceived duration, such that the physically short visual stimulus appears to be the long one and vice versa. Both of these effects would affect hit and false alarm rates, and thus, *d′*.

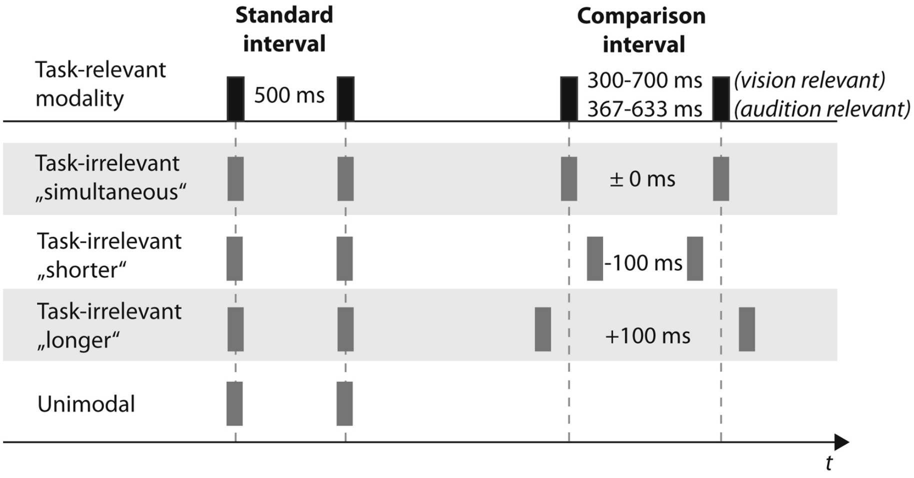

Figure 1. Temporal positions of relevant and irrelevant markers for the standard and the comparison interval in the four bias conditions of Experiment 1.

Stimuli and Apparatus

All stimuli were generated and presented on a Macintosh Computer using the Psychtoolbox extensions (Brainard, 1997; Pelli, 1997) for Matlab. Visual stimuli were presented on a CRT monitor running at a refresh rate of 150 Hz, and all auditory stimuli were presented binaurally via headphones. Participants were seated in a distance of approximately 45 cm from the screen.

A white dot (diameter 1 mm) served as fixation and a white disk (diameter 8 mm) served as visual onset and offset marker. They were presented at the screen center in white (80 cd/m^2) on a black background (< 1 cd/m^2). A 800 Hz sine tone with an amplitude of 72 db SPL served as auditory onset and offset marker. Its duration was 13.3 ms, with onset and offset ramps of 5 ms.

Procedure

Each participant performed two experimental sessions, one with a visual and the other with an auditory duration discrimination task. Each session took approximately 1 hr 20 minutes and the order of the sessions was counterbalanced across participants.

In the visual task session, the task-relevant onset and offset markers of the standard and comparison intervals were presented in the visual modality. Specifically, each trial started with the presentation of a fixation dot at the center of the screen. After 1,000 ms, the fixation dot was replaced with a white disk (standard onset marker) for 13.3 ms (cf. Figure 1, first row). The 500 ms standard interval began after disk offset: During this interval, the screen remained empty. The end of the standard interval was marked by the onset of a second white disk (standard offset marker), which remained on the screen for 13.3 ms. Subsequently, the fixation dot was presented for an interstimulus interval of 1,000 ms. Then, again, a white disk was presented for 13.3 ms (comparison onset marker). At disappearance of the disk, the comparison interval started and it ended with the appearance of a fourth white disk (comparison offset marker, 13.3 ms). The duration of the comparison interval varied randomly from trial to trial and could take one of the following values: 300, 340, 380, 420, 460, 500, 540, 580, 620, 660, or 700 ms. After the offset marker of the comparison interval, the fixation dot reappeared until participants gave their response. Participants were to press with their left and right index finger either the "y" or the "m" key of a standard German keyboard, in order to indicate whether the first or the second presented interval was the longer one, respectively.

In the auditory task session, the course of each trial was virtually identical, with only two exceptions: First, the task-relevant onset and offset markers of the standard and the comparison intervals were presented in the auditory modality (13.3 ms sine tone pulses) instead of the visual modality. Second, the duration of the comparison intervals varied between 367, 394, 420, 447, 474, 500, 527, 554, 580, 607, and 634 ms in order to compensate for the typically better temporal resolution of the auditory modality.

In order to manipulate multisensory bias, pulses were also presented in the task-irrelevant modality. Specifically, in the visual duration discrimination condition, the relevant visual pulses were accompanied by task-irrelevant auditory ones, and in the auditory duration discrimination condition, the relevant auditory pulses were accompanied by task-irrelevant visual pulses.

In each trial, irrelevant pulses were presented simultaneously with the task-relevant onset and offset markers of the standard interval duration. Along with the comparison interval, the task-irrelevant pulses were presented according to four different bias conditions (cf. Figure 1): First, in "unimodal" trials, no task-irrelevant pulses were presented with the task-relevant markers of the comparison interval. Second, in "simultaneous" trials, task-irrelevant pulses were presented simultaneously with the task-relevant comparison interval markers. In "shorter" trials, task-irrelevant pulses were presented 50 ms *after* the task-relevant onset marker and 50 ms *before* the task-relevant offset marker of the comparison interval, respectively. In "longer" trials, task-irrelevant pulses were presented 50 ms *before* the task-relevant onset marker and 50 ms *after* the task-relevant offset marker of the comparison interval, respectively. All four trial types were presented in random order.

Participants were instructed to ignore the task-irrelevant stimulation and base their judgment on the task-relevant modality only. In order to emphasize this instruction, a relatively high proportion of unimodal trials (50% of all trials) was presented. Since these trials contained no information about the duration of the comparison interval in the task-irrelevant modality, in 50% of all trials it would have been impossible for participants to accomplish their task by relying on the task-irrelevant information alone. The remaining 50% of trials were equally split into "simultaneous," "longer," and "shorter" trials. Overall, 20 practice trials (picked at random from all possible trials) and 990 experimental trials were administered in each task session, thus yielding 45 trials per comparison level in unimodal trials and 15 trials per comparison level in each of the "simultaneous," "shorter," and "longer" conditions. After each block of 66 trials, written feedback containing the percentage of correct responses in the last block was presented and participants could take a self-terminated break.

Results and Discussion

For each participant, Task Modality, and Bias Condition, the proportion of "comparison longer" responses (P) was computed separately for all comparison durations (x_i). Then, a logistic psychometric function (cf. Bush, 1963),

$$P(x_i) = \frac{1}{1 - \mathrm{e}^{-(x_i - \mathrm{PSE})/(0.91 \times \mathrm{JND})}},$$

was fitted to these data by employing a maximum-likelihood procedure, and by using the MATLAB function "fminsearch," which is an implementation of the Nelder-Mead algorithm (Nelder & Mead, 1965). From this function, PSE and JND can be derived. PSE corresponds to the comparison duration at which the participant perceives comparison and standard to be equal in duration and thus, PSE reflects the perceived duration of the comparison stimulus. JND amounts to half the interquartile range of the psychometric function and thus reflects the steepness of the psychometric function, which is a measure of the sensitivity with which a participant can discriminate between the standard and the comparison stimulus. The steeper the psychometric function, and thus, the lower the value of JND, the higher is discrimination sensitivity. Figure 2 depicts the observed response frequencies and fitted psychometric functions for an exemplary participant, in order to illustrate a typical data pattern.

Separate repeated-measures ANOVAs with the factors Bias Condition (shorter, simultaneous, longer, and unimodal) and Task Modality (visual vs. auditory) were then conducted on the estimates of PSE and JND. The ANOVA on PSE revealed a strong main effect of Bias Condition, $F(3, 33) = 80.71$, $p < .001$, $\eta_p^2 = 0.88$, as well as an interaction between Task Modality and Bias Condition, $F(3, 33) = 42.55$, $p < .001$, $\eta_p^2 = 0.79$. PSE was not affected by Task Modality alone, $F(1, 11) = 1.14$, $p = .31$, $\eta_p^2 = 0.09$. In order to examine the observed interaction between Task Modality and Bias Condition more closely, separate repeated-measures ANOVAs with factor Bias Condition were conducted for the two task modalities.

As expected, for the visual duration discrimination task, there was a strong effect of Bias Condition on PSE, $F(3, 33) = 68.06$, $p < .001$, $\eta_p^2 = 0.86$ (cf. Figure 3). Importantly, a larger PSE corresponds to a shorter perceived duration, that is, the presented comparison must be relatively long in order to be perceived as being equal to the duration of the standard stimulus. Likewise, a relatively small PSE corresponds to a relatively long perceived duration. Thus, the observed pattern of results clearly indicates that perceived visual interval duration was strongly biased toward the duration of the task-irrelevant auditory interval. Bonferroni-corrected post hoc comparisons showed that PSE in the "shorter" condition (555 ms) was larger than PSE in the "simultaneous" (468 ms), the "unimodal" (451 ms), and in the "longer" (397 ms) conditions (all ps < .001). Likewise, PSE in the "longer" condition was smaller than PSE in the "simultaneous" and the "unimodal" condition (all ps < .001). The only nonsignificant post hoc comparison was the one between PSE in the "simultaneous" and in the "unimodal" condition ($p = 1.00$). In addition, one-sample t-tests confirmed that PSEs in the "simultaneous" and the "unimodal" conditions were smaller than the standard duration, $t(11) = 6.46$, $p < .001$ and $t(11) = 3.38$, $p < .01$, respectively, reflecting a typical negative time-order error (e.g., Allan, 1977; Hellström, 1985).

For the auditory duration discrimination task, there was also a clear effect of Bias Condition on PSE, $F(3, 33) = 10.04$, $p < .001$, $\eta_p^2 = 0.48$. As can be seen in Figure 3, the direction of this effect is similar, although its magnitude is less pronounced, to the effect observed for the visual task condition. PSEs in the "longer," "simultaneous," "shorter," and "unimodal" conditions were 460, 472, 482, and 481 ms, respectively. Multiple Bonferroni-corrected post hoc t-tests showed that PSE in the "longer" condition was smaller than in the "simultaneous" ($p = .01$), "shorter" ($p < .01$), and "unimodal" ($p = .02$) conditions. No other post hoc comparisons were significant, despite a slight tendency for PSE in the "shorter" condition to be larger than in the "simultaneous" condition ($p = .13$), all other ps > .50. In sum, these results show that visually presented irrelevant intervals can also affect the perceived duration of auditory intervals, although to a lesser extent than in the reverse direction. Additional one-sample t-tests confirmed that PSEs in the "simultaneous" and the "unimodal" conditions were smaller than the standard duration, $t(11) = 6.80$, $p < .001$, and $t(11) = 2.88$, $p = .01$, respectively, again pointing to a negative time-order error.

In order to assess the effects of multisensory integration on discrimination sensitivity, a further repeated-measures ANOVA with factors Task Modality and Bias Condition was conducted on JND estimates. As can be seen in Figure 4, JND was smaller in the auditory (36 ms) than in the visual (52 ms) task modality, $F(1, 11) = 14.27$, $p = .003$, $\eta_p^2 = 0.56$. Also, Bias Condition affected JND

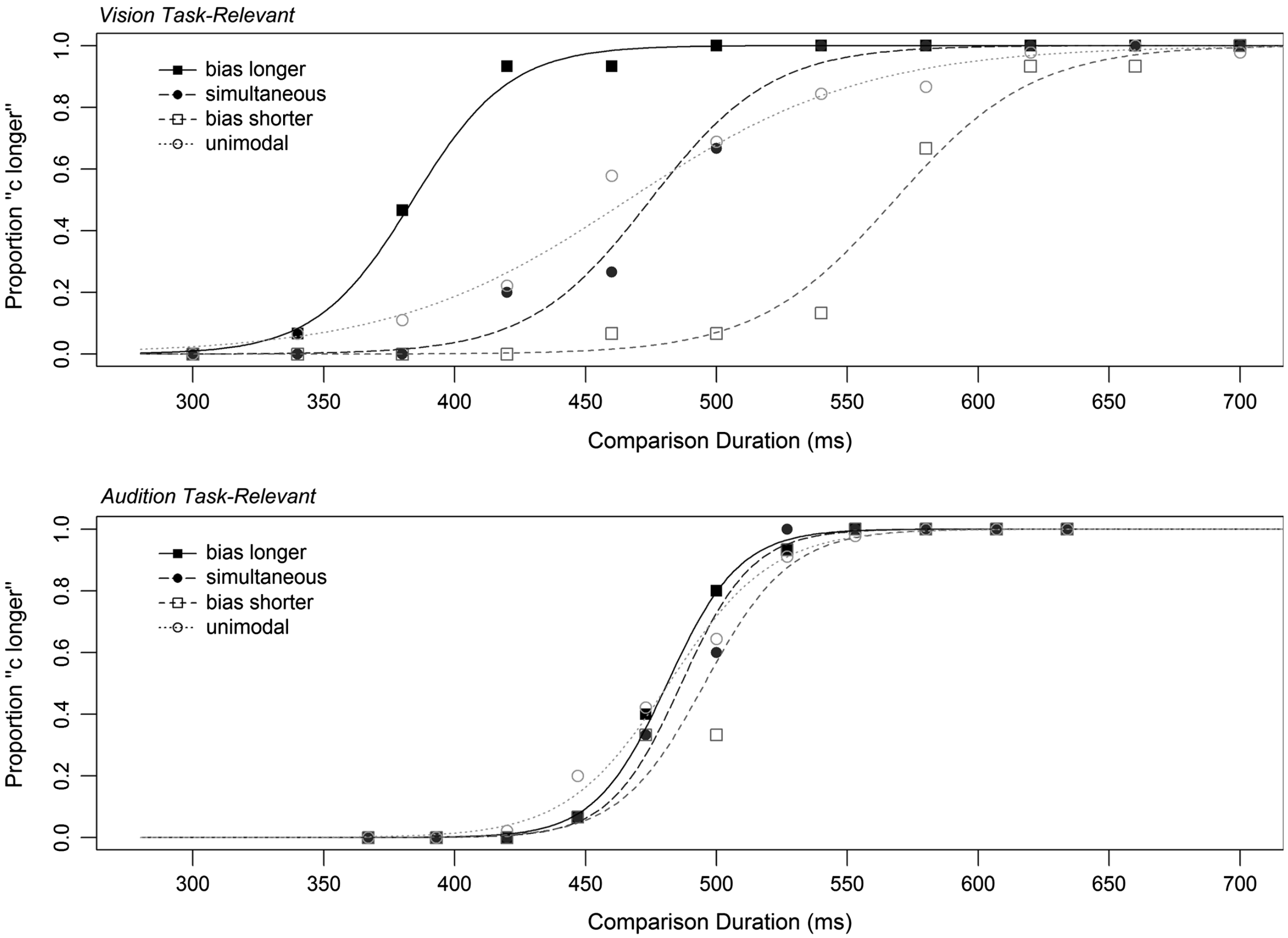

Figure 2. Proportion of ''comparison longer'' responses and fitted psychometric functions for a single participant in Experiment 1. Depicted are all bias conditions for the visual (upper panel) and the auditory (lower panel) task condition.

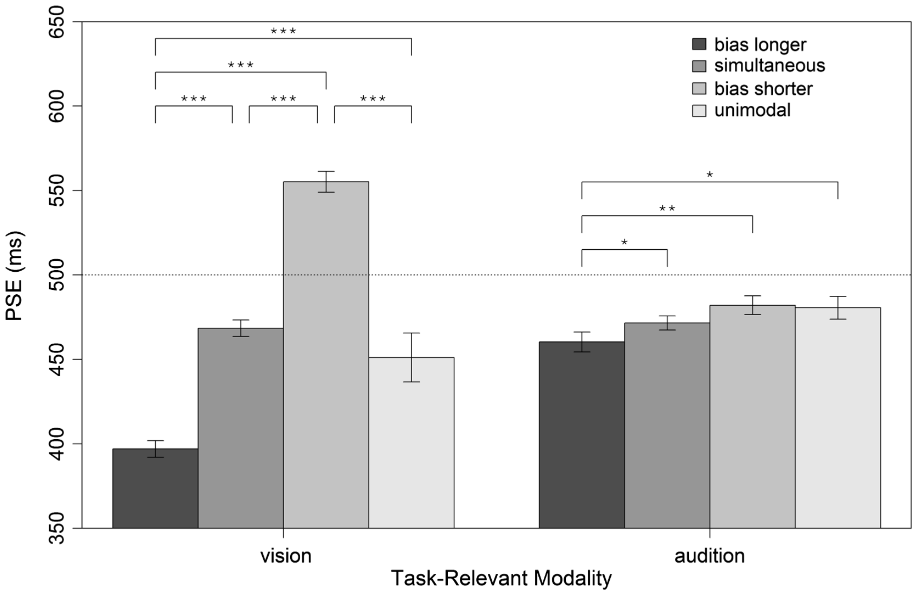

Figure 3. Mean Point of Subjective Equality (PSE) in ms ± 1 standard error of the mean, depending on task modality and bias condition in Experiment 1. Asterisks indicate the significant results of Bonferroni-corrected post hoc comparisons which were conducted separately for each task condition.

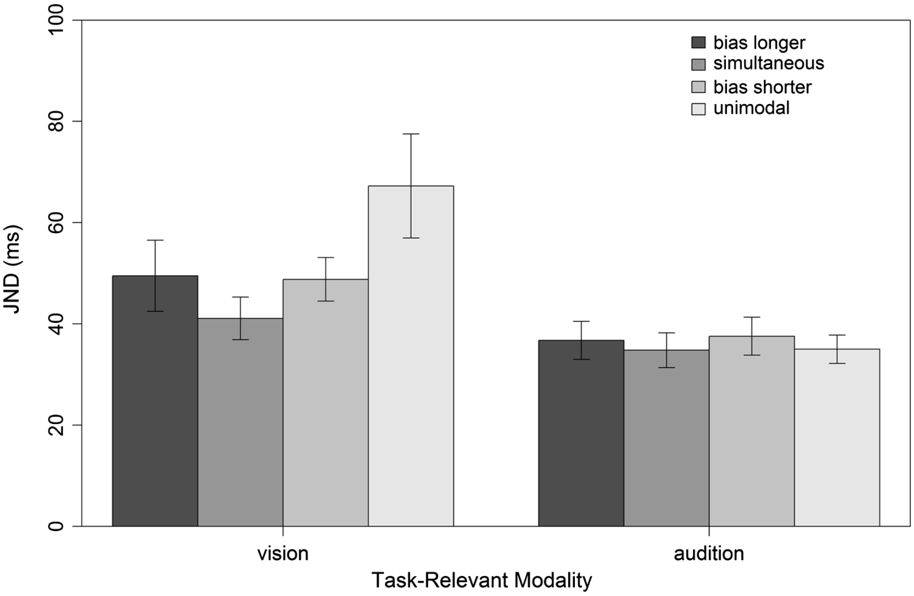

Figure 4. Mean Just Noticeable Difference (JND) in ms ± 1 standard error of the mean, depending on task modality and bias condition in Experiment 1.

estimates, $F(3, 33) = 5.15$, $p = .005$, $\eta_p^2 = 0.32$, and this effect was qualified by a Task Modality × Bias Condition interaction, $F(3, 33) = 5.41$, $p = .004$, $\eta_p^2 = 0.33$.

To further investigate this interaction, separate repeated-measures ANOVAs with factor Bias Condition on JND were conducted for the two task modalities. For the visual duration discrimination task, there was a main effect of Bias Condition, $F(3, 33) = 5.84$, $p = .003$, $\eta_p^2 = 0.35$. Bonferroni-corrected post hoc *t*-tests showed that JND tended to be larger (67 ms) in the "unimodal" condition than in the "longer" (50 ms), $p = .067$, and in the "simultaneous" (41 ms), $p = .087$, condition. No other comparisons were significant, all *p*s > .35. A corresponding ANOVA for the auditory task modality showed that there was no effect of Bias Condition on JND, $F(3, 33) = 0.71$, $p = .55$, $\eta_p^2 = 0.06$.

An additional post hoc *t*-test showed that JND in unimodal trials was considerably lower in the auditory task condition than in the visual task condition, $t(11) = 3.90$, $p = .002$. A parallel comparison for the simultaneous condition did not show any meaningful difference between the two tasks, $t(11) = 1.35$, $p = .20$.

In summary, for perceived duration it was demonstrated that auditory interval duration strongly affected perceived visual duration, such that shorter accompanying auditory intervals led to a massive underestimation of visual interval duration, and longer accompanying auditory intervals led to a massive overestimation of visual interval duration. In comparison to the perceived duration in the simultaneous condition, the under- and overestimation amounted to 87 and 71 ms, respectively, which corresponds to 87 and 71% of the physical duration difference (±100 ms) between relevant and irrelevant intervals. Interestingly, this influence was not unidirectional, as a strict auditory dominance account would suggest, but could also be observed – even though to a much reduced extent – for the discrimination of auditory intervals, when these were accompanied by irrelevant visually marked intervals. Here, the conflicting interval duration (±100 ms) caused over- and underestimation of approximately 10–12%. Thus, perceived audiovisual duration was not exclusively determined by the perceived auditory duration, but also codetermined by the visual duration information. This argues for the notion that multisensory integration of duration information involves a weighting process which combines information from different modalities to different extents (see also Klink et al., 2011), as has been previously suggested for other types of information as, for example, in audiovisual location perception, or visuo-haptic height perception (Alais & Burr, 2004; Ernst & Banks, 2002; Heron et al., 2004).

When regarding the mean PSE value observed in the "unimodal" condition, one might be led to conclude that there is an overestimation of the unimodal comparison stimulus with respect to the bimodal standard stimulus. This seems to be inconsistent with previous studies showing that the perceived duration of unimodal visual intervals is typically shorter than perceived duration of audiovisual intervals (Chen & Yeh, 2009; Walker & Scott, 1981). However, in this regard it is important to note that in a classical reminder design, as employed in the present study, absolute values of PSE should not be interpreted, because they may be subject to response bias or time-order error (Allan, 1977; Hellström, 1985). Such biases clearly seem to be present in our study, because even in the "simultaneous" condition, the comparison seemed to be overestimated in comparison to the – also "simultaneous" and thus identical – standard stimulus. Therefore, one should interpret PSE values obtained in a reminder design exclusively in relation to other PSE values which are assessed against the same standard stimulus. For this reason, our results concerning the unimodal condition cannot be interpreted as incongruent with the results of previous studies. Most importantly, our study was specifically designed to investigate the integration of multimodal congruent versus incongruent intervals. All our conclusions regarding this multimodal biasing effect on perceived duration are based on relative comparisons of PSE values between these biasing conditions, and are thus justified by our experimental design.

Interestingly, the pronounced biasing effect of the "longer" and "shorter" conditions on perceived duration was not accompanied by a corresponding decline in discrimination sensitivity. Specifically, JNDs were comparable between the "longer," "shorter," and "simultaneous" bias conditions, and this was true irrespective of whether vision or audition was the task-relevant modality. Only in the unimodal visual task condition, sensitivity was lower than in all other conditions. This is in line with previous studies comparing discrimination performance for unimodal auditory and visual intervals (Gamache & Grondin, 2010; Ulrich et al., 2006), which show typically lower sensitivity for visual interval discrimination than for auditory interval discrimination. Yet, even though participants were instructed to perform the duration discrimination task on basis of the visual information only, adding irrelevant auditory temporal information increased their sensitivity for the discrimination. Most remarkably, this multisensory integration benefit occurred irrespective of whether the irrelevant information was congruent with the relevant information, as in the "simultaneous" condition, or incongruent, as in the "longer" and "shorter" conditions. Thus, the integration of temporal information from multiple senses functionally seems to lead to especially stable internal stimulus representations, even if the information is not identical across the different modalities. As has been previously suggested, multisensory integration might even allow for an optimization of perceptual discrimination processes (Alais & Burr, 2004; Ernst & Banks, 2002; Ernst & Bülthoff, 2004).[2]

As should be noted, however, all present conclusions depend critically on participants' adherence to the instruction to respond to the temporal information from the task-relevant modality. As an extreme counterexample, participants might have simply closed their eyes in the visual task condition and attended to the auditory stimulation alone. Then, of course, the observed bias effects cannot be attributed to multisensory integration. To counteract the use of such harmful strategies, an especially high proportion of "unimodal" trials (50%) was administered, in which no temporal information about comparison duration was presented in the irrelevant modality. These trials were randomly interspersed and thus could be identified at the earliest at onset of the comparison stimulus. Therefore, it is rather unlikely that participants would have relied on temporal information from the irrelevant modality for their judgments. Alternatively, there might be differential response biases, such that participants are more likely to respond "longer" ("shorter") in the presence of a longer (shorter) task-irrelevant interval, which would also affect PSE. Therefore, it would be especially convincing if a multimodal integration effect on perceived duration could be demonstrated in a situation in which the irrelevant modality contains potentially biasing stimulation, but no information about interval duration per se. In Experiment 2, we created such an experimental situation.

Experiment 2

In this experiment, we aimed at ruling out an alternative explanation according to which the effects observed in Experiment 1 were due to noncompliant attention allocation rather than multisensory integration. Specifically, this alternative explanation holds that contrary to the instructions, participants responded to the auditory stimuli despite performing the visual task condition, and – at least in some of the trials – also to the visual stimuli, despite performing the auditory task condition.

Actually, many studies investigating multisensory integration are subject to similar criticism; however, there are some examples that overcome this limitation by disentangling the biasing information from the response dimension. For example, in a clever study which investigated temporal ventriloquism effects in a temporal order judgment task (Morein-Zamir et al., 2003, see also General Discussion), participants were required to judge the temporal order of two visual signals by responding whether the one above or the one below fixation appeared first. Either simultaneous to, or before and after the two visual stimuli, irrelevant sounds were presented. All sounds originated from the same spatial source and thus did not contain any information about the location of the task-relevant visual stimuli. Importantly, even though the auditory stimuli could thus not be used as a basis for the required spatial judgment, their temporal occurrence strongly affected participants' performance.

---

[2] Due to the prediction of statistical optimality of this account, one can compute the expected variance of the perception of multimodal simultaneous stimuli (here, indicated by JND) from the observed variances of the perception of each modality alone. This expected variance will be typically smaller than both observed unimodal variances. Only when the variance associated with one modality is much smaller than the one associated with the other modality, as is true for our unimodal auditory condition, multimodal variance will be close to the smallest observed unimodal variance. As can be seen in Figure 4 and our corresponding analyses, the variances from our bimodal conditions were comparable to the auditory unimodal variance and tendentially smaller than the unimodal visual variance, and thus could be principally in line with the predictions of Bayesian integration. However, studies on Bayesian integration typically concentrate on discrimination of unimodal versus bimodal congruent stimuli; not on incongruent ones as in the present study. Moreover, participants are typically informed to respond to the compound stimulus (instead of ignoring one of the modalities as in the present experiment), a condition that may likely be a prerequisite of statistically optimal integration. Therefore, we refrain from performing more specific analyses concerning the predictions of Bayesian integration. Nonetheless, future studies might try to investigate in more detail whether statistical optimality holds for the integration of durations and whether it can be observed independent of specific task instructions and incongruence of the multimodal information.

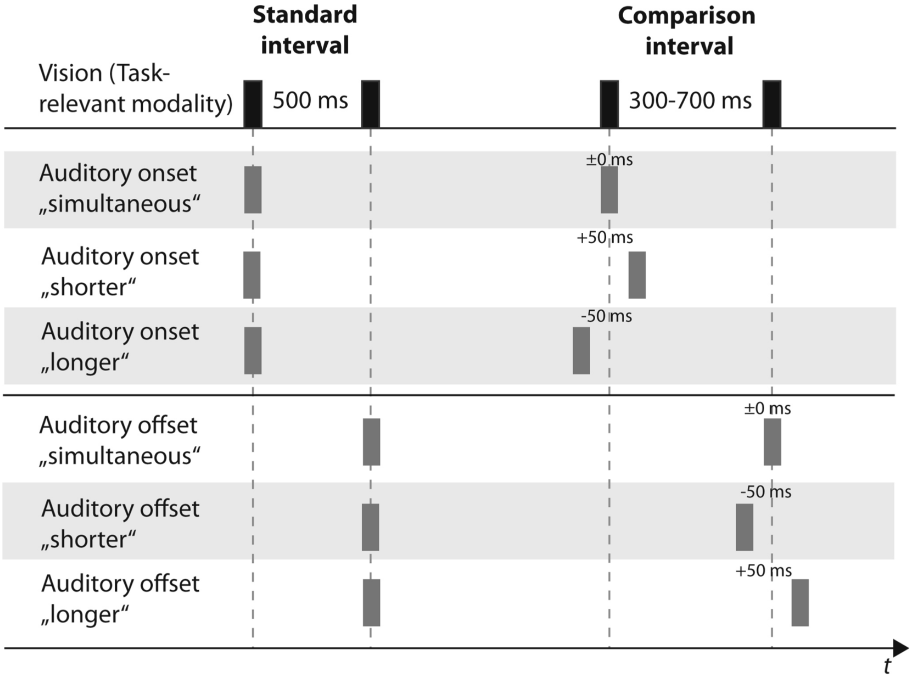

Figure 5. Temporal positions of visual and auditory markers for the standard and the comparison interval in the three bias conditions of Experiment 2.

In the present experiment, we created a similar experimental setup by providing biasing auditory information along with the visually marked intervals, yet, the auditory modality did not convey any information about interval duration. Specifically, along with the visual onset and offset markers of the standard and comparison, only one auditory pulse per interval was presented – either shortly before, simultaneous with, or after the visual onset or offset marker. Accordingly, the auditory pulse could potentially still affect the perceived occurrence of one of the visual markers, and thus perceived visual duration, but it could not per se be used as a basis for the judgment about perceived duration.

Method

Participants

Fourteen participants took part in this experiment. Two participants of the original sample had to be replaced later due to poor task performance, which resulted in an estimation of JND larger than the sample mean + 2 *SD*. The final sample included 10 women, 13 right-handers, and the mean age was 22.9 (*SD* = 2.5) years.

Stimuli and Apparatus

All stimuli and apparatus were identical to the ones employed in Experiment 1.

Procedure

The procedure was identical to the one of Experiment 1 with the following exceptions: First, only the visual task condition was tested in a single session. Second, along with each visually marked interval only one auditory marker was presented (cf. Figure 5). In half of the trials, the auditory pulses accompanied the visual onset markers, and in the other half of the trials they accompanied the visual offset markers. These ‘‘onset’’ and ‘‘offset’’ trials were presented randomly intermixed. Similar to Experiment 1, the auditory marker was always presented simultaneously with the respective visual (onset or offset) marker of the standard interval. For the comparison interval in ‘‘onset’’ trials, the auditory marker was presented with a delay of 0 ms, +50 ms, or −50 ms relative to the visual onset marker in the simultaneous, shorter, and longer condition, respectively. For the comparison interval in ‘‘offset’’ trials, the auditory marker was presented with a delay of 0 ms, −50 ms, or +50 ms relative to the visual offset marker in the simultaneous, shorter, and longer condition, respectively.

Due to this modification, the auditory modality did not convey information about interval duration on which participants might base their duration judgments. Since it was thus not necessary anymore to control for a potentially harmful strategy of responding to the auditory information alone, as a third change, the unimodal condition was omitted from the experimental design. Finally, in order to avoid that participants might infer some information about the current bias condition from the relative timing of the two remaining auditory markers, the interstimulus interval between standard offset and comparison onset was now varied randomly between 800 and 1,200 ms in steps of 6.6 ms. Each of these possible values could appear with equal probability.

Overall, participants performed 20 practice trials (picked at random from all possible trials) and 990 experimental

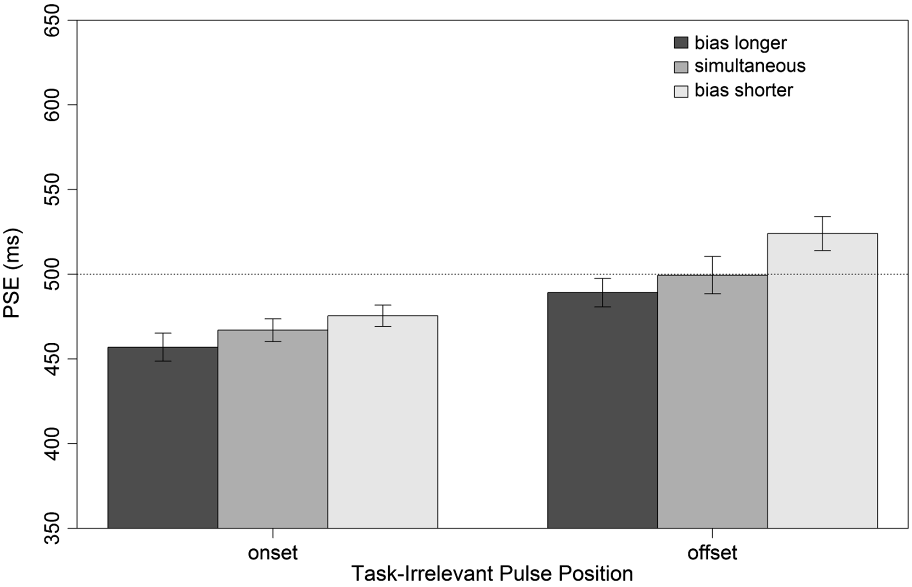

Figure 6. Mean Point of Subjective Equality (PSE) in ms ± 1 standard error of the mean, depending on pulse position and bias condition in Experiment 2.

trials. These were segregated into 15 blocks of 66 trials, each consisting of one presentation per comparison duration (11 levels) for each combination of the factors Pulse Position (onset vs. offset) and Bias Condition (simultaneous, shorter, longer). After each block, feedback about their percentage of correct responses in the last block was displayed to the participants and they could initiate the next block with a self-timed key press.

Results and Discussion

As in Experiment 1, for each participant, Pulse Position and Bias Condition, the percentage of "comparison longer" responses was computed separately for all comparison durations. Again, a logistic psychometric function (cf. Bush, 1963) was fitted to these data by employing a maximum-likelihood procedure in order to derive PSE and JND as dependent measures.

Separate repeated-measures ANOVAs with the factors Bias Condition (shorter, simultaneous, and longer) and Pulse Position (onset vs. offset) were then conducted on the estimates of PSE and JND. The ANOVA on PSE revealed a strong main effect of the position of the auditory pulse, that is, intervals were generally judged as being longer when the task-irrelevant auditory pulse accompanied the task-relevant onset markers (467 ms) than when they accompanied the task-relevant offset markers (504 ms), $F(1, 13) = 28.49$, $p < .001$, $\eta_p^2 = 0.69$ (Figure 6). Most importantly, there was again strong evidence for a multimodal bias, $F(2, 26) = 23.98$, $p < .001$, $\eta_p^2 = 0.65$, and this effect was independent of the position of the task-irrelevant auditory pulse, $F(2, 26) = 1.90$, $p = .17$, $\eta_p^2 = 0.13$. Bonferroni-corrected post hoc t-tests showed that perceived duration in the longer condition (PSE = 473 ms) was significantly longer than in the simultaneous (PSE = 483 ms, $p = .01$) and in the shorter condition (PSE = 500 ms, $p < .001$). Also, perceived duration in the simultaneous condition was longer than in the shorter condition ($p = .002$).

In addition, one-sample t-tests were conducted to compare PSE in the simultaneous condition with the standard duration (500 ms). In the onset condition, PSE was significantly below the standard duration, $t(13) = 4.93$, $p < .001$; whereas in the offset condition, PSE did not differ from the standard duration, $t(13) = 0.05$, $p = .96$.

A further ANOVA with factors Bias Condition and Pulse Position on JND estimates (cf. Figure 7) revealed that sensitivity was slightly higher when the task-irrelevant auditory pulse accompanied the task-relevant onset marker (66 ms) rather than the offset marker (75 ms), $F(1, 13) = 12.41$, $p = .004$, $\eta_p^2 = 0.49$. Bias Condition did not affect discrimination sensitivity, $F(2, 26) = 0.90$, $p = .42$, $\eta_p^2 = 0.06$, and there was no interaction of Pulse Position and Bias Condition, $F(2, 26) = 1.83$, $p = .18$, $\eta_p^2 = 0.12$.

Overall, these results are highly consistent with the outcome of Experiment 1. The timing of the single, irrelevant auditory pulse affected perceived visual duration and this was independent of whether the irrelevant auditory pulse was paired with the onset or the offset pulse of the visual intervals. Compared to the simultaneous condition, perceived duration was prolonged by 10 ms in the "longer" condition and shortened by 17 ms in the "shorter" condition. Thus, the magnitude of the perceived bias amounted to 20–34% of the physical bias (±50 ms) and thus was smaller than the one observed in Experiment 1. This is not especially surprising given that multisensory integration critically depends on the contingency of the information from the different modalities (Klink et al., 2011; Sarmiento et al., 2012). Thus, integration might be less effective when only every second visual stimulus is accompanied by an auditory one. Most important for the present purposes, however, participants would not have been able to accomplish the task by attending only to the auditory information, and the auditory information did not contain

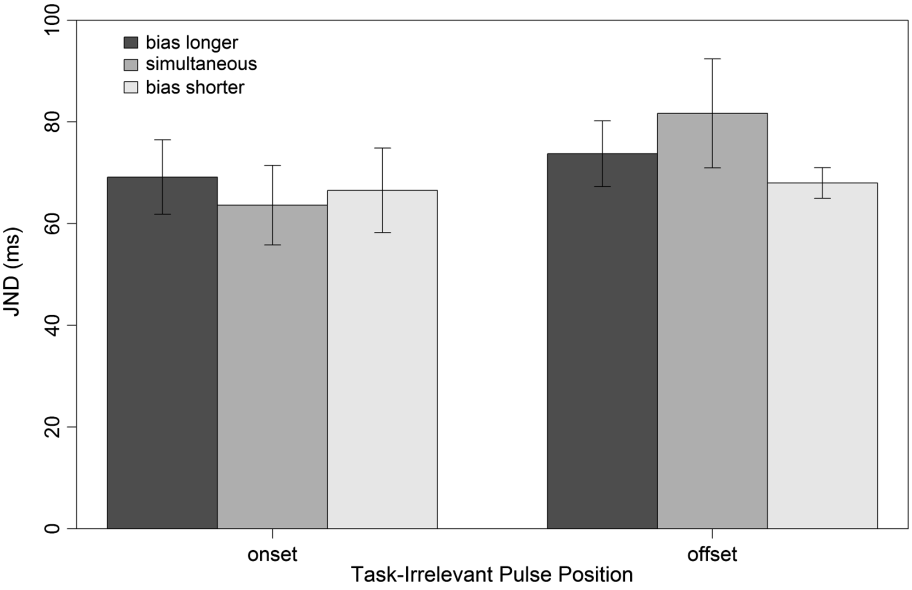

Figure 7. Mean Just Noticeable Difference (JND) in ms ± 1 standard error of the mean, depending on pulse position and bias condition in Experiment 2.

response-relevant information about interval duration, which could have biased participants' responses. Thus, the observed effects are most likely due to a multimodal integration bias rather than strategic responding based on the auditory information. Finally, again, the observed bias effect on perceived duration did not come with a cost regarding discrimination sensitivity when the auditory and visual pulses were incongruent rather than congruent.

Interestingly, there was a general difference in perceived duration depending on the position of the task-irrelevant pulses. Specifically, intervals were generally judged as longer when the auditory irrelevant signals were presented contingent on the onset rather than the offset of the intervals. This difference might be due to a response bias or even be mediated by attentional processes; for example, a stimulus that is presented relatively early during the interval might serve as a warning or accessory signal (Hackley & Valle-Inclán, 1998; Niemi & Näätänen, 1981) or even help to open an attentional gate, that has been postulated to operate in temporal perception (Zakay & Block, 1997). Since all these conditions are known to shorten perceptual latency of a given signal and increase perceived duration (Bausenhart, Rolke, Seibold, & Ulrich, 2010; Grondin & Rammsayer, 2003; Seifried, Ulrich, Bausenhart, Rolke, & Osman, 2010), similar processes might have been at play in the present experiments. Most importantly, however, such attentional processes are unlikely to be the sole source of the influence of Bias Condition on perceived duration. First, the effects of Bias Condition were independent of whether the irrelevant pulse accompanied the onset or the offset of the presented interval. Second, a prolongation of perceived duration was even observed when the irrelevant pulse followed the relevant offset pulse. This effect can neither be explained by temporal attention nor by a distraction effect and thus points strongly to the existence of a "true" multisensory integration of temporal information.

General Discussion

Overall, these results corroborate the results of previous studies (Chen & Yeh, 2009; Klink et al., 2011; Romei et al., 2011) showing that task-irrelevant auditory intervals of conflicting duration strongly alter the perception of visually presented time intervals. With the present results, however, this influence can be generalized to the perception of empty temporal intervals, and at the same time, specified with regard to its effects on perceived duration, as indexed by PSE, and discrimination sensitivity, as indexed by JND.

Both experiments showed that irrelevant auditory intervals or pulses can strongly affect perceived duration of relevant visual intervals. Likewise, Experiment 1 showed that a considerably smaller, yet reliable effect in the reverse direction (visual intervals affecting perceived duration of auditory intervals) is also present. Thus, multisensory integration of interval duration indeed seems to lead to the formation of a single, unified percept, and the magnitude (i.e., duration) of this percept is codetermined by the inputs from the auditory and the visual modality. Since the impact of vision on the combined audiovisual percept is relatively small compared to the impact of audition, however, it is conceivable that respective evidence can only be gathered with sufficiently fine-grained analysis methods. In this regard, the employed psychophysical approach, in which discrimination performance was assessed for a range of different comparison intervals, proved to be especially useful. Thus, in the present study we could expand previous findings which indicated a unidirectional influence of auditory intervals on the perceived duration of visual intervals (Chen & Yeh, 2009; Klink et al., 2011) to an asymmetric, but bidirectional pattern of multisensory influences.

Furthermore, the present experiment also informs about the effects of multisensory integration on discrimination sensitivity. Specifically, the biasing effects on perceived duration are not accompanied by costs regarding discrimination sensitivity. Most interestingly, even when the irrelevant auditory information was inconsistent with the relevant one (i.e., longer or shorter interval duration), the presence of this information led to an improvement in sensitivity for interval discrimination compared to the unimodal visual information (Experiment 1). When comparing JND estimates in the "longer" and "shorter" conditions to the "simultaneous" condition, both experiments showed that the latter unbiased condition does not lead to higher discrimination sensitivity than the conditions with incongruent information.

This latter result seems to be in contrast, for example, to the results of Romei et al. (2011), who demonstrated improved perceptual sensitivity (as measured by d') for bimodal *congruent* intervals compared to unimodal visual ones, and a clearly decreased perceptual sensitivity for bimodal *incongruent* intervals compared to unimodal visual ones. In their study, however, the irrelevant auditory information in the congruent condition always corresponded to the correct response (e.g., the short visual interval was always accompanied by the short auditory interval), whereas in the incongruent condition, it always corresponded to the wrong response (e.g., the short visual interval was always accompanied by the long auditory interval). Thus, auditory congruent information would have increased the probability for a correct response, thereby leading to a performance increment, and auditory incongruent information would have decreased the probability for a correct response, thereby leading to a performance decrement, as compared to the unimodal condition. Thus, even though this result clearly indicates that the perception of time intervals in one modality is affected by temporal information from another modality, it confounds the actual effects of multisensory integration on perceived duration and on sensitivity. The present study, in contrast, did overcome this limitation by assessing full psychometric functions and by introducing shorter as well as longer "incongruent" information. Thereby, we were able to demonstrate clearly dissociable effects of multisensory integration on perceived duration and sensitivity.

As has been suggested by Klink et al. (2011), various mechanisms might contribute to the integration of interval durations from different modalities. First, the temporal ventriloquism illusion, in which the perceived temporal occurrence of a particular signal is biased toward the temporal occurrence of a signal from another modality, might play a crucial role in this respect. For example, Morein-Zamir et al. (2003) asked their participants to discriminate the temporal order of two visual stimuli to assess the temporal resolution of the perceptual system. Importantly, Morein-Zamir et al. found that temporal resolution for the pair of visual stimuli did not only depend on the visual information but was also affected by a concurrent presentation of auditory stimuli. Specifically, better temporal resolution was observed when auditory stimuli were presented shortly before the first visual stimulus and shortly after the second visual stimulus (asynchronous condition) than when the auditory stimuli were presented simultaneously with the visual stimuli (simultaneous condition). The authors explained this by suggesting that the auditory stimuli temporally captured the visual ones, and thus "pulled" them toward their temporal occurrence. Specifically, the visual stimuli were pulled apart in the asynchronous condition, thereby their perceived temporal separation increased, and this enabled an improved discrimination of their temporal order. Such capturing of the temporal occurrence of visual stimuli by asynchronous auditory stimuli has been repeatedly demonstrated within various experimental setups and different tasks (e.g., Aschersleben & Bertelson, 2003; Fendrich & Corballis, 2001; Shipley, 1964; Vroomen & de Gelder, 2004). Likewise, in the present experiment, the perceived onset and offset of the presented relevant intervals might have been pulled toward the on- and offset of the task-irrelevant intervals. Thereby, perceived duration of the relevant interval would be stretched in the "longer" condition and compressed in the "shorter" condition, compared to the "simultaneous" condition. Thus, the temporal ventriloquism account is highly consistent with the results of the present study. This explanation also seems to fit well with the results of our Experiment 2, in which only a single pulse was presented in the biasing modality, and the magnitude of the biasing effect was approximately half as large as the one obtained in the visual task condition of Experiment 1.

Even though this account is a plausible explanation for the present results, other factors might also have contributed to the observed results. For example, the presence of auditory irrelevant stimuli at different time points before or during the presentation of the relevant intervals might have accelerated the pulse rate of an internal clock and thus might have affected visual duration judgments differentially (Chen & Yeh, 2009; Klink et al., 2011). It is even conceivable that, in the "longer" condition, the irrelevant temporal onset signal might have acted as a temporal warning signal for interval presentation, and thereby have contributed to a prolonged perceived duration (cf. Grondin & Rammsayer, 2003; Mo & George, 1977). Likewise, in the "shorter" condition, the signals from the irrelevant condition might have acted as a distractor, pulling attention away from the to-be-timed modality, and thereby have shortened perceived duration (cf. Brown, 2008), even though, as outlined in the Discussion of Experiment 2, such attentional influences cannot explain the complete pattern of results. Future research will hopefully elucidate the respective contributions of temporal ventriloquism, arousal, and attention to the observed multisensory integration effects in perceived duration.

To sum up, multimodal integration of incongruent time intervals leads to clearly dissociable effects on perceived duration and sensitivity. Specifically, inconsistent duration information from an irrelevant modality can substantially bias perceived duration in a relevant modality, but it does not lead to an accompanying decline in discrimination sensitivity. Rather, the perceptual system seems to combine the different sources of information in order to obtain especially stable interval representations, and accordingly high discrimination sensitivity.

Acknowledgments

This study was supported by Grant BA4110/3-1 of the German Research Foundation (DFG). We thank Julia Feyerabend and Laura Schneider for their assistance in data acquisition and four anonymous reviewers for helpful comments and suggestions.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262. doi: 10.1016/s0960-9822(04)00043-0

Allan, L. G. (1977). The time-order error in judgments of duration. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 31*, 24.

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension: 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology, 50*, 157–163. doi: 10.1016/s0167-8760(03)00131-4

Bausenhart, K. M., Rolke, B., Seibold, V. C., & Ulrich, R. (2010). Temporal preparation influences the dynamics of information processing: Evidence for early onset of information accumulation. *Vision Research, 50*, 1025–1034. doi: 10.1016/j.visres.2010.03.011

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory – visual spatial discordance. *Perception & Psychophysics, 29*, 578–584.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Bratzke, D., Seifried, T., & Ulrich, R. (2012). Perceptual learning in temporal discrimination: Asymmetric cross-modal transfer from audition to vision. *Experimental Brain Research, 221*, 205–210.

Brown, S. W. (2008). Time and attention: Review of the literature. In S. Grondin (Ed.), *Psychology of Time* (pp. 111–138). Bingley, UK: Emerald Group.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198*, 49–57. doi: 10.1007/s00221-009-1933-z

Bush, R. R. (1963). Estimation and evaluation. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 429–469). New York, NY: Wiley.

Chen, K.-M., & Yeh, S.-L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychologica, 130*, 225–234. doi: 10.1016/j.actpsy.2008.12.008

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*, 162–169. doi: 10.1016/j.tics.2004.02.002

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63*, 718–725.

Gamache, P.-L., & Grondin, S. (2010). Sensory-specific clock components and memory mechanisms: Investigation with parallel timing. *European Journal of Neuroscience, 31*, 1908–1914. doi: 10.1111/j.1460-9568.2010.07197.x

Goldstone, S., & Lhamon, W. T. (1972). Auditory-visual differences in human temporal judgment. *Perceptual and Motor Skills, 34*, 623–633.

Grondin, S., & Rammsayer, T. (2003). Variable foreperiods and temporal discrimination. *Quarterly Journal of Experimental Psychology, 56A*, 731–765. doi: 10.1080/02724980244000611

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science, 16*, 228–235. doi: 10.1111/j.0956-7976.2005.00808.x

Hackley, S. A., & Valle-Inclán, F. (1998). Automatic alerting does not speed late motoric processes in a reaction-time task. *Nature, 391*, 786–788.

Hellström, A. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin, 97*, 35–61.

Heron, J., Whitaker, D., & McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction. *Vision Research, 44*, 2875–2884. doi: 10.1016/j.visres.2004.07.001

Kanai, R., Lloyd, H., Bueti, D., & Walsh, V. (2011). Modality-independent role of the primary auditory cortex in time estimation. *Experimental Brain Research, 209*, 465–471. doi: 10.1007/s00221-011-2577-3

Klink, P. C., Montijn, J. S., & van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention, Perception, & Psychophysics, 73*, 219–236. doi: 10.3758/s13414-010-0010-9

Mo, S. S., & George, E. J. (1977). Foreperiod effects on time estimation and simple reaction time. *Acta Psychologica, 41*, 47–59.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*, 154–163. doi: 10.1016/s0926-6410(03)00089-2

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7*, 308–313.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89*, 133–162.

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Romei, V., De Haas, B., Mok, R. M., & Driver, J. (2011). Auditory stimulus timing influences perceived duration of co-occurring visual stimuli. *Frontiers in Psychology, 2*, 215. doi: 10.3389/fpsyg.2011.00215

Sarmiento, B. R., Shore, D. I., Milliken, B., & Sanabria, D. (2012). Audiovisual interactions depend on context of congruency. *Attention, Perception, & Psychophysics, 74*, 563–574.

Seifried, T., Ulrich, R., Bausenhart, K. M., Rolke, B., & Osman, A. (2010). Temporal preparation decreases perceptual latency: Evidence from a clock paradigm. *The Quarterly Journal of Experimental Psychology, 63*, 2432–2451. doi: 10.1080/17470218.2010.485354

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science, 145*, 1328–1330. doi: 10.1126/science.145.3638.1328

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport: For Rapid Communication of Neuroscience Research, 12*, 7–10. doi: 10.1097/00001756-200101220-00009

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Crossmodal temporal discrimination: Assessing the predictions of a general pacemaker-counter model. *Perception & Psychophysics, 68*, 1140–1152.

Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 513–518. doi: 10.1037/0096-1523.30.3.513

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1327–1339. doi: 10.1037/0096-1523.7.6.1327

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why "sounds are judged longer than lights": Application of a model of the internal clock in humans. *The Quarterly Journal of Experimental Psychology, 51B*, 97–120.

Wickens, T. D. (2001). *Elementary signal detection theory*. Oxford, UK: Oxford University Press.

Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Sciene, 6*, 12–16.

Received November 30, 2012
Revision received September 19, 2013
Accepted September 24, 2013

Karin M. Bausenhart

Fachbereich Psychologie
University of Tübingen
Schleichstrasse 4
D-72076 Tübingen
Germany
Tel. +49-7071-29-75582
Fax +49-7071-29-2410
E-mail karin.bausenhart@uni-tuebingen.de

APPENDIX B – Study 2

De la Rosa, M. D., & Bausenhart, K. M. (2013). Multimodal integration of interval duration: Temporal ventriloquism or changes in pacemaker rate? *Timing and Time Perception*, *1*, 189–215. https://doi.org/10.1163/22134468-00002015

Running head: MULTIMODAL INTEGRATION OF INTERVAL DURATION

Multimodal Integration of Interval Duration:

Temporal Ventriloquism or Changes in Pacemaker Rate?

Maria Dolores de la Rosa & Karin M. Bausenhart

University of Tübingen

Word count (main text without references): 9,338 words

Keywords: Time perception, multisensory integration, temporal ventriloquism, pacemaker-accumulator model.

Address editorial correspondence to Maria Dolores de la Rosa Gamiz, Fachbereich Psychologie, University of Tübingen, Schleichstrasse 4, D-72076 Tübingen, Germany.

Phone: +49 (0)7071 29-76423, FAX: +49 (0)7071 29-2410, email: maria-dolores.de-la-rosa-gamiz@uni-tuebingen.de

Abstract

Previous studies demonstrated that perceived duration of visual intervals is strongly influenced by conflicting auditory intervals. However, it remains unclear which mechanisms underlie this multimodal integration of interval duration. To investigate this issue, we employed a reproduction task with empty (Experiment 1) or filled (Experiment 2) intervals, and a paired-comparison task (Experiment 3) to assess perceived duration of visually marked intervals, which could be accompanied by auditory marked intervals with congruent (same) or conflicting (longer and shorter) durations. First, we predicted that conflicting auditory intervals would bias the perceived duration of visual intervals towards the duration of the auditory ones. Second, according to pacemaker-accumulator models, two different mechanisms might contribute to multimodal interval integration: changes in pacemaker rate or changes of the switch component based on temporal ventriloquism effects. In the former case, the multimodal bias effect should increase with increasing interval duration. In the latter case, the effect should remain constant across interval durations. All experiments showed a strong influence of auditory interval duration on perceived visual duration. In Experiments 1 and Experiment 2, this bias effect unexpectedly decreased with increasing interval duration, which might be due to limitations of the employed duration reproduction method. In Experiment 3, however, the observed multimodal bias effect remained clearly constant across all interval durations. This finding supports the idea that multimodal integration of conflicting time intervals is mainly determined by a temporal ventriloquism effect, which affects the switch component of a pacemaker-accumulator mechanism. In addition, we demonstrated that bimodal congruent intervals are perceived as longer than unimodal visual ones. For filled (Experiment 2), but not for empty (Experiment 1) intervals, this effect seems to be caused by an increase in pacemaker rate.

Introduction

Many natural objects or events can be perceived via multiple senses and the corresponding information has to be recombined across senses in order to form a coherent and comprehensive perceptual impression. However, under certain circumstances such a multisensory integration fails to provide a veridical perception of the underlying event. For instance, it has been shown that perception of time durations can be distorted when temporal information is delivered by different sensory modalities (e.g., Chen & Yeh, 2009; Goldstone & Lhamon, 1972; Klink, Montijn, & van Wezel, 2011; Walker & Scott, 1981; Wearden, Edwards, Fakhri, & Percival, 1998). In such cases, auditory temporal information plays an important role since it strongly dominates visual temporal information and thus determines perception of multimodal events (Bausenhart, de la Rosa, & Ulrich, submitted; Chen & Yeh, 2009; Klink et al., 2011; Romei, De Haas, Mok, & Driver, 2011; Walker & Scott, 1981). It is unclear, however, which mechanisms exactly underlie the multisensory integration of interval duration.

A recent line of evidence investigated this issue by presenting multimodal temporal intervals with conflicting durations (Bausenhart et al., submitted; Klink et al., 2011; Romei et al., 2011). In these studies, it was demonstrated that perceived duration of an interval in one modality can be strongly affected by the presentation of a task-irrelevant interval of conflicting duration in another modality. For example, Bausenhart et al. (submitted) employed a discrimination task in which empty standard and comparison intervals were presented, each marked by two pulses in a task-relevant modality. Participants were asked to decide in each trial which interval (i.e., standard or comparison) was the longer one. Both intervals were accompanied by task-irrelevant empty intervals in another modality (i.e., visual relevant interval markers accompanied by auditory irrelevant interval markers, or vice versa). For the standard intervals, markers in both modalities were always presented simultaneously. For the comparison intervals, however, the task-irrelevant markers were presented according to different bias conditions: In the “simultaneous” condition, task-irrelevant markers were presented at the same time as task-relevant markers. In the “longer” condition, task-irrelevant markers were presented slightly before and after the task-relevant onset and offset markers, respectively. Finally, in the “shorter” condition, task-irrelevant markers were presented slightly after and before the task-relevant onset and offset markers, respectively (cf. Figure 1 for an illustration of similar conditions). The duration of the task-irrelevant intervals severely influenced perceived duration in the task-relevant modality even though participants were explicitly instructed to ignore the stimulation in the task-irrelevant modality. Specifically, perceived duration was shortened in the “shorter” condition and prolonged in the “longer” condition compared to the simultaneous condition. This multimodal bias was especially pronounced when vision was task-relevant, suggesting that the task-irrelevant auditory markers strongly affected perceived visual duration. A similar, yet much smaller effect was found for the reverse condition, in which audition was task relevant. Thus, for multimodal temporal intervals with conflicting duration, audition seems to play an especially important role in determining the perceived duration of the multimodal compound stimulus (see also Klink et al., 2011).

This and similar results (Klink et al., 2011; Romei et al., 2011) might be interpreted in terms of the *temporal ventriloquism* effect (cf. Morein-Zamir, Soto-Faraco, & Kingstone, 2003). This phenomenon is observed when two short stimuli are presented in different modalities (e.g. vision and audition) with a slight temporal asynchrony between their respective onsets. In this case, participants perceive one coherent point in time at which the combined audiovisual stimulus occurs, and this time point is usually biased towards the appearance of the auditory stimulus (Bertelson & Aschersleben, 2003; Getzmann, 2007; Morein-Zamir et al., 2003). Building on this temporal ventriloquism effect, Klink et al. (2011) suggested that the integration of multimodal intervals is based on a similar mechanism. They proposed that the auditory and visual stimuli access the brain as separate bits of information. They are then processed in a grouping stage where both inputs are either intra- or crossmodally grouped (e.g., based on temporal proximity between the stimuli presented in the different modalities). If the stimuli are crossmodally grouped, a temporal ventriloquism-like effect causes the visual interval onset to be “pulled” towards the auditory interval onset (and likewise for the offset). Importantly, the perceived multimodal onset and offset then determine the switch control of a pacemaker-accumulator model (Gibbon, 1977; Gibbon, Church, & Meck, 1984) and thereby affect perceived duration. Specifically, pacemaker-accumulator models of time perception (Gibbon, 1977; Gibbon et al., 1984; Wearden et al., 1998) incorporate several components such as an internal pacemaker, a switch mechanism, an accumulator, a reference memory, and a comparator stage. First, the pacemaker continuously generates pulses. Whenever a to-be-timed stimulus is detected, the switch closes and consequently, the pulses are forwarded to the accumulator stage. When the to-be-timed stimulus ceases, the switch opens again and pulses are no longer forwarded to the accumulator. The number of accumulated pulses represents the duration of the stimulus, and it can be compared to the contents of the reference memory (e.g., the number of pulses that corresponds to a previously presented stimulus). Thus, if a temporal ventriloquism effect causes the switch of this mechanism to close earlier and/or open later, more pulses will be forwarded to the accumulator stage and thus, the perceived duration of the corresponding interval is prolonged. Likewise, if the switch is caused to close later and/or to open earlier, a shorter perceived duration will result.

Moreover, Klink et al. (2011) suggested that a second mechanism might also contribute to the perceived duration of audiovisual intervals. In addition to the switch control, the authors proposed a multimodal stimulus processing stage where the pacemaker’s pulse rate would be predominantly determined by the auditory information (Chen & Yeh, 2009; Penton-Voak, Edwards, Percival, & Wearden, 1996; Wearden et al., 1998). Specifically, it is well-known that auditory stimuli are typically perceived as longer than visual ones (Goldstone & Lhamon, 1972; Ulrich, Nitschke, & Rammsayer, 2006; Walker & Scott, 1981; Wearden et al., 1998). This effect has been ascribed to different pulse rates of the internal pacemaker component (e.g., Wearden et al., 1998). Specifically, auditory intervals are assumed to produce a higher pulse rate, resulting in a larger number of accumulated pulses and hence, a longer perceived duration than visual intervals. Similarly, the pulse rate for a combined audiovisual event might be determined by the auditory stimulation rather than the visual one.

Several empirical results provide evidence for this notion (e.g., Chen & Yeh, 2009; Walker & Scott, 1981; see also Klink et al., 2011, Exp. 3). For example, Chen and Yeh (2009) reported an oddball experiment in which participants had to compare a train of repeatedly presented visual standard intervals with a single comparison interval presented in a different color (the oddball). This oddball was either unimodal or bimodal (i.e., accompanied by a tone of identical duration). Not only was the infrequent oddball interval perceived as longer than the frequent standard, but more important for the present purposes, bimodal oddballs were also perceived as longer than the unimodal visual ones.[1] Similar results were reported by Walker and Scott (1981), who measured duration reproductions for uni- and bimodal filled intervals. They showed that reproduced durations of audiovisual congruent intervals were closer to the reproduced duration of unimodal auditory intervals than to unimodal visual intervals.

Most intriguingly, in the studies of Chen and Yeh (2009) and Walker and Scott (1981), the difference in perceived duration between unimodal visual and bimodal intervals even increased with increasing interval duration. This latter effect supports the notion of a higher pacemaker rate for audiovisual compared to unimodal visual stimuli, since according to the pacemaker-accumulator model, the magnitude of any effect caused by differences in pacemaker rate should be proportional to interval duration (see also Penton-Voak et al., 1996; Wearden et al., 1998). Specifically, a higher pacemaker rate should evoke a prolonged perceived duration of any given interval, and this effect should get more pronounced the longer the interval duration is.

In summary, two distinct mechanisms have been suggested to play a role in multimodal integration of temporal intervals: switch effects mediated by temporal ventriloquism and pacemaker effects. Specifically, for multimodal intervals of *conflicting* duration, temporal ventriloquism might affect the switch mechanism, and thus the moments at which the switch closes/opens in order to start/terminate the accumulation of pulses. Yet, so far, there is no direct evidence for this claim. It is likewise conceivable that multimodal integration effects are due to different pacemaker rates of an internal clock, as has been demonstrated for audiovisual filled events of *congruent* duration (Chen & Yeh, 2009; Walker & Scott, 1981).

Therefore, the aim of the present study was to investigate which mechanisms contribute to perceived duration of multimodal intervals with conflicting duration, by assessing perceived duration across a wide range of interval durations. Thus, based on the results of previous studies (Bausenhart et al., submitted; Klink et al., 2011; Walker & Scott, 1981), we expected to find a strong influence of conflicting auditory interval duration on perceived visual duration. Importantly, if the multimodal bias is independent of interval duration, it can be attributed to temporal ventriloquism effects on the switch component. Otherwise, if the magnitude of the multimodal bias increases with increasing interval duration, as has been observed so far only for filled multimodal intervals of congruent duration, this bias likely depends on differences in the auditory and visual pacemaker rates (cf., Chen & Yeh, 2009; Wearden et al., 1998).

Experiment 1

In Experiment 1, we employed a time reproduction task in which participants had to reproduce visually marked empty interval durations ranging from 500 ms to 1,253 ms.We employed empty intervals similar as in a previous study (Bausenhart et al., submitted), in order to replicate and extend these previous results with a different method. The intervals were either unimodal or accompanied by congruent (same) or conflicting (shorter or longer) auditory marked intervals. We predicted that reproductions of the visual intervals would be biased towards the durations of the conflicting auditory intervals. If this bias in reproduced duration is due to a temporal ventriloquism effect, it should have a similar magnitude independent of interval duration. Otherwise, if the bias in perceived duration depends mainly on differences in the pulse rate of the pacemaker, the bias should increase with increasing interval duration. Unimodal visual intervals were included for two reasons: First, participants were instructed to base their reproductions on the duration of the visually marked intervals alone (cf., Bausenhart et al., submitted; Klink et al., 2011). An alternative strategy of responding to the auditory stimulation should be strongly discouraged by presenting randomly interspersed unimodal visual intervals, because in these trials participants cannot rely on auditory stimulation alone to perform their task. Second, these trials would allow a direct comparison of reproduced duration between unimodal visual and bimodal congruent intervals (cf., Chen & Yeh, 2009; Walker & Scott, 1981).

Method

Participants. Twelve undergraduate students (6 females; age range: 20-33 years; mean age: 22.75 years) from the University of Tübingen voluntarily took part in Experiment 1 in exchange of either course credits or money. [2]

Apparatus and Stimuli. The experiment was run on a Macintosh computer connected to a CRT monitor running at 150 Hz. The presentation of all stimuli and data recording were controlled by Matlab with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). Responses were collected with the *m* key of a standard German keyboard. Participants were seated in a dimly lit room at a viewing distance of approximately 56 cm from the computer screen. All visual stimuli were presented in white (100 cd/m²) on a black background (< 1 cd/m²) at the center of the screen. In each trial, a white dot (1.8 mm diameter, 0.18° visual angle) defined the fixation point and two white disks (6.4 mm diameter, 0.65° visual angle) marked the interval to be estimated. The auditory stimulus consisted of a pure sinusoidal tone (800 Hz, 79 dB, 13.3 ms) with ramped 5 ms on- and offsets, presented binaurally through headphones. An oscilloscope was employed to assess the duration of visual and auditory intervals and to ensure the required synchronicity (or asynchronicity) between the two modalities.

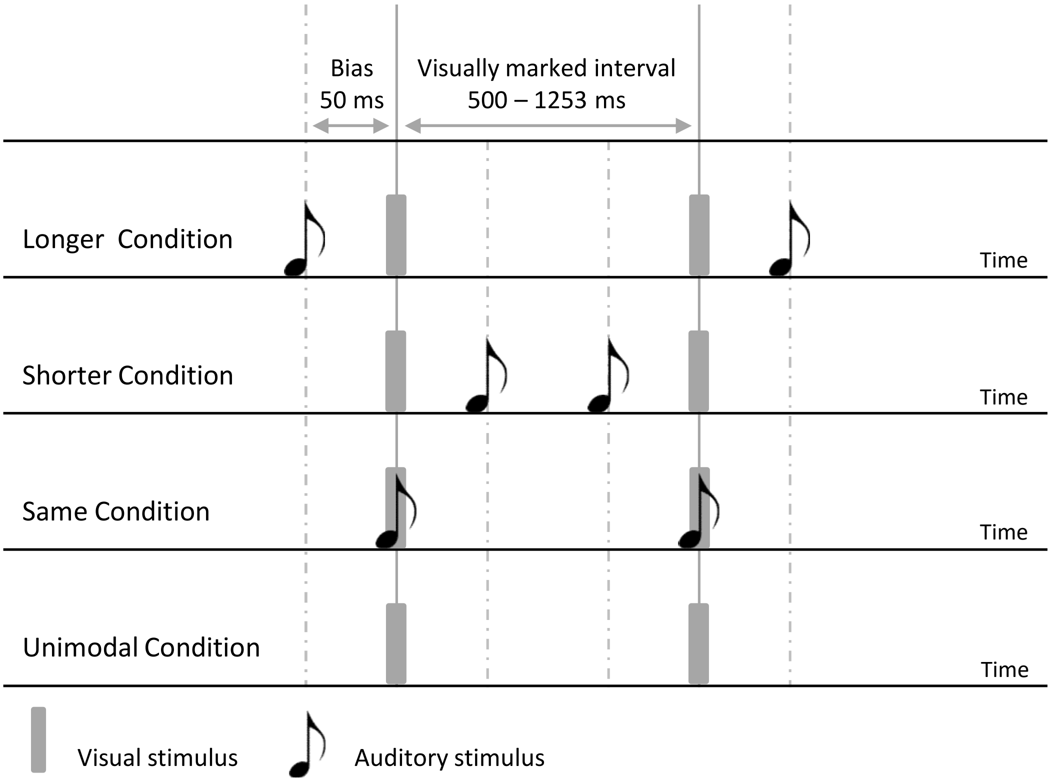

Figure 1. Schematic representation of visual and auditory markers in the four bias conditions.

Procedure. Participants were instructed to make time reproductions as close as possible to the interval duration defined by the visual markers (i.e., the duration from first marker offset to second marker onset) and to ignore any auditory stimuli.

Each trial started with the presentation of a blank screen for 1,000 ms. Then, the fixation point was presented for 1,000 ms preceding the visual onset marker that appeared for 13.3 ms at the center of the screen, concentric with the fixation point. After the onset marker, the screen remained black for an interval of either 500, 753, 1,000 or 1,253 ms. Each interval duration was equally likely and randomized across trials. The interval duration was followed by the visual offset marker presented for 13.3 ms. Four different conditions were introduced to manipulate multisensory bias (Figure 1). In unimodal trials, no auditory stimuli were presented during the trial. However, in bimodal trials, auditory stimuli accompanied the visual ones according to three different conditions: “longer”, “same” and “shorter”. In the “longer” condition, the first tone appeared 50 ms *before* the visual onset marker and the second one 50 ms *after* the visual offset marker. In the “shorter” condition, the tone appeared 50 ms *after* the visual onset marker and 50 ms *before* the visual offset. In the “same” condition, the auditory stimuli appeared *simultaneously* with both visual markers. All four conditions (unimodal, longer, shorter, and same) were equally likely and randomized across trials.

The visual offset marker was followed by a 500-ms blank screen. Then, the word “Jetzt” (“Now”) was displayed at the center of the screen requesting the participants to reproduce the visually marked empty interval duration with a single key press of their right index finger. This response prompt disappeared as soon as the key was pressed. Then, the screen remained empty and the next trial started 500 ms after the key was released.

The experiment consisted of two practice blocks and 16 experimental blocks composed of 32 trials each. At the end of each block, the absolute percentage of deviation of the reproduced durations from the presented interval durations was presented. This feedback information prevented any hint of whether participants had over- or underestimated the interval durations. Participants were instructed to keep this score as low as possible during the experiment, since lower values would correspond to more accurate reproductions. Finally, participants could initiate the next block by pressing the space bar.

Results

Practice blocks did not enter data analysis. For each participant and condition, mean and standard deviation (SD) of the reproduced duration were computed. Trials with reproduced durations below or above the condition mean +/- 2*SD were discarded from data analysis (4.3% of trials). Repeated-measures analyses of variance (ANOVA) were conducted with the factors Bias Condition (unimodal, longer, shorter, and same conditions) and Stimulus Duration (500, 753, 1,000, and 1,253 ms) for the mean reproduced duration, the Constant Error, and SD.

Reproduced Duration. Participants reproduced longer durations with increasing stimulus duration, $F(3, 33) = 108.7$, $p < .001$. Furthermore, reproduced duration was longest in the “longer” condition (858 ms), intermediate in the “same” condition (816 ms), and shortest in the “unimodal” (767 ms) and in the “shorter” (764 ms) condition, $F(3, 33) = 50.87$, $p < .001$. Additional bonferroni-corrected post-hoc *t*-tests for this main effect of bias condition (long vs. same vs. short vs. unimodal) showed significant differences for every comparison (all $p < .01$), except for the difference between the unimodal and the shorter condition, which was not significant (p = 1.00). Most importantly, the ANOVA revealed a significant interaction between Stimulus Duration and Bias condition, $F(9, 99) = 3.03$, $p = .003$.

To investigate this interaction in more detail, two additional analyses were conducted. First, a repeated-measures ANOVA was performed on a subset of the data that included reproduced duration for the “longer” and the “shorter” condition only. This analysis again showed that reproduced duration increased with increasing stimulus duration, $F(3,33) = 114.9$, $p < .001$. Longer reproductions were observed in the “longer” than in the “shorter” condition, $F(1,11) = 90.43$, $p < .001$. These two factors interacted, $F(3,33) = 6.80$, $p = .001$. Surprisingly, the difference in reproduced duration between the “longer” and “shorter” condition decreased with increasing stimulus duration (see Figure 2). Nonetheless, post-hoc t-tests revealed that the difference between the “longer” and “shorter” condition was highly significant for the shortest stimulus duration (500 ms), $t(11) = 7.62$, $p < .001$, as well as for the longest stimulus duration (1,253 ms), $t(11) = 4.69$, $p = .001$.

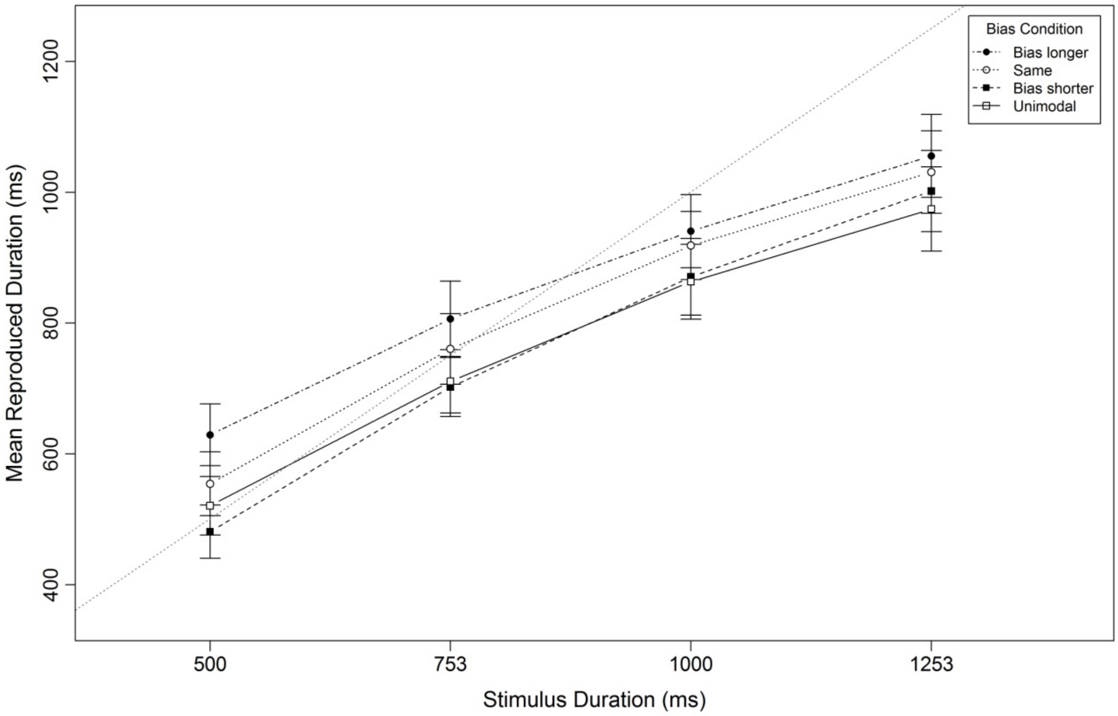

Figure 2. Mean reproduced duration in Experiment 1 as a function of Bias Condition (longer, shorter, same, and unimodal) and Stimulus Duration. Error bars represent the standard error of the mean. The veridical line is depicted in gray.

Second, we assessed the *bias effect* by computing the difference between the mean of reproduced durations for each bias condition (longer, shorter and unimodal) and the "same" condition, which served as baseline against which the magnitude of the multimodal bias effect was assessed. An ANOVA was conducted on these data with the novel factor Bias Condition (longer-same, shorter-same and unimodal-same) and Stimulus Duration (500, 753, 1,000 and 1,253 ms). The results showed that the overall bias effect remained constant across stimulus duration, $F < 1$. Furthermore, the "longer-same" condition showed an overestimation (41.91 ms), whereas the "shorter-same" and the "unimodal-same" condition showed an underestimation of perceived duration with respect to the same condition (-52.05 ms and -48.66 ms, respectively), $F(2, 22) = 56.71$, $p < .001$. Most importantly, both factors interacted significantly, $F(6, 66) = 3.69$, $p < .01$ (see Figure 3).

To investigate this interaction in more detail, further repeated-measure ANOVAs with the factor Stimulus Duration were individually performed on the bias effect of the "longer-same", the "shorter-same" and the "unimodal-same" conditions. First, the bias effect of the "longer-same" condition decreased with increasing stimulus duration, $F(3, 33) = 4.78$, $p < .01$. Second, the bias effect of the "shorter-same" condition also showed a decrement with increasing stimulus duration (see Figure 3) although this difference only approached significance, $F(3, 33)=2.27$, $p=.10$. Finally, the bias effect of the "unimodal-same" condition, in contrast to the previous ones, remained constant across stimulus duration, $F<1$.

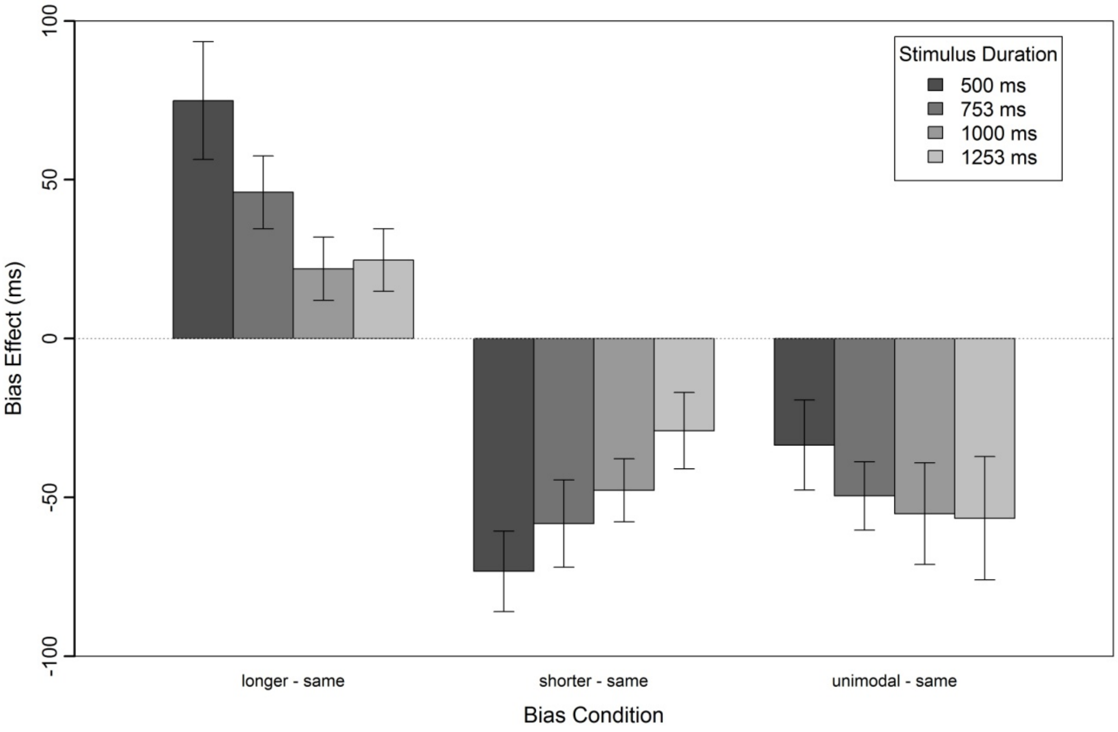

Figure 3. Bias Effect in Experiment 1 for the longer-same, shorter-same, and unimodal-same conditions and for each stimulus duration. Error bars represent one standard error of the mean.

Constant Error. An ANOVA was performed on the absolute deviation of the reproduced duration from the veridical duration (i.e., Constant Error, *CE*). This analysis showed a pronounced effect of stimulus duration on *CE*, $F(3,33) = 40.62$, $p < .001$, that is, a Vierordt effect (Lejeune & Wearden, 2009; Vierordt, 1868). Specifically, this effect reflects an overestimation of the shortest stimulus duration by 46.23 ms and an underestimation of the longest stimulus durations by 237.22 ms. This analysis also confirmed an effect of Bias Condition, and an interaction of both factors (both $p <.01$), which basically correspond to the multimodal bias effects reported above.

Variability of Reproductions. The ANOVA on SD of reproduced duration revealed that the variability of reproductions significantly increased with increasing stimulus duration, $F(3, 33) = 14.36$, $p < .001$. Furthermore, there was a significant main effect of Bias Condition, $F(3, 33)$

= 4.63, *p* = .01. The variability of reproductions was largest in the “longer” (153 ms) and “unimodal” (150 ms) condition, intermediate in the “same” (145 ms) condition, and smallest in the “shorter” (132 ms) condition. Finally, the interaction between Stimulus Duration and Bias Condition was not significant, even though a slight tendency was present, *F*(9,99) = 1.74, *p* = .09 (see Figure 4).

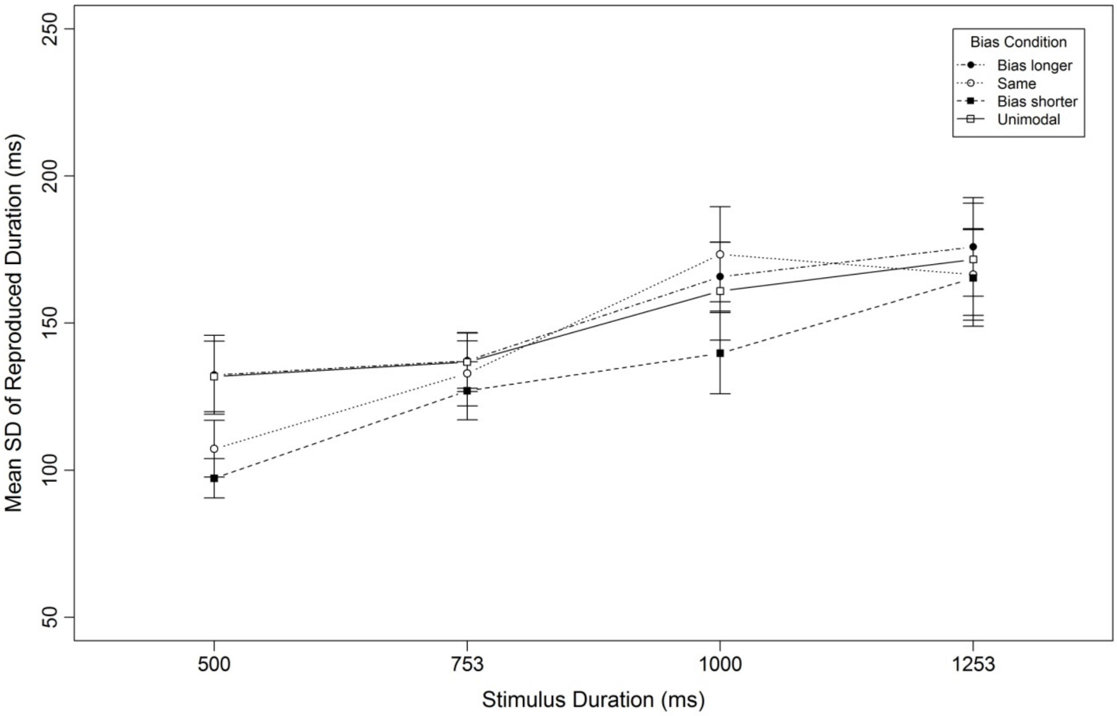

Figure 4. Mean standard deviation (SD) of reproduced duration in Experiment 1 as a function of Bias Condition (longer, shorter, same, and unimodal) and Stimulus Duration. Error bars represent the standard error of the mean.

Discussion

The results of Experiment 1 replicated previous studies (Chen & Yeh, 2009; Klink et al., 2011; Romei et al., 2011) showing a strong effect of auditory bias on perceived visual duration. Specifically, participants reproduced the longest durations when visual intervals were accompanied by “longer” auditory ones and the shortest reproduced durations when visual intervals were accompanied by “shorter” auditory ones. Since the true physical difference between the auditory intervals in the “shorter” and “longer” conditions was 200 ms, the observed difference of 94 ms between the reproduced duration in these two conditions amounts to an observed multimodal bias effect of 47 %. Thus, the auditory intervals strongly affected the perceived duration of the visual intervals, even though this effect was not as strong as the one observed in a previous study using the paired comparison method (Bausenhart et al., submitted). Nonetheless, a strong multisensory integration of temporal intervals of conflicting duration could be observed with the present time reproduction method.

Interestingly, the observed multimodal bias in perceived duration decreased with increasing interval duration, when comparing the “shorter” and the “longer” bias condition. This is most certainly inconsistent with the assumption that multimodal integration evokes its biasing effect through an increase in pacemaker rate, because this account predicts an increase of the multimodal bias with increasing interval duration. According to the alternative hypothesis, multimodal bias effects are due to temporal ventriloquism effects operating on the switch component of the pacemaker-accumulator model. In this case, the bias effect should be of comparable size independent of interval duration. Again, this has not been observed in the present experiment.

How can the observed underadditivity of the biasing effect with interval duration be explained? Regarding this question, it is interesting to note that the physically longest presented interval was clearly underestimated by participant’s reproductions, whereas there was a slight overestimation of the shortest physical interval duration (cf. the corresponding analysis on *CE*). This effect is consistent with Vierordt’s law (Vierordt, 1868) and has repeatedly been demonstrated in studies using reproduction methods (Wearden, 2003; Woodrow, 1934; for an overview, see Lejeune & Wearden, 2009). We suggest that it is possible that such a Vierordt effect could distort the effects of an experimental manipulation across different temporal intervals. Specifically, a “true” biasing effect of a given magnitude might be enlarged at short interval durations because of the general overestimation of short intervals. At the same time, it might be decreased at long interval durations because of the general underestimation of such long intervals. Therefore, a truly additive (or even overadditive) effect might become underadditive in the presence of a Vierordt effect.

Based on the present data, it cannot be decided whether this explanation is actually feasible or not, yet it might provide an account of why we observed an underadditive interaction of multimodal bias condition and interval duration, rather than the expected patterns of additivity (in case of a switch effect) or overadditivity (in case of a pacemaker rate effect). Therefore, it is difficult to clearly interpret our observed pattern of results in favor of any of these two hypotheses. Nonetheless, the sizable magnitude of the observed underadditivity of the multimodal bias effect renders an underlying additive pattern of results more plausible. This is because a very strong Vierordt effect would be necessary to counteract an overadditive effect and even reverse it to an underadditive effect. However, this would be a clearly speculative interpretation.

In this regard it is interesting to examine the difference between the “same” and the “unimodal” conditions of the present experiment. In support of previous studies (Chen & Yeh, 2009; Walker & Scott, 1981), unimodal visual intervals were reproduced as shorter than “same” intervals, that is, bimodal audiovisual intervals with congruent duration. As has been suggested repeatedly (e.g., Wearden, Norton, Martin, & Montford-Bebb, 2007), the auditory portion of the stimulation might have led to increased arousal. But, in contrast to these previous studies, in our Experiment 1 this effect was additive with increasing interval duration rather than overadditive (even though a slight numerical trend towards overadditivity was observed, cf. Fig. 3). There are several possible explanations for this discrepancy. First, and as discussed above, the existence of a Vierordt effect might have counteracted a similar overadditivity in our experiment, thus leading to an additive pattern of results. Second, both previous studies employed filled durations instead of empty intervals. It is possible that especially filled auditory intervals evoke more arousal than intervals that are only marked by short auditory pulses (Wearden et al., 2007). The enhanced arousal associated with filled auditory stimuli would lead to an increase in pacemaker rate, and thus, the effect of the combined audiovisual intervals should increase with increasing interval duration, while such an effect would not be observable for empty intervals. Indeed, some indirect evidence for this claim can be found in Experiment 4 of Walker and Scott (1981), which showed no such overadditivity when participants had to reproduce the duration of gaps in otherwise continuous tones and lights. Therefore, to investigate the discrepancy between our results with empty intervals and those reported for filled duration, we conducted a second experiment which employed filled auditory and visual intervals with conflicting durations.

Experiment 2

In Experiment 1, two important results were found: a pattern of underadditivity when comparing the "longer" and the "shorter" condition and an additive effect when comparing the "same" and the "unimodal" condition. This would suppose that different mechanisms are underlying the crossmodal integration of congruent and conflicting interval durations. It is possible that a Vierordt effect might have prevented, on the one hand, a possible additive effect (based on temporal ventriloquism effect) between long and short bias conditions and, on the other hand, a possible underlying overadditivity (based on a pacemaker effect) between congruent ("same") and unimodal intervals. However, it might also be possible that multimodal integration of empty intervals is mediated by different mechanisms than multimodal integration of filled intervals (e.g., because empty intervals might produce less arousal than filled ones), and thus produces different pattern of results (Wearden et al., 2007). Therefore, we replicated Experiment 1 with the sole difference of using filled instead of empty intervals. Returning to our main hypotheses, if filled auditory durations are really associated with an increase in arousal that consequently increases the pacemaker rate, especially the difference between the “same” and the “unimodal” condition should increase with stimulus duration.

Method

Participants. A fresh sample of twelve undergraduate students (12 females; age range: 19-44 years; mean age: 23 years) from the University of Tübingen voluntary participated in Experiment 2 in exchange of either course credits or money.

Apparatus, Stimuli, and Procedure. Stimuli, apparatus and procedure were identical to those employed in Experiment 1 except for the use of filled (continuous) auditory and visual intervals of either 500, 753, 1,000 or 1,253 ms, (each +/-100 ms for the conflicting biasing intervals), instead of empty intervals marked by brief pulses.

Results

For each participant and condition, mean and standard deviation (SD) of the reproduced duration were computed. 4.3% of trials were discarded from data analysis employing the same outlier criteria as in Experiment 1. Repeated analyses of variance (ANOVA) were conducted with the factors Bias Condition (unimodal, longer, shorter, and same condition) and Stimulus Duration (500, 753, 1,000, and 1,253) for the mean reproduced duration, *CE*, and SD.

Reproduced Duration. Participants reproduced longer durations with increasing stimulus duration, $F(3, 33) = 95.33$, $p < .001$. Reproduced duration was longest in the “longer” condition (1,051 ms), intermediate in the “same” condition (985 ms), and shortest in the “shorter” (913 ms) and in the “unimodal” (819 ms) condition, $F(3, 33) = 38.72$, $p < .001$. Additional bonferroni-corrected post-hoc *t*-tests for the main effect of Bias Condition (longer vs. same vs. shorter vs. unimodal) showed significant differences for every comparison (all $p < .05$). Importantly, the interaction between Stimulus Duration and Bias Condition was significant, $F(9, 99) = 5.66$, $p < .001$.

As Experiment 1, we conducted two additional analyses to investigate this interaction. First, a repeated-measures ANOVA with the factors Bias Condition (longer vs. shorter) and Stimulus Duration (500, 753, 1,000, and 1,253 ms) showed that reproduced duration increased with increasing stimulus duration, *F*(3, 33) = 106.9, *p* < .001. Reproduced durations were longer in the "longer" condition than in the "shorter" condition, *F*(1, 11) = 77.67, *p* < .001. Furthermore, the interaction between these two factors was significant, *F*(3, 33) = 5.24, *p* < .01. As Experiment 1, the difference in reproduced duration between both conditions decreased with increasing stimulus duration (see Figure 5). However, subsequent post hoc *t*-tests showed that the difference between the "longer" and "shorter" condition was significant for the shortest stimulus duration (500 ms), *t*(11) = 9.96, *p* < .001, as well as for the longest stimulus duration (1,253 ms), *t*(11)=6.08, *p* < .001.

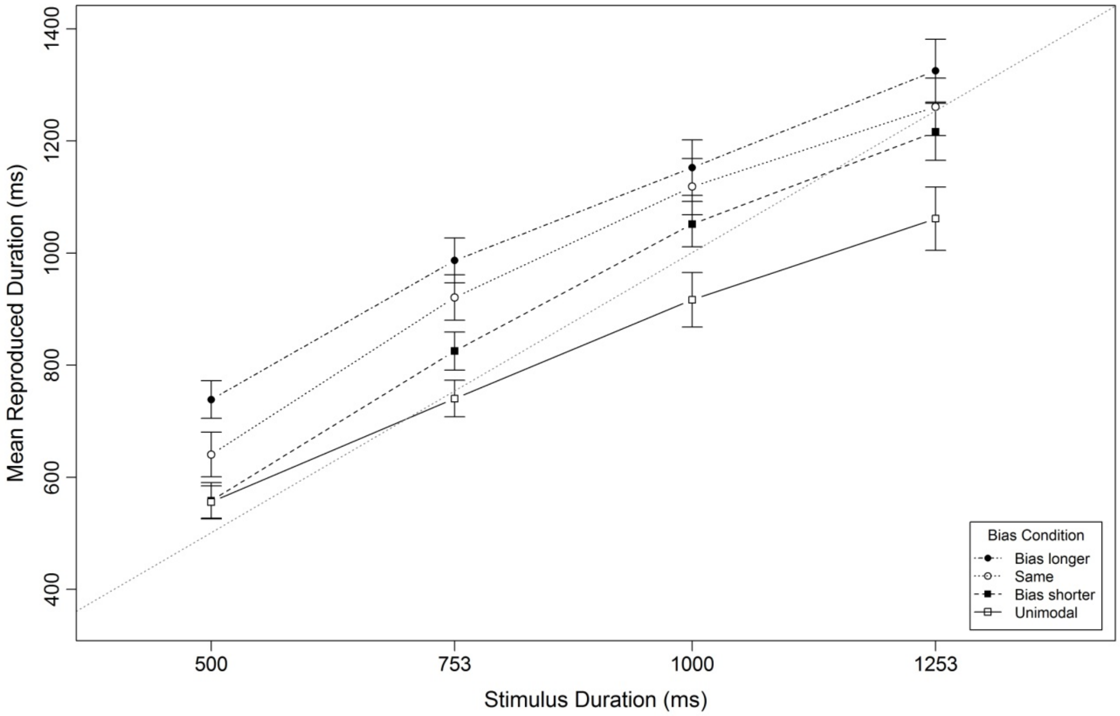

Figure 5. Mean reproduced duration in Experiment 2 as a function of Bias Condition (longer, shorter, same, and unimodal) and Stimulus Duration. Error bars represent the standard error of the mean. The veridical line is depicted in gray.

Second, and similarly to Experiment 1, an additional repeated-measures ANOVA was performed on the bias effect computed as the difference of reproduced duration in each condition (longer, shorter and unimodal) from reproduced duration in the “same” condition. The corresponding ANOVA showed that the bias effect differed slightly between stimulus durations, *F*(3, 33) = 3.02, *p* = .04. Specifically, it was smallest (-22.94 ms) for the 500 ms stimulus, intermediate (-59.8 ms) for the 1,253 ms stimulus and largest for the 753 ms (-70.04 ms) and the 1,000 ms stimulus (-78.19 ms). Moreover, the bias effect of the “longer-same” condition showed an overestimation of perceived duration (65.69 ms), whereas the bias effect of the “shorter-same” and the “unimodal-same” condition showed an underestimation (-72.33 ms and -166.59 ms, respectively), *F*(2, 22) = 41.46, *p* < .001. Most importantly, both factors interacted significantly, *F*(6, 66) = 6.72, *p* < .001 (see Figure 6).

Further repeated-measures ANOVAs with the factor Stimulus Duration were individually performed for the bias effect in the “longer-same”, the “shorter-same” and the “unimodal-same” conditions. First, the bias effect of the “longer-same” condition decreased with increasing stimulus duration (except for the longest stimulus duration), *F*(3, 33) = 3.76, *p* < .05. Second, the bias effect of the “shorter-same” condition also showed a slight decrement with increasing stimulus duration (see Figure 6), but this effect was not significant, *F*(3, 33) = 1.71, *p* = .18. Finally, the bias effect of the “unimodal-same” condition, in contrast to the previous ones, significantly increased with increasing stimulus durations, *F*(3, 33) = 6.25, *p* < .01.

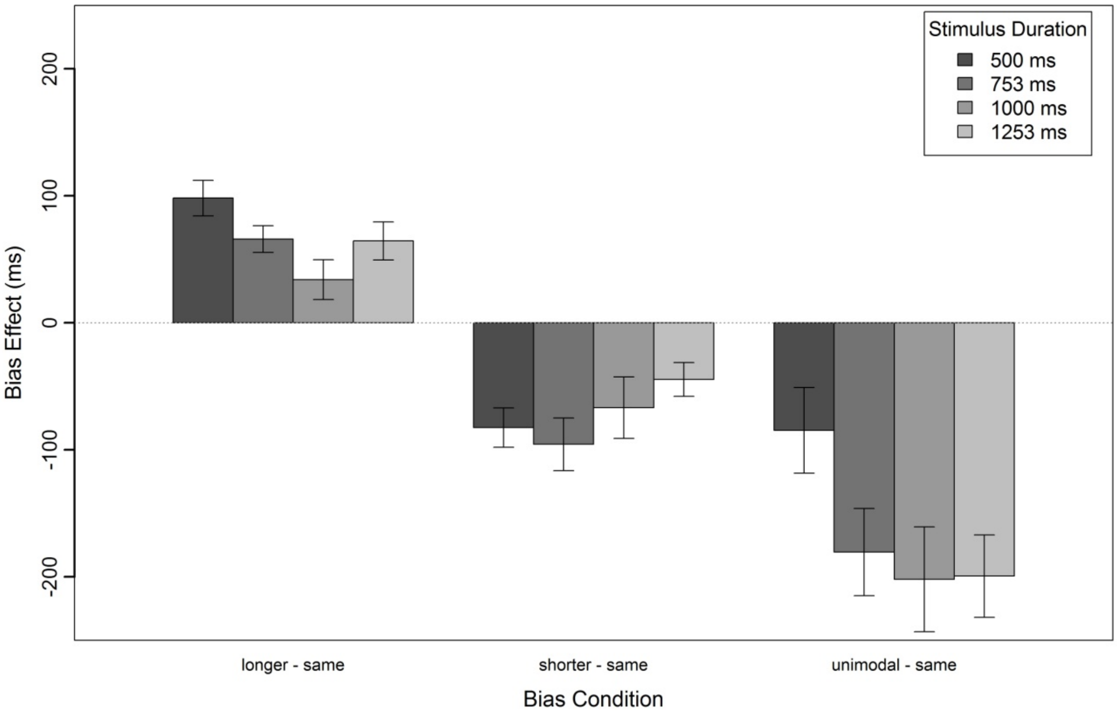

Figure 6. Bias Effect in Experiment 2 for the longer-same, shorter-same, and unimodal-same conditions and for each stimulus duration. Error bars represent one standard error of the mean.

Constant Error. As Experiment 1, a repeated-measures ANOVA was performed on the *CE*. This analysis also showed a pronounced Vierordt effect, that is, an effect of stimulus duration on *CE*, $F(3,33) = 7.94$, $p < .001$. Specifically, the short stimulus duration was overestimated by 123.21 ms and the longest stimulus duration was underestimated by 36.97 ms. This analysis also confirmed an effect of Bias Condition, and an interaction of both factors (both $p < .001$), which basically corresponds to the multimodal bias effects reported above.

Variability of Reproductions. The ANOVA on SD of reproduced duration revealed that the variability of reproductions significantly increased with increasing stimulus duration, $F(3, 33) = 13.96$, $p < .001$.

However, neither the main effect of Bias Condition nor the interaction between both factors showed a significant result, *F*(3, 33) = 2.02, *p* = .13 and *F*(9,99) = 1.17, *p* =.32, respectively (see Figure 7).

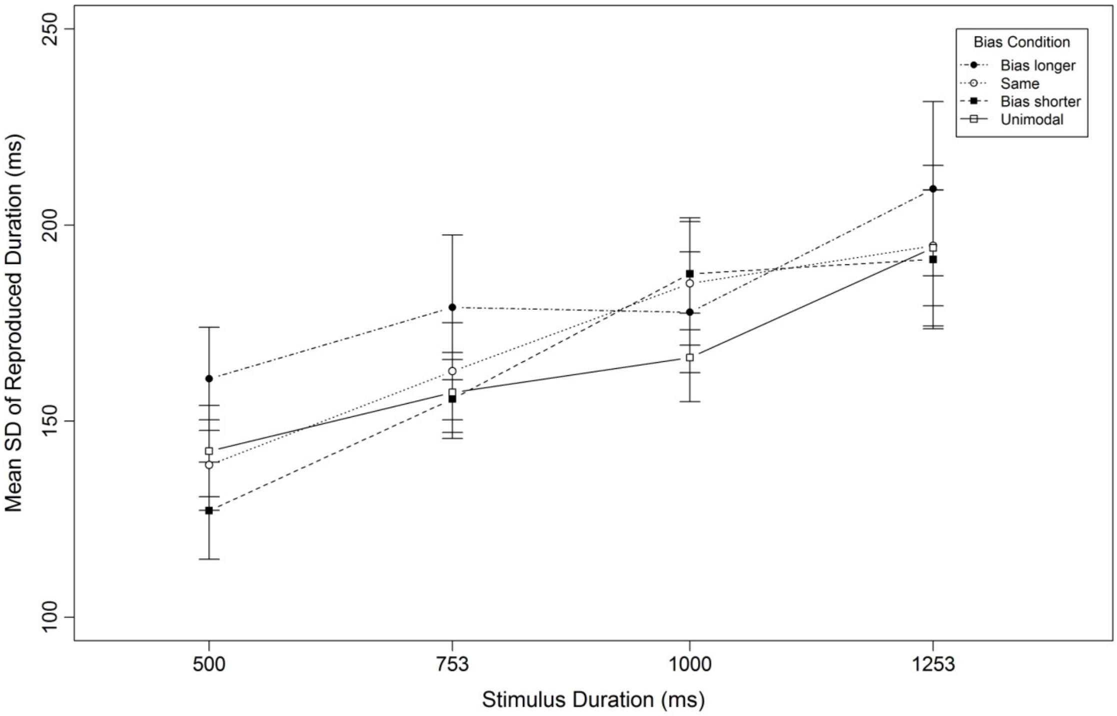

Figure 7. Mean standard deviation (SD) of reproduced duration in Experiment 2 as a function of Bias Condition (longer, shorter, same, and unimodal) and Stimulus Duration. Error bars represent the standard error of the mean.

Discussion

The strong effect of auditory bias on perceived duration was again observed in Experiment 2, which replicated our main finding of Experiment 1. Accordingly, reproduced duration was prolonged in the "longer" and shortened in the "shorter" condition, in comparison to the "same" condition. Importantly, the mean observed difference between reproduced durations for the "longer" and the "shorter" condition was 138 ms, which constitutes an observed multimodal bias effect of 69%. Thus, an even stronger multisensory integration of conflicting temporal intervals than in Experiment 1 was obtained in Experiment 2 employing filled audiovisual durations (see also Footnote 3).

In contrast to Experiment 1, however, the difference between the “same” and the “unimodal” condition showed a clear pattern of overadditivity across stimulus duration, which supports previous studies with congruent filled auditory and visual intervals (Chen & Yeh, 2009; Walker & Scott, 1981). Thus, our results add evidence in favor of an increased arousal evoked by filled auditory intervals (Wearden et al., 2007), which leads to an accelerated pacemaker rate, which in turn increases the effect of combined audiovisual intervals across stimulus durations. Accordingly, different mechanisms might be at play for the integration of filled (Experiment 2) vs. empty (Experiment 1) *congruent* intervals. Moreover, these results demonstrate that our experimental design is appropriate to reveal an existing overadditivity. Specifically, even though in this Experiment a Vierordt effect was also present, the overadditivity was strong enough to be still pronounced and well observable.

Regarding the observed multimodal integration of *conflicting* filled intervals, however, again a tendential (in comparison to the “same” condition) or significant (comparison of the “longer” and the “shorter” condition) underadditivity was observed with increasing interval durations. Thus, for multimodal intervals of conflicting duration, the integration process seems to be similar for empty and filled intervals. As discussed in Experiment 1, this pattern of results is neither consistent with the pacemaker rate notion by which an overadditivity would be expected, nor with the notion of a temporal ventriloquism effect on the switch component. Unfortunately, it is still hard to interpret the underadditive bias effect. Again, the observed Vierordt effect might have reduced an underlying additive bias effect, but this is still fairly speculative. Therefore, since Experiment 1 and 2 are inconclusive with respect to our hypotheses regarding the crossmodal integration of conflicting interval durations, we conducted a third experiment by employing a method that is less likely to be affected by distortions such as the Vierordt effect.

Experiment 3

In Experiment 1 and 2, the effect of bias condition on reproduced duration unexpectedly decreased with increasing interval duration. Speculatively, this pattern of results may be codetermined by the employed duration reproduction method, which can be subject to severe distortions as, for example, the Vierordt effect (Vierordt, 1868). To overcome this methodological limitation, we conducted a third experiment by employing a reminder task (cf., Bausenhart et al., submitted) in which a standard duration is always presented before a variable comparison stimulus. In this task, participant's judgments about perceived duration are always based on a direct comparison between a standard and a comparison interval, and thus, the judgments are less prone to biasing influences like the Vierordt effect. Furthermore, in this experiment, we increased the range of the employed interval durations from 500 – 1,253 ms in Experiment 1 and 2 to 500 – 2,000 ms. Any interaction of perceived duration with interval duration that is due to an increased pacemaker rate should thus get more easily evident for this larger range of interval durations. Since a much larger number of trials is typically required for the paired-comparison method, we also decided to omit the "same" condition and base our analysis on comparisons between a "unimodal", "shorter", and a "longer" condition only. These three conditions were presented in random order, whereas the three standard durations were tested in separate experimental sessions. Such blocking of stimulus duration might also help to prevent the occurrence of a Vierordt effect (Bausenhart, Dyjas, & Ulrich, in press; Jazayeri & Shadlen, 2010; but see Woodrow, 1934).

Method

Participants. Twelve participants (8 female; age range: 19-31 years; mean age: 22.58 years) took part in Experiment 3 in exchange of either course credits or money. Two of them had previously participated in Experiment 1. Two participants were excluded from data analysis because they had performed close to chance level in at least one experimental condition, thus leading to unrealistically large estimates of *JND* (> 1,000 ms).

Stimuli & Apparatus. All employed stimuli and apparatus were identical to the ones employed in Experiment 1, with the following exception: Responses were collected via the *y* and the *m* key of a standard German keyboard.

Procedure. Each participant performed three experimental sessions on separate days which differed in the duration of the employed standard and comparison intervals. The order of the three sessions was counterbalanced across participants. Apart from this, task and procedure were identical for all three sessions. Participants were instructed to ignore the auditory stimuli presented during the experiment and to base their responses on the visual stimulation only.

At the beginning of each experimental trial, a fixation dot was presented at the center of the screen for 1,000 ms and was then replaced with a white disc (standard onset marker) for 13.3 ms. Then, the screen remained empty for the standard duration, which varied between 500, 1,000, and 2,000 ms in different experimental sessions. After the standard duration, the white disc (standard offset marker) was again presented for 13.3 ms, before the fixation point reappeared for an interstimulus interval of 1,500 ms. Afterwards the white disc reappeared for 13.3 ms (comparison onset marker). When this marker disappeared, the comparison interval started. In the 500 ms standard session, the duration of this comparison interval varied randomly from trial to trial, in 11 equally spaced comparison levels ranging from 300 to 700 ms. In the 1,000 and 2,000 ms standard sessions, comparison levels ranged from 600 to 1,400 ms and from 1,200 to 2,800 ms, respectively. These ranges were chosen as to cover +/- 40 % of the standard interval duration. After the comparison interval duration, the white disc (comparison offset marker) was presented for 13.3 ms. Following the offset marker of the comparison interval, the fixation dot reappeared and remained on the screen until participants gave their response. Participants were asked to press the ‘y’ key with their left index finger if they judged the first interval as longer and the ‘m’ key with their right index finger if they had had perceived the second interval as longer. The next trial started 500 ms after the key press.

Multimodal bias was manipulated in three different conditions: “unimodal”, “shorter”, and “longer”. To foster multisensory integration in all three conditions, auditory markers were always presented simultaneously with the visual onset and offset markers of the standard interval. For the comparison interval, multimodal bias was introduced similar as in Experiment 1: (1) Only visual marker stimuli were presented in the “unimodal” condition, (2) auditory markers were presented 50 ms *after* and *before* the visual onset and offset markers in the “shorter” condition, respectively, (3) auditory markers were presented 50 ms *before* and *after* the visual onset and offset markers in the “longer” condition, respectively. All three bias conditions were presented in random order.

In each of the experimental sessions, that is, for each of the three standard durations, participants performed 20 practice trials (picked at random from all possible trials) and 15 blocks of 33 trials each (11 Comparison Durations * 3 Bias Conditions). After each block, written feedback containing the percentage of correct responses in the last block was presented. Overall, participants performed 495 experimental trials, thus yielding 15 trials per comparison level in each of the experimental conditions.

Results

For each participant, standard duration and bias condition, the percentage of “comparison longer” responses was computed separately for all comparison durations. Then, a logistic psychometric function (Bush, 1963),

$$P(x_i) = \frac{1}{1 + e^{-(x_i - PSE)/(0.91 \cdot JND)}},$$

was fitted to these data by employing a maximum-likelihood procedure. In this logistic function, the probability *P* of the response “comparison longer” is associated with the comparison level x_i. The Point of Subjective Equality (*PSE*) is defined as the location of this psychometric function, that is, the duration at which the probability for a “comparison longer” response amounts to .5, and thus, the duration at which the participant perceives the comparison to be equal in duration to the standard stimulus. Thus, the *PSE* reflects the perceived duration of the comparison stimulus, with smaller *PSEs* corresponding to longer perceived duration of the comparison stimulus. The Just Noticable Difference (*JND)* amounts to half the interquartile range of the psychometric function. It reflects the steepness of the psychometric function, and thus, the sensitivity with which a participant can discriminate between the standard and the comparison stimulus. The lower the value of *JND*, the higher the discrimination sensitivity. Subsequently, separate repeated-measures ANOVAs with the factors Bias Condition (“unimodal”, “shorter”, and “longer”) and Standard Duration (500, 1,000, and 2,000 ms) were conducted on the estimates of *PSE*, the *Constant Error*, and *JND*.

Point of Subjective Equality. PSE increased with increasing standard duration, *F*(2, 18) = 517.02, *p* < .001. Specifically, mean *PSE* values of 490, 978, and 1,913 ms were observed for the 500, 1,000, and 2,000 ms standard condition. Accordingly, participants generally performed well in the discrimination task, with a slight overestimation of the comparison stimulus for all three standard durations. Furthermore, *PSE* was strongly affected by Bias Condition, *F*(2, 18) = 61.25, *p* < .001. Mean *PSE* was largest in the "shorter" bias condition (1,212 ms), intermediate in the "unimodal" condition (1,124 ms), and smallest in the "longer" condition (1,046 ms). Bonferroni-corrected post-hoc *t*-tests indicated that all bias conditions differed significantly from each other (all *p* < .01). Most importantly, these differences in perceived duration were completely independent of standard duration, *F*(4, 36) = 0.10, *p* = .98 (see Figure 8).

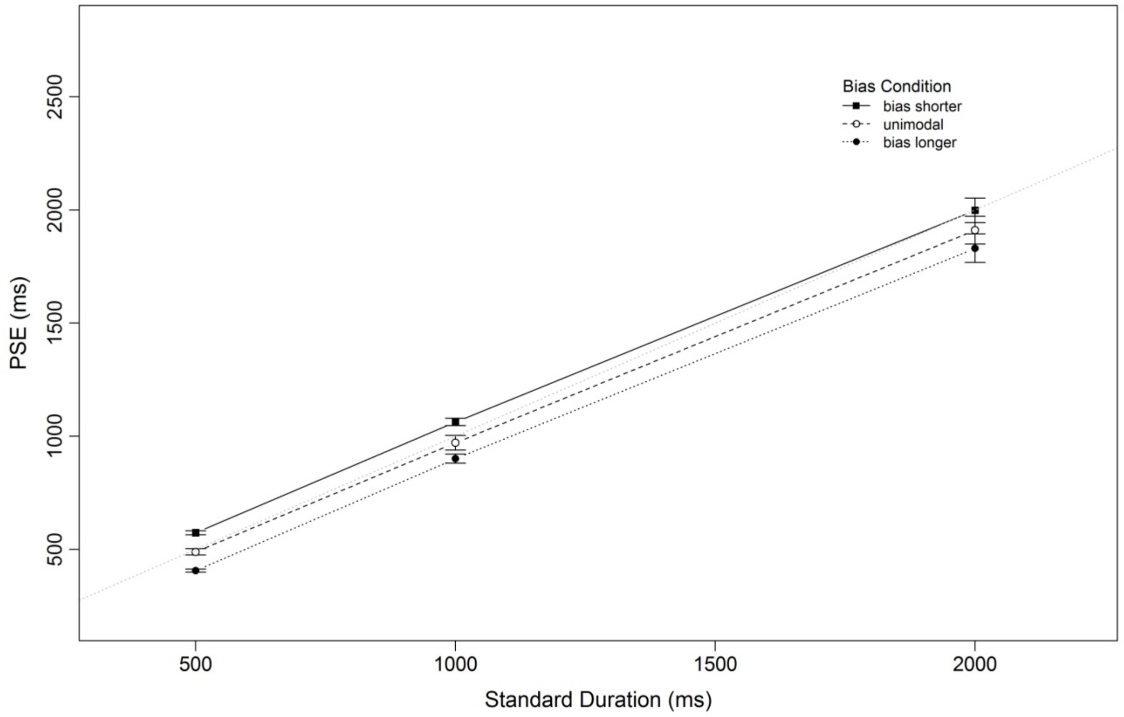

Figure 8. Mean *PSE* in Experiment 3 as a function of Bias Condition and Standard Duration. Error bars represent the standard error of the mean. The veridical line is depicted in gray.

Constant Error. CE in each condition was computed as the difference between *PSE* and the standard duration. In contrast to Experiments 1 and 2, a repeated-measures ANOVA with factors Standard Duration and Bias Condition did not show an effect of standard duration on *CE*, *F*(2, 18) = 1.68, *p* = .22. Thus, no Vierordt effect was present in this experiment. In addition, this analysis showed an effect of Bias Condition (*p* < .001), and no interaction of both factors (*F* < 1), thus replicating the main results of the previous analysis of *PSE*.

Just Noticeable Difference. Unsurprisingly, *JND* increased with increasing standard duration, *F*(2, 18) = 100.49, *p* < .001, with mean values of 52, 101, and 196 ms for the 500, 1,000, and 2,000 ms standard durations. These values correspond to a constant Weber Fraction close to 0.1 for each of the different standard durations. Furthermore, there was a main effect of Bias Condition on *JND*, *F*(2, 18) = 11.04, *p* < .001, and in contrast to *PSE*, this effect was modulated by Standard Duration, *F*(4, 36) = 3.43, *p* = .02. This interaction seems to be mainly caused by *JND* in the "longer" bias condition for the 1,000 ms standard (cf. Figure 9). Indeed, a separate ANOVA comparing *JND*s in the "unimodal" and the "shorter" bias condition for the three different standard durations revealed consistently larger *JND*s, and thus, worse discrimination performance, for the "unimodal" condition, *F*(1, 9) = 15.65, *p* < .01, and no interaction of these conditions with standard duration, *F*(2, 18) = 2.33, *p* =.13. A parallel ANOVA on *JND*s of the "unimodal" and the "longer" bias condition also showed larger *JND*s in the "unimodal" condition, *F* (1, 9) = 9.55, *p* = .01, and in addition, an interaction of these conditions with standard duration, *F* (2, 18) = 3.72, *p* = .04. Post-hoc *t*-tests confirmed higher *JND*s in the "unimodal" condition than in the "longer" condition for the 500 ms, *t*(9) = 3.54, *p* < .01, and the 2,000 ms standard duration, *t*(9) = 2.89, *p* = .02, but not for the 1,000 ms standard duration, *t*(9) = 0.21, *p* = .84.

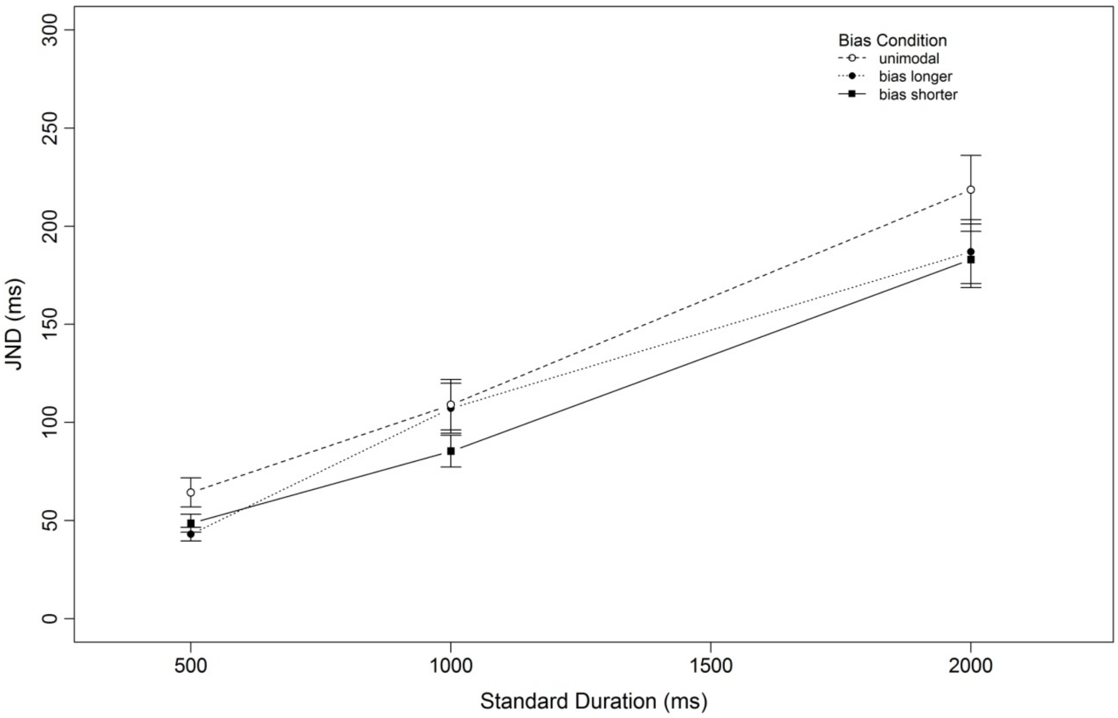

Figure 9. Mean *JND* in Experiment 3 as a function of Bias Condition and Standard Duration. Error bars represent the standard error of the mean.

Discussion

The overall effect of bias condition is consistent with the outcome of our previous experiments. Most importantly, bias condition affected perceived duration, as indexed by *PSE*, such that perceived duration of the visually marked temporal intervals was biased towards the duration of the conflicting auditory intervals. This multimodal bias effect was even more pronounced than in Experiments 1 and 2. Specifically, the true physical difference between the auditory intervals in the "shorter" and "longer" conditions again was 200 ms. Thus, the observed difference of 166 ms between *PSE*s in these two conditions amounts to an observed multimodal bias effect of 83 % of the true physical difference. Therefore, in this experiment, the auditory information seemed to dominate the visual temporal information almost completely.

Most importantly, and in contrast to Experiment 1, this multimodal bias effect was highly similar for all employed standard durations. Thus, irrespective of interval duration, a physical asynchrony between the onset and offset markers of audiovisual intervals translated to changes in perceived duration of comparable magnitude. This is highly consistent with the notion of a temporal ventriloquism effect in perceived duration (Klink et al., 2011; Morein-Zamir et al., 2003). Moreover, this result weakens the alternative account, according to which differences in pacemaker rate play a major role in the perception of audiovisual conflicting intervals.

General Discussion

The present experiments were designed to investigate which mechanisms underlie the multisensory integration of temporal intervals. To this end, we employed a reproduction task (Experiment 1 and Experiment 2) and a paired-comparison task (Experiment 3) to investigate perceived duration of multimodal audiovisual intervals with congruent (same) and conflicting (shorter or longer) durations.

Overall, all Experiments show a strong influence of asynchronously presented auditory intervals on the perceived duration of visual intervals. Specifically, independent of the method used to assess duration perception, perceived duration of visual intervals appeared longest when the irrelevant auditory intervals were longer, and it appeared shortest when the irrelevant auditory intervals were shorter than the relevant visual ones. This pattern of results indicates a multisensory integration of the audiovisual temporal information, such that the auditory duration strongly affected the visual percept. Especially for the paired-comparison task, the auditory duration almost completely dominated the perceived visual duration, whereas this influence was less strong, yet still pronounced, for the reproductions in Experiment 1 and intermediate in Experiment 2[3]. These results therefore support previous evidence (Bausenhart et al., submitted; Klink et al., 2011; Romei et al., 2011) on the multisensory integration of conflicting audiovisual time intervals. Alternatively, however, the results could also be interpreted as reproductions based on the auditory intervals instead of the visual ones in part of the trials. This is rather unlikely, however, because in all Experiments, a high (25% in Experiment 1 and 2, and 33% in Experiment 3) percentage of randomly interspersed unimodal trials would have made such a strategy quite detrimental to performance. Interestingly, the strongest multimodal bias was observed in the Experiment 3, which included the highest proportion of unimodal trials – and in which it would therefore be most harmful to attend only to the auditory stimulation. In addition, Bausenhart et al. (submitted) employed a similar paired-comparison task as in Experiment 3, but presented only one auditory marker (e.g., onset marker alone). Thus, there were no auditory intervals to which participants could respond to. Interestingly, an auditory biasing effect on perceived visual duration was still observed. Therefore, this finding also supports the notion the multisensory bias effect is indeed due to a biased perception of the visual intervals rather than a mere reproduction or judgment of the auditory intervals alone.

Most important for the purpose of the present study was the examination of this multisensory integration effect across a wide range of interval durations. As outlined in the Introduction, the multisensory integration effect should increase with increasing interval duration if it is based on an influence of the auditory information on the pacemaker rate of a pacemaker-accumulator model, whereas an influence on the switch component of this model would lead to a constant effect across interval durations. Surprisingly and contrary to these predictions, the observed multimodal bias in reproduced duration (Experiment 1 and 2) decreased with increasing interval duration. Since this result might have been codetermined by the existence of a pronounced Vierordt effect (Vierordt, 1868), it is hard to draw any decisive conclusions about the mechanisms underlying this effect (see Discussion of Experiment 1 and 2). However, the paired-comparison method with blocked standard durations employed in Experiment 3 prevented the occurrence of a Vierordt effect, and most crucially, the multisensory integration effect was clearly additive with interval duration. Thus, this latter result is highly consistent with the assumption that multimodal integration of conflicting audiovisual time intervals affects the switch component of a pacemaker-accumulator model. Specifically, according to this explanation, in “longer” trials, the switch that regulates the start and end of the accumulation of pulses generated in the pacemaker would close earlier and open later than in “same” trials, and this would result in a higher number of accumulated pulses for a given interval duration. Likewise, in “shorter” trials, the switch would close with some delay and open earlier than in “same” trials, and thus, this would result in a smaller number of accumulated pulses. As suggested by, for example, Klink et al. (2011), this might be caused by a temporal ventriloquism effect (Bertelson & Aschersleben, 2003; Morein-Zamir et al., 2003). Accordingly, if the auditory and visual marker stimuli are presented slightly asynchronously, the perceived onset and offset of the visual interval would be shifted in time towards the onset and offset of the auditory interval. As a consequence, closing and opening of the switch, and thus the number of accumulated pulses, would be determined by the onset and offset of the combined audiovisual percept and not by the physical onset and offset of the visual interval alone. Thus, temporal ventriloquism could readily account for the observed multisensory integration effects on perceived duration of conflicting audiovisual intervals.

Another important theoretical aspect of our data is concerned with the observed difference between the “same” and the “unimodal” conditions. In Experiment 1, this difference was additive with increasing stimulus duration and thus seemed to be at odds with the overadditivity reported in previous studies (Chen & Yeh, 2009; Walker & Scott, 1981). However, this might be explained by the use of empty intervals instead of filled ones as in these previous studies. For that reason, we conducted Experiment 2 by using filled visual and auditory intervals. Here, the difference between the “same” and the “unimodal” condition showed a clear overadditivity with stimulus duration. This result is consistent with the notion of an enhanced arousal evoked by continuous auditory stimuli (Wearden et al., 2007), which leads to an increased pacemaker rate. Finally, however, one might argue that the results of Experiment 3 are at odds with the observed underestimation of “unimodal” compared to “same” intervals in Experiments 1 and 2. Specifically, even though a “same” condition was omitted in Experiment 3, a similar situation was produced in the “unimodal” condition, when participants had to discriminate between the audiovisual congruent standard and the unimodal comparison. In that case, a smaller PSE than the standard duration was observed, indicating that participants judged the unimodal visual comparison even as slightly longer than the audiovisual standard interval. Particularly, this result might be interpreted as either an overestimation of the unimodal visual comparison or an underestimation of the audiovisual standard, and it is not possible to distinguish these alternatives. Since a reminder design was employed in Experiment 3, in which the order of the standard and the comparison is fixed across trials, *PSE* values could have even been inflated by simple response biases (see Green & Swets, 1966) or time order errors (Hellstrom, 1985; Michels & Helson, 1954). Therefore, in this paradigm, absolute *PSE* values, and thus, the appearent overestimation of unimodal intervals, should not be interpreted. Most importantly, however, this does not hamper the interpretation of our results, because our conclusions are not based on absolute *PSE* values, but on a *relative* comparison of *PSE* values, which are all assessed in comparison to a standard stimulus that is identical across conditions. Thus, taking the results of all three Experiments together, unimodal visual intervals are perceived as shorter than bimodal congruent ones. For filled (Experiment 2), but not for empty (Experiment 1) intervals, this effect seems to be caused by an increase in pacemaker rate.

More central to the aim of the present study is the multimodal integration of conflicting bimodal intervals. Interestingly, even though the irrelevant auditory information in these trials was different from the relevant visual duration information, and thus potentially misleading, this discrepancy did not lead to a decline of discrimination sensitivity. Specifically, the variability of duration reproductions in bimodal trials of Experiment 1 and 2 was consistently similar to or even smaller than the variability of unimodal visual reproductions. Likewise, in Experiment 3, *JNDs* were generally larger in the “unimodal” condition than in the bimodal conditions. The only exception from this pattern was found for the combination of the “longer” condition and the 1,000 ms standard duration, in which *JND* did not differ from *JND* of the “unimodal” condition. This effect appears rather unsystematic and is therefore hard to interpret. If this effect was directly associated with the specific 1,000 ms standard interval duration, it should have been observed for the “shorter” bias condition as well. Likewise, it cannot be due to the presentation of irrelevant auditory stimuli, which might have a distracting influence, because then it should also be evident for other standard interval durations. Even though we have no conclusive explanation for this effect, it should be stressed, however, that there was no cost associated with the presentation of conflicting auditory information. Thus, all experiments suggest that multisensory integration of conflicting intervals leads to relatively stable percepts without accompanying costs, but rather benefits, regarding discrimination sensitivity (Bausenhart et al., submitted).

This finding is also consistent with previous studies from other domains, such as spatial localization or size discrimination (Alais & Burr, 2004; Ernst & Banks, 2002; Ernst & Bulthoff, 2004). Specifically, Ernst and Banks (2002) investigated the crossmodal integration of visual and haptic stimuli in size discrimination. These authors quantified the degree to which vision or haptics dominate the resulting multimodal percept by measuring the variances associated with each modality. Based on a Maximum-Likelihood process, they suggested that the variance associated with a given modality determines the degree of dominance exerted by this modality, in a statistically optimal fashion. Applying this model to our results, it might be suggested that smaller variability observed in bimodal intervals might be attributed to the typically smaller variance associated with auditory duration judgments compared to visual duration judgments, and the resulting relative auditory dominance over visual information. A dedicated investigation of this account, however, would require the assessment of the variability of unimodal visual and auditory as well as bimodal (audiovisual) stimuli. Since we did not include a unimodal auditory condition, we cannot ascertain whether our data conform to the principle of statistically optimal integration or not, even though the general pattern of results seems to be well in line with this assumption. Future studies might address this issue in more detail, and try to determine whether temporal ventriloquism, for example, is mediated by such a statistically optimal integration process.

The present results are not only important regarding multisensory integration of temporal information, but they also may be especially interesting from the viewpoint of time perception research, because eventually they may inform about more basic mechanisms of time perception per se. Specifically, even though there is evidence for effects of several manipulations on pacemaker speed (e.g., Chen & Yeh, 2009; Penton-Voak et al., 1996; Wearden & Penton-Voak, 1995), there is astonishingly little research on the role of the switch mechanism of the pacemaker-counter model. This is especially surprising with regard to the central role of this model component for the perception of time. The switch mechanism is typically assumed to be triggered directly by physical input (e.g., stimulus onset or offset). The timing of physical stimulus input, however, cannot be manipulated independently of interval duration. If a stimulus is, for example, modified to end later, its duration will consequently also be longer. This natural confound might actually have prevented researchers from paying much attention to a more direct investigation of the role of the switch mechanism for time perception so far (for a rare exception, cf. Bendixen, Grimm, & Schröger, 2006). Establishing temporal-ventriloquism-like effects for conflicting multimodal intervals might thus not only be interesting with regard to multisensory integration, but eventually also provide a means to manipulate and study the proposed switch mechanism of the pacemaker-counter model more directly.

In summary, we demonstrated that auditory information influences perceived visual duration, such that visually marked intervals were biased towards conflicting auditory ones and, more importantly, these bias effects showed a decreasing or constant magnitude with the lengthening of interval duration. Thus, our study supports the existence of a temporal ventriloquism effect in perceived multimodal duration. This effect biases the moment at which the switch of a pacemaker-accumulator mechanism opens and closes towards the auditory interval markers and consequently, evokes a biased perception of visual duration.

Acknowledgments

This study was supported by Grant BA4110/3-1 of the German Research Foundation (DFG). We thank Julia Feyerabend and Laura Schneider for their assistance in data acquisition and Anja Fiedler and three anonymous reviewers for helpful comments on a previous version of this manuscript.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol, 14*, 257-262. doi: 10.1016/s0960-9822(04)00043-0

Bausenhart, K. M., de la Rosa, M. D., & Ulrich, R. (submitted). Multimodal integration of time: visual and auditory contributions to perceived duration and sensitivity.

Bausenhart, K. M., Dyjas, O., & Ulrich, R. (in press). Temporal reproductions are influenced by an internal reference: Explaining the Vierordt effect. *Acta Psychol*. doi: 10.1016/j.actpsy.2013.06.011

Bendixen, A., Grimm, S., & Schröger, E. (2006). The relation between onset, offset, and duration perception as examined by psychophysical data and event-related brain potentials. *J Psychophysiol, 20*, 40-51. doi: 10.1027/0269-8803.20.1.40

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *Int J Psychophysiol, 50*, 147-155. doi: 10.1016/S0167-8760(03)00130-2

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*, 433-436.

Bush, R. R. (1963). Estimation and evaluation. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 429-469). New York: Wiley.

Chen, K.-M., & Yeh, S.-L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychol, 130*, 225-234. doi: 10.1016/j.actpsy.2008.12.008

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn Sci, 8*, 162-169. doi: 10.1016/j.tics.2004.02.002

Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception, 36*, 1089-1103. doi: 10.1068/P5741

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychol Rev, 84*, 279-325. doi: 10.1037/0033-295x.84.3.279

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. *Ann N Y Acad Sci, 423*, 52-77. doi: 10.1111/j.1749-6632.1984.tb23417.x

Goldstone, S., & Lhamon, W. T. (1972). Auditory-visual differences in human temporal judgment. *Percept Mot Skills, 34*, 623-633. doi: 10.2466/pms.1972.34.2.623

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hellstrom, A. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychol Bull, 97*, 35-61. doi: 10.1037/0033-2909.97.1.35

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nat Neurosci, 13*, 1020-1026. doi: 10.1038/nn.2590

Klink, P. C., Montijn, J. S., & van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Atten Percept Psychophys, 73*, 219-236. doi: 10.3758/s13414-010-0010-9

Lejeune, H., & Wearden, J. H. (2009). Vierordt's the experimental study of the time sense (1868) and its legacy. *Eur J Cogn Psychol, 21*, 941-960. doi: 10.1080/09541440802453006

Michels, W. C., & Helson, H. (1954). A quantitative theory of time-order effects. *Am J Psychol, 67*, 327-334.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cogn. Brain Res, 17*, 154-163. doi: 10.1016/s0926-6410(03)00089-2

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis, 10*, 437-442.

Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *J Exp Psychol Anim Behav Process, 22*, 307-320. doi: 10.1037/0097-7403.22.3.307

Romei, V., De Haas, B., Mok, R. M., & Driver, J. (2011). Auditory stimulus timing influences perceived duration of co-occurring visual stimuli. *Front Psychol, 2*, 215. doi: 10.3389/fpsyg.2011.00215

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Crossmodal temporal discrimination: assessing the predictions of a general pacemaker-counter model. *Percept Psychophys, 68*, 1140-1152. doi: 10.3758/BF03193716

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen [The experimental study of the time sense]*. Tübingen, Germany: Laupp.

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *J Exp Psychol Hum Percept Perform, 7*, 1327-1339.

Wearden, J. H. (2003). Applying the scalar timing model to human time psychology: progress and challenges. In H. Helfrich (Ed.), *Time and mind II: Information processing perspectives.* (pp. 21-39). Ashland, OH: Hogrefe & Huber Publishers.

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why 'sounds are judged longer than lights': application of a model of the internal clock in humans. *Q J Exp Psychol B, 51B*, 97-120.

Wearden, J. H., Norton, R., Martin, S., & Montford-Bebb, O. (2007). Internal clock processes and the filled-duration illusion. *J Exp Psychol Hum Percept Perform, 33*, 716-729. doi: 10.1037/0096-1523.33.3.716

Wearden, J. H., & Penton-Voak, I. S. (1995). Feeling the heat: body temperature and the rate of subjective time, revisited. *Q J Exp Psychol B, 48B*, 129-141.

Woodrow, H. (1934). The temporal indifference interval determined by the method of mean error. *J Exp Psychol, 17*, 167-188. doi: 10.1037/h0070235

Footnotes

[1] Yet, due to the nature of the oddball task employed by Chen and Yeh (2009), it remains somewhat unclear whether the observed difference in perceived duration between unimodal visual and bimodal stimuli emerged because the auditory information within the bimodal stimuli increased the pacemaker rate. Alternatively, adding the auditory information to the visual stimulus might have also increased the perceived “oddness” of the oddball stimulus. In this case, the increase in pacemaker rate would be due to an enhanced oddball effect rather than to the multimodal stimulation per se.

[2] This and following experiments were conducted in accordance with the Helsinki Declaration and the ethics guidelines for scientific work of the University of Tübingen.

[3] This was supported by Bonferroni corrected two-sample *t*-tests conducted on the amount of multisensory bias. Therefore, the observed difference between the auditory intervals in the “shorter” and “longer” conditions was converted into a percentage, with the true physical difference of 200 ms corresponding to 100%. According to these analyses, the multimodal bias in Experiment 3 was more pronounced than in Experiment 1 ($p < .01$). In Experiment 2, the multimodal bias tended to be larger than in Experiment 1 ($p = .08$), but did not differ from Experiment 3 ($p = .51$).

APPENDIX C – Study 3

De la Rosa, M. D., & Bausenhart, K. M. (2018). Enhancement of letter identification by concurrent auditory stimuli of varying duration. *Acta Psychologica*, *190*, 38–52. https://doi.org/10.1016/j.actpsy.2018.07.001

Enhancement of Letter Identification by Concurrent Auditory Stimuli of Varying Duration

Maria D. de la Rosa & Karin M. Bausenhart

University of Tübingen

Address editorial correspondence to Maria Dolores de la Rosa Gamiz, Fachbereich Psychologie, University of Tübingen, Schleichstrasse 4, D-72076 Tübingen, Germany.

Phone: +49 (0)7071 29-76423, FAX: +49 (0)7071 29-2410, email: maria-dolores.de-la-rosa-gamiz@uni-tuebingen.de

Abstract

Previously it has been shown that the concurrent presentation of a sound can improve processing of visual information at higher perceptual levels, for example, in letter identification tasks. Moreover, increasing the duration of the concurrent sounds can enhance performance in low-level tasks as contrast detection, which has been attributed to a sustained visual activation corresponding to the duration of the sound. Yet, the role of sound duration has so far not been investigated in higher-level visual processing. In a series of five Experiments, we again demonstrated that the mere presence of a concurrent sound can enhance the identification of a masked, centrally presented letter compared to unimodal presentation, even though this benefit was absent in one experiment for high-contrast letters yielding an especially high level of task-performance. In general, however, the sound-induced benefit was not modulated by a variation of target contrast or by the duration of the target-to-mask interstimulus interval. Taking individual performance differences into account, a further analysis suggested that the sound-induced facilitation effect may nevertheless be most pronounced at specific performance levels. Beyond this general sound-induced facilitation, letter identification performance was not further affected by the duration of the concurrent sounds, even though in a control experiment it could be established that letter identification performance improved with increasing letter duration, and perceived letter duration was prolonged with increasing auditory duration. The results and their interpretation with respect to the large observed interindividual performance differences are discussed in terms of potential underlying mechanisms of multisensory facilitation, as preparedness enhancement, signal enhancement, and object enhancement.

Keywords: backwards-masking paradigm; letter identification; multisensory integration; duration perception

Enhancement of Letter Identification by Concurrent Auditory Stimuli of Varying Duration

Research on multisensory perception has shown that concurrent auditory stimulation can facilitate visual perceptual sensitivity. Importantly, existing evidence shows that this auditory benefit can be observed at different levels of visual perceptual processing, from simple detection of separable features at low-level perceptual stages to identification of complex object representations comprised of combinations of stimulus features at higher-level perceptual stages.

For example, sounds facilitate low-level perceptual processes, resulting in an increased sensitivity for rather simple near- or supra-threshold visual stimuli such as Gabor patches, lines, or simple shapes, defined by elementary features as luminance, contrast, color, or orientation (Y. C. Chen, Huang, Yeh, & Spence, 2011; de Haas, Cecere, Cullen, Driver, & Romei, 2013; Frassinetti, Bolognini, & Làdavas, 2002; Lippert, Logothetis, & Kayser, 2007; Ngo & Spence, 2010; Noesselt et al., 2010; Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008; Pérez-Bellido, Soto-Faraco, & López-Moliner, 2013; Senkowski, Saint-Amour, Höfle, & Foxe, 2011; Stein, London, Wilkinson, & Price, 1996; Vroomen & Gelder, 2000; Vroomen & Keetels, 2009). More scarce is evidence regarding the influence of sounds on processing in higher stages of visual perception, regarding processing of more complex stimuli (such as, e.g., letter identification). Yet, Y. C. Chen & Spence (2011) were able to demonstrate such an influence by employing a backward masking paradigm. They presented a peripheral target letter for 40 ms, followed by a masking letter after a variable interstimulus interval (ISI) between 0 and 133 ms. Importantly, the target letter was presented either alone or together with a 27-ms sound, synchronized to letter onset. The authors found that the presence of the sound enhanced letter identification performance, but only in a specific range of ISI durations between 27 and 40 ms (Experiment 1). Similar facilitatory effects at the 40-ms ISI were observed for high-intensity targets combined with low-intensity masks and for low-intensity targets combined with high-intensity masks (Experiment 2). Moreover, the authors found that presenting the sound concurrently with the masking letter also improved target letter identification (Experiment 3). Additionally, this crossmodal facilitation depended on a reliable temporal coincidence of the target letter and sound (Experiment 4) but, otherwise, their spatial consistency was not crucial (Experiment 5).

Taking together these results, the authors concluded that concurrent sounds facilitate the identification of masked letters by two mechanisms. First, a bottom-up mechanism might serve to produce crossmodal facilitation within a specific range of ISI durations. Accordingly, whenever the ISI is long enough to warrant that the target can be temporally segregated from the mask, such that two separable object representations are created, sounds would automatically enhance letter identification by strengthening the object representation of the target letter. This *object-enhancement* account was preferred over a simpler *signal-enhancement* account and a *preparedness-enhancement* account. The former idea would imply performance improvements not at intermediate but especially at very brief ISIs, when the perceptual signal corresponding to the target is most weakened by the masking stimulus. The latter idea, in contrast, would imply performance improvements based on improved alertness or preparation for target processing, across a wide range of ISIs from very brief to longer ones - as long as performance is below ceiling. Furthermore, the authors assumed a second, more controlled top-down process, which might underlie the necessary reliable temporal coincidence of target and sound for crossmodal facilitation. Specifically, this mechanism would be based on a “unity assumption” (Y. C. Chen & Spence, 2017; Welch & Warren, 1980). According to this assumption, an integration of auditory and visual stimuli only takes place if they are assumed to emerge from the same perceptual object.

This finding on crossmodal facilitation of letter identification implies that higher-level visual perceptual processes, involved in combining simple features to create complex object representations (e.g., Madec, Rey, Dufau, Klein, & Grainger, 2011; Treisman & Gelade, 1980), under certain conditions, can be affected by the concurrent presentation of uninformative sounds. In their study, however, Y.-C. Chen & Spence (2011) only employed a single duration (27 ms) for sounds which was shorter than letter duration (40 ms). Based on their finding that reliable temporal coincidence is crucial to observe the crossmodal facilitation effect, one might wonder whether masked letter identification might be modulated by varying the duration of the concurrent sound relative to the visual target duration. Actually, several previous studies have demonstrated that not only the presence, but also the duration of accompanying sounds can affect perception of visual stimuli. For example, perceived duration of visual intervals is strongly biased by concurrently presented auditory intervals with conflicting durations (Asaoka & Gyoba, 2016; Bausenhart, De la Rosa, & Ulrich, 2014; K. M. Chen & Yeh, 2009; De la Rosa & Bausenhart, 2013; Hartcher-O'Brien, Di Luca, & Ernst, 2014; Klink, Montijn, & van Wezel, 2011; Walker & Scott, 1981). Specifically, visual intervals are perceived as longer or shorter, compared to a unimodal visual baseline, when they are accompanied by longer or shorter auditory intervals, respectively. Based on the observation of constant effect magnitudes across different interval durations, this effect has been explained in terms of a temporal ventriloquism effect (Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Hereby, it is assumed that the onset/offset of the visual stimulus is biased towards the auditory on- and offset (De la Rosa & Bausenhart, 2013; see also Klink et al., 2011), and thus, the visual duration percept is biased towards the auditory duration.

Importantly, de Haas et al. (2013) investigated the effect of sound durations also on the sensitivity of visual perceptual processing by employing a low-level perceptual task. Specifically, their participants had to detect a Gabor patch that could be presented in one of two consecutive intervals of dynamic white noise. The task was divided into two different conditions. In the unimodal condition only the Gabor patch was presented, lasting for one of eight different durations between ~24 and ~190 ms. In the audiovisual condition, the Gabor patch lasted for a fixed duration of ~24 ms and was always accompanied by a continuous pure tone whose duration varied between the same eight durations as the visual target in the unimodal task. Unsurprisingly, in the unimodal task, target detection sensitivity increased with increasing duration of the visual target. More important, also in the audiovisual condition, detection sensitivity for the 24-ms Gabor patch increased with increasing sound duration from 24 ms up to 60 and 96 ms. The authors concluded that early audiovisual interactions might be responsible for this effect. Specifically, the increasing sound duration might have produced sustained visual activation, resulting in a longer visual target representation, which in turn would facilitate target detection. Thus, this proposed mechanism is basically consistent with the temporal ventriloquism explanation of biased duration perception as outlined above. However, a controversial aspect of the results concerns the comparison across the unimodal and the audiovisual condition at the shortest stimulus duration (24 ms). This comparison showed that visual detection sensitivity was actually impaired by presenting a same duration auditory stimulus, in comparison to the unimodally presented visual stimulus. This finding seems somewhat inconsistent with previous studies reporting sound-induced facilitation of stimulus processing (e.g., Frassinetti et al., 2002; Noesselt et al., 2010, 2008), and presumably points to an additional disruptive effect of sound presentation on the processing of the visual information.

In summary, the presentation of concurrent sounds can influence different stages of visual perceptual processing: from rather low-level processing such as detection of simple stimuli defined by elementary features, to higher-level processing required for the identification of complex stimuli, such as letters. Importantly, it has been also shown that the duration of the sound can further modulate low-level processing by improving contrast detection with increasing sound duration. However, no such evidence has been reported so far for higher-level visual processing.

The aim of the present study was two-fold. First, we attempted a conceptual replication of the sound-induced enhancement of letter identification observed by Y. C. Chen & Spence (2011) in order to investigate whether this effect is robust to variations of the specific conditions of stimulation. In fact, only few studies have so far reported related sound-induced effects on processing of rather complex stimulus configurations, but within rather dissimilar experimental paradigms (i.e., letter processing within the attentional blink, Kranczioch & Thorne, 2013; or discrimination of complex dot configurations, Takeshima & Gyoba, 2014). Therefore, we combined a letter identification task with a backwards-masking paradigm (Kinsbourne & Warrington, 1962; Rolke, 2008; Turvey, 1973). Specifically, participants had to identify a centrally presented letter followed by a mask consisting of visual random noise (masking-by-noise). Similar to the study of Y. C. Chen & Spence (2011), the target letter could be either presented unimodally or accompanied by a concurrent sound. Second, we wanted to investigate whether the duration of the concurrent sounds would further modulate the influence of the sound on letter identification. Therefore, we employed sound of either the same duration as the visual target or conflicting (shorter or longer) durations. Based on Y. C. Chen & Spence (2011), the presentation of any sound should strengthen the object representation of the target letter, and thus facilitate letter identification performance compared to the unimodal condition. In addition, if the duration of the auditory stimulus modulates the persistence of the internal object representation of the target letter, letter identification performance should be further modulated by sound duration (cf. de Haas et al., 2013).

Experiment 1

In Experiment 1, we investigated the influence of sound on the identification of letters in a backwards-masking paradigm. Therefore, the target letter was presented either unimodally or simultaneously accompanied by a sound with congruent (same) or conflicting (longer or shorter) duration. According to Y. C. Chen & Spence (2011), the presentation of any sound should increase the accuracy of letter identification in comparison with the unimodal condition, and this effect should emerge especially at intermediate ISIs. Additionally, if the duration of the concurrent sound can modulate letter identification by temporally shortening or expanding the corresponding visual object representation (cf. de Haas et al., 2013), letter identification performance should increase with the duration of the concurrent sounds.

Method

Participants. Twenty-four participants (21 female; age range: 18-47 years; mean age: 24.38 years) from the University of Tübingen voluntarily took place in Experiment 1 in exchange of either course credit or money (8€/hour). Please note that this and all following experiments were conducted according to the ethics guidelines for scientific work of the University of Tübingen. All participants of this and the following experiments provided written informed consent before performing the experiment, and reported normal or corrected-to-normal vision and hearing abilities.

Apparatus and Stimuli. The experiment was run on a Macintosh computer connected to a CRT monitor (iiyama Vision Master^{TM} Pro 454) running at 150 Hz. The presentation of all stimuli and data recording were controlled by Matlab with the Psychophysics Toolbox extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Participants were seated in a dimly lit room at a viewing distance of approximately 55 cm from the computer screen.

All visual stimuli were presented in white (99 cd/m^2) on a black background (< 1 cd/m^2), resulting in a luminance difference of > 0.98 in units of Michelson contrast (computed as $(L_{max} - L_{background})/(L_{max} + L_{background})$). These visual stimuli consisted of a fixation point defined by a small dot (0.10° visual angle), the target stimuli composed of 8 possible letters (A, S, D, F, G, H, J, or K, all approximately 0.51° x 0.41°), a question mark (0.51° x 0.31° visual angle), and a static random noise mask. This mask consisted of 200 white squares (0.1° visual angle) which were drawn on the pixel-level (not smoothed) and randomly distributed on a black square of a 2.05° x 2.05° visual angle. In each trials, a new random pattern of dot locations was created. All these stimuli were presented at the center of the screen and, except for the question mark, centered within a white frame (2.05° x 2.05° visual angle) in order to reduce spatial uncertainty.

Auditory stimuli consisted of pure sinusoidal tones (800 Hz, 65 dB(A)) played binaurally through headphones. The duration of the tones were 7, 20 or 33 ms[1], presented with ramped 2 ms on- and offsets. Participants responded to the target letters by pressing the corresponding adjacent keys of a standard German keyboard. Correctness of stimulus timing and synchrony / asynchrony between the visual and auditory stimulation was checked by oscilloscopic measurement.

Procedure. At the beginning of the experiment participants received verbal and written instructions to press the key corresponding to the perceived target stimulus (i.e., either A, S, D, F, G, H, J or K) as accurately as possible, as soon as the question mark would appear. They were also informed that sometimes also sounds would be presented, but that these were task-irrelevant.

Each trial started with the presentation of the fixation point together with the white frame. After 1000 ms, the fixation point was replaced by one of the possible target letters for 20 ms. The target letter could be presented in four different conditions (Figure 1). In the no-sound condition, the target letter was presented unimodally. In the audiovisual conditions, the target letter was accompanied by a sound that varied in duration according to three different conditions: longer, same and shorter. In the longer

---

[1] All durations employed in this and the following experiments were multiples of the minimum refresh duration of the employed display (i.e. 6.667 ms), but were rounded to whole milliseconds in the main text.

condition, the (33-ms) sound started 7 ms before the target letter onset and ended 7 ms after the target letter offset. In the same condition, the (20-ms) sound and the target letter started and ended synchronously. In the shorter condition, the (7-ms) sound started 7 ms after the target letter onset and ended 7 ms before the target letter offset. At target stimulus offset, the white frame remained empty for an ISI of either 0, 7, 13, 20, 27, or 33 ms. At the end of the ISI, the white frame was filled with a random visual noise pattern for 500 ms. Afterwards, the mask and white frame were replaced by the question mark which prompted participants to respond, for a maximum response time of 2000 ms. After the response was registered or after the 2000 ms had expired, participants received feedback about their responses for 500 ms: the word “Korrekt!” (“Correct”) colored in green indicated a correct response and “Falsch!” (“False”) colored in red indicated a wrong response. In case of no key pressed, “zu langsam” (“too slow”) was presented.

192 different trials resulted from the combination of 8 target stimuli, 4 sound conditions and 6 ISIs. Each combination was presented 5 times, resulting in 960 trials overall, in random order. The experiment consisted of a practice block with 20 trials (chosen randomly from all possible trials) and the 960 experimental trials, subdivided in 24 experimental blocks with 40 trials each. At the end of each block, participants received feedback about the percentage of correct responses in the preceding block. The whole experiment took around 1 hour.

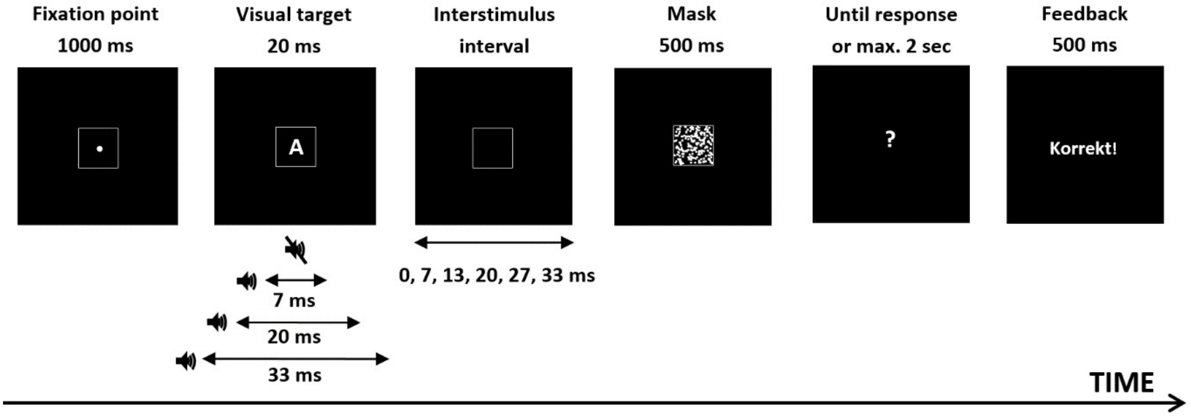

Figure 1. Schematic representation of events in a trial in Experiment 1.

Results

Practice blocks and trials with no response (0.02 %) were discarded from data analysis. A repeated-measures analysis of variance (ANOVA) was conducted with the factors Sound Condition (longer, same, shorter, and no sound) and ISI (0, 7, 13, 20, 27, and 33 ms) for the mean percentage of correct responses)[2]. For the following analyses, when appropriate, *p* values were adjusted for violations of the sphericity assumption using the Greenhouse-Geisser correction.

Overall, participants´ mean percentage of correct responses was 76.79 %, ranging from 45.93 % to 94.69 % between participants. More correct responses were obtained with increasing ISI, $F(5, 115) = 140.33$, $p <. 001$, $\eta^2_p = .86$ (see Fig. 2), with percentages monotonically increasing from 33.04% (*SD* = 18.41%) at the shortest ISI to 95.57% (*SD* = 4.17%) at the longest ISI. However, letter identification was not affected by Sound Condition, even though it showed a tendency, $F(3, 69) = 2.45$, $p = .07$, $\eta^2_p = .10$. Numerically, participants achieved 75.94 (*SD* = 13.11), 77.08 (*SD* = 12.35), 76.70 (*SD* = 12.54), and 77.43 (*SD* = 12.59) percent of correct responses in the no-sound, shorter, same, and longer conditions, respectively. Finally, the interaction between Sound Condition and ISI was not significant, $F(15, 345) = 0.96$, $p = .50$, $\eta^2_p = .04$.

Based on the results of Y. C. Chen & Spence (2011), one would expect that the presence of a sound would affect letter identification performance only at intermediate ISIs producing performance levels of approximately 60-70%. In our design, this corresponds to an ISI of 7 ms (cf. Fig. 2). Therefore, to mirror the analysis of Y. C. Chen & Spence (2011), we conducted an additional *t*-test comparing the percentage of correct responses between the no-sound and the same condition at this ISI. This analysis showed only marginally better performance in the same ($M = 63.23\%$, $SD = 25.03\%$) than in the no-sound ($M = 59.48\%$, $SD = 27.50\%$) condition, $t(23) = 2.04$, $p = .05$ (at all other ISIs, this comparison was far from significant, all *p*s > .43).

---

[2] An anonymous reviewer pointed out that the observed data could alternatively be analysed by fitting psychometric functions to the proportion of correct responses (PC) depending on ISI. We have conducted such analyses and present their results in a supplementary material to this manuscript. In summary, with a single exception, these analyses are in line with all results of the (pre-planned) ANOVAs on percentage of correct responses reported in the main text.

A second pre-planned analysis was conducted to specifically address the influence of sound duration on letter identification. Therefore, we explicitly compared performance between the three bimodal conditions with an additional ANOVA. Clearly, letter identification performance did not differ between the three sound conditions, and this effect was also not modulated by ISI (both *Fs* < 1), while overall performance again increased across ISIs, $F(5, 115) = 134.22, p < .001, \eta^2_p = .85$.

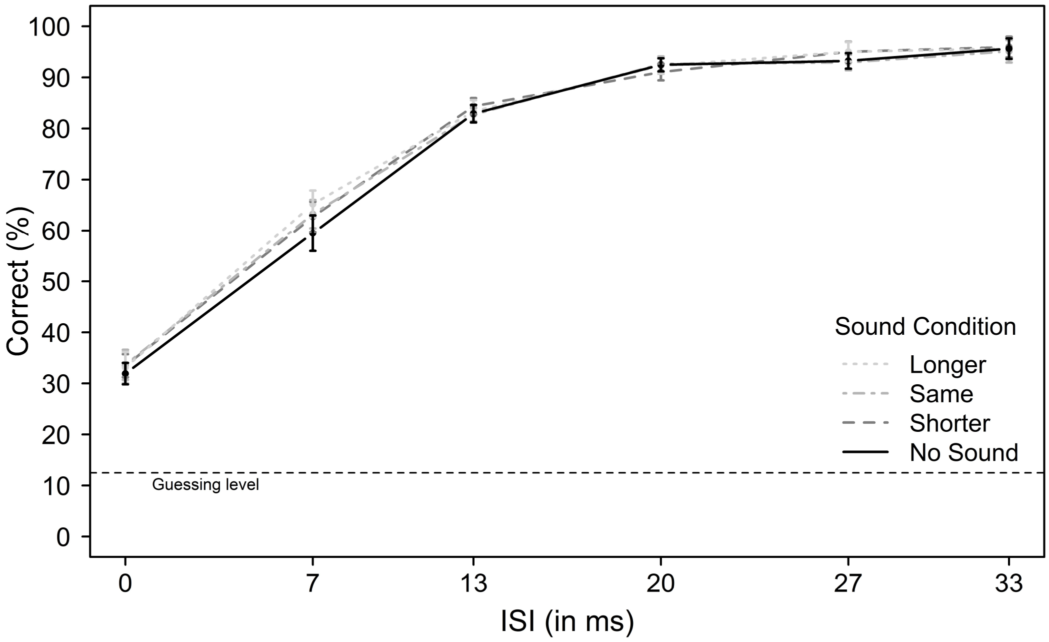

Figure 2. Mean percentage of correct responses in Experiment 1 as a function of Sound Condition and ISI. Error bars represent ±1 within-subject standard error of the mean with a correction for between-subjects variability according to Morey (2008).

Discussion

Experiment 1 was designed to test two major hypotheses: First, we wanted to test for a sound-induced facilitation of letter identification similar to the one observed by Y. C. Chen & Spence (2011). Second, we wanted to investigate whether the duration of the concurrent sounds would further modulate their effect on letter identification.

The results show only a slight trend for sounds affecting the accuracy of letter identification. Basically, the mere presence of a sound in the same condition, as compared to the unimodal condition, produced only a marginal improvement of letter identification performance, and even this weak influence was constricted to the 7-ms ISI. Clearly, this effect must be interpreted with care, since it is based on a post-hoc analysis and only marginally significant. Supposing its validity, this result would reflect the basic result pattern of Y. C. Chen & Spence (2011) quite well: sound-induced enhancement of letter identification at intermediate ISIs producing about 60-70% of correct responses, but not briefer or longer ones.

While Y. C. Chen & Spence (2011) found this crossmodal facilitation within a specific range of ISIs between 27 and 40 ms, participants already reached a similar level of performance at the 7 ms ISI in our study. For all longer ISIs, accuracy already reached relatively high levels above 80% (83.33, 92.10, 94.04, and 95.57% at ISIs of 13, 20, 27, and 33 ms, respectively). Presumably some aspects of our specific employed masking paradigm yielded a relative low task difficulty, leading to a much steeper initial slope of the masking-ISI function, compared to the one of Y. C. Chen & Spence (2011). This might in turn have hampered the chance to observe a clearer sound-induced enhancement of letter identification across a wider range of ISIs.

There are several crucial differences between the basic masking paradigms employed in our and Y. C. Chen & Spence´s (2011) study that might have contributed to the overall higher performance level and lack of clear multisensory integration effect in our study. First, we presented target letters always at the screen centre, as opposed to target letters presented laterally at the left or right of fixation. Central target presentation typically leads to enhanced perceptual performance regarding tasks requiring spatial resolution such as letter identification, since processing at the fovea is characterized by higher spatial resolution and contrast sensitivity (e.g., Bouma, 1970; Pointer & Hess, 1989; Rijsdijk, Kroon, & van der Wildt, 1980). Also, compared to the task of Y. C. Chen & Spence (2011), our stimulus layout evoked less spatial uncertainty and thus processing may have benefitted from preallocated spatial attention (e.g., Carrasco, Williams, & Yeshurun, 2002; Henderson, 1991; Pestilli & Carrasco, 2005). Moreover, according to studies investigating the distribution of neurons which are sensitive to bimodal stimulation across the visual receptive field, multisensory integration per se may increase with increasing eccentricity of the target (Allman & Meredith, 2007). Second, we used a random noise pattern as masking stimulus (i.e., masking-by-noise), as opposed to a structurally similar masking letter (i.e., masking-by-structure), which was quite often confused with and thus misreported as the target letter in the study of Y. C. Chen & Spence (2011, cf. Fig. 1). Interestingly, and consistent with the present results, the largest performance impairment (i.e., the minimum of the ISI-masking function) is usually obtained at briefer ISIs for noise pattern masks than for structure pattern masks (see also, Agaoglu, Agaoglu, Breitmeyer, & Ogmen, 2015). Third, we employed a higher contrast between background and target/mask than Y. C. Chen & Spence (2011). Even though these authors could demonstrate that at least manipulating the relative luminances of target and mask does not necessarily affect the sound-induced enhancement effect, their experiment 2 also indicates that a relatively lower target-background contrast leads to much lower letter identification performance at brief ISIs (see also Rolke, 2008).

Interestingly, there is empirical evidence that stimulus contrast seems to be an important factor modulating the sound-induced enhancement of visual processing (Noesselt et al., 2010). Specifically, some studies report an auditory influence on visual sensitivity exclusively for low or intermediate levels of stimulus contrast (Y. C. Chen et al., 2011; Noesselt et al., 2010; Senkowski et al., 2011). For example, Noesselt et al. (2010) measured the sensitivity of participants to detect a Gabor patch presented either with low or high contrast and presented either unimodally or with a concurrent same-duration sound. The results revealed an enhancement of visual detection sensitivity specifically when the sound co-occurred with a low-contrast, but not with a high-contrast visual target. This specific sound-induced facilitation of low-contrast visual targets is consistent with the principle of “inverse effectiveness” (Meredith & Stein, 1983, 1986; Stein & Meredith, 1993). Considered as a principle of multisensory integration, this phenomenon describes that the maximal enhancement of a multisensory response occurs when the constituting unisensory stimuli evoke a weak response in isolation, and thus implies early multisensory interactions through which a sound may boost the perceptual representation of a relatively weak visual signal. Accordingly, it might be possible that the high-contrast letter identification required in our study did not strongly profit from the sound presentation because the unisensory visual signals provided reliable information that led to high accuracy by itself. In other, rather descriptive, terms, the sounds in our study may not have led to clear sound-induced benefits because of a ceiling effect, as is indicated by the asymptotic course of the ISI-masking function starting from an ISI duration of approximately 20 ms (cf. Fig. 2).

Given this result, it is also not possible to evaluate our second empirical question: Since the sound did not or at best slightly modulate letter discrimination performance, one cannot draw any conclusions concerning the (lack of) effect of sound duration on discrimination performance. Therefore, in Experiment 2, we employed a lower contrast between target and background, in order to reduce overall performance levels and obtain a shallower slope of the ISI-masking function indexing multisensory masked letter discrimination.

Experiment 2

In Experiment 1, there was no clear evidence for a sound-induced benefit on letter identification, and the duration of the auditory stimuli did not modulate letter identification at all. Possibly, the high contrast letters employed in Experiment 1 yielded a high performance level, thus preventing clear additional benefits evoked by the concurrent sound presentation. Therefore, we conducted a second experiment employing the same masked visual letter identification task as Experiment 1, but presenting letters at a lower contrast. Returning to our main hypothesis, we expected that the presentation of any sound would produce an enhancement of letter identification compared with the unimodal condition. Moreover, if sound duration can modulate the duration of the perceptual representation of the letter, letter identification performance should also increase with increasing sound duration.

Method

Participants. Twenty-four participants (19 female; age range: 18-33 years; mean age: 22.42 years) from the University of Tübingen voluntarily took place in Experiment 2 in exchange of either course credit or money (8€/hour). Three participants with an average percentage of correct responses below 25% were discarded from data analyses. Therefore, 21 participants remained for data analyses.

Apparatus, Stimuli, and Procedure. Apparatus, stimuli und procedure were identical to those employed in Experiment 1, with the following exceptions: Different luminances for the target stimulus (97 cd/m^2) and for the background (46.6 cd/m^2) were employed, resulting in a Michelson contrast of 0.35, while frame, mask, fixation point and feedback were still presented with 99 cd/m^2. Also, the target stimuli were presented with a slightly smaller size of 0.41° x 0.31° visual angle.

Results

As in Experiment 1, practice blocks and trials with no recorded response (0.09%) were discarded from data analyses, and a repeated-measures analysis of variance (ANOVA) was conducted with the factors Sound Condition (longer, same, shorter, and no sound), and ISI (0, 7, 13, 20, 27, and 33 ms) for the mean percentage of correct responses. Again, whenever appropriate, *p* values were adjusted for violations of the sphericity assumption using the Greenhouse-Geisser correction.

In Experiment 2, mean percentage of correct responses for the final sample was 62.11 %, ranging from 32.85 % to 82.08 % for individual participants. Letter identification performance again improved with increasing ISI, $F(5, 100) = 233.04$, $p <. 001$, $\eta^2_p = .92$, with percentages monotonically increasing from 24.11% (*SD* = 9.13%) at the shortest ISI to 88.10% (*SD* = 10.78%) at the longest ISI. Moreover, Sound Condition showed a significant main effect, $F(3, 60) = 7.00$, $p < .001$, $\eta^2_p = .26$. Numerically, participants achieved 62.88 (*SD* = 12.97), 63.31 (*SD* = 12.64), 62.07 (*SD* = 13.73), and 60.17 (*SD* = 13.54) % of correct responses in the longer, same, shorter, and no-sound conditions, respectively. Finally, the interaction of Sound Condition and ISI was not significant, $F(15, 300) = 1.04$, $p = .42$, $\eta^2_p = .05$.

Additional ANOVAS were conducted to investigate the main effect of Sound Condition in more detail. First, an ANOVA comparing the no-sound vs. same condition revealed that the no-sound condition differed significantly from the same condition: $F(1, 20) = 19.47$, $p < .001$, $\eta^2_p = .49$. Second, an ANOVA including all bimodal conditions showed no effect of the sound duration, *F*(2, 40) = 1.64, $p = .21$, $\eta^2_p = .08$. Consistent with the omnibus ANOVA above, in none of these analyses, the interaction of sound condition with ISI was significant (all *p*s > .22).

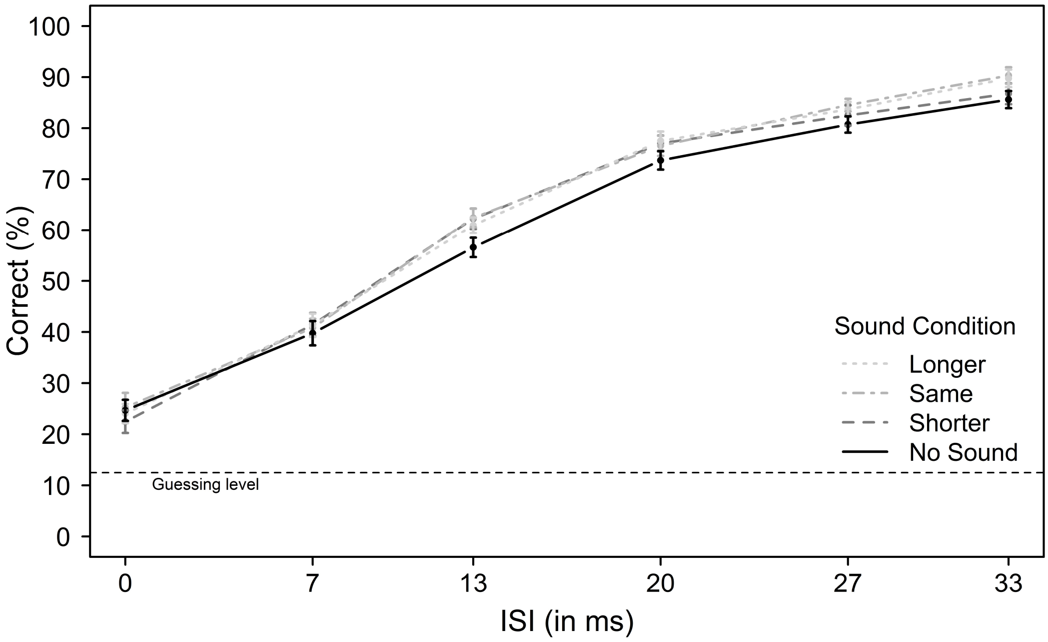

Figure 3. Mean percentage of correct responses in Experiment 2 as a function of Sound Condition and ISI. Error bars represent ±1 within-subject standard error of the mean with a correction for between-subjects variability (Morey, 2008).

Discussion

Overall, performance in Experiment 2 (*M* = 62.11%) was clearly impaired in comparison to Experiment 1 (*M* = 76.79%), demonstrating that, as expected, presenting the visual target letters at a lower contrast hampered letter identification. Most importantly, in this experiment, the presentation of any sound enhanced letter identification performance across a broad range of ISI durations in comparison with the unimodal condition. This result indicates that concurrent sounds facilitate letter identification, at least when the target is presented at a comparably low contrast and therefore, the target provides a relatively weak visual signal, resulting in performance well below ceiling. This interpretation is consistent with the principle of “inverse effectiveness” for multisensory integration that has been already demonstrated employing low level tasks as contrast detection (Noesselt et al., 2010; Senkowski et al., 2011).

However, despite the clear sound-induced benefit on letter identification, performance in this task again was not modulated by the *duration* of the concurrent sounds. This contrasts with the evidence in favour of such a duration-dependent improvement for low-level perceptual tasks as detection of Gabor patches (de Haas et al., 2013). Therefore, it seems that at least for higher-level perceptual tasks as letter discrimination, sound duration does not modulate the duration of the internal representation of the target letters. As mentioned in Introduction, many previous studies have already shown that the duration of concurrent sounds can bias the perceived duration of a visual stimulus (e.g., Bausenhart et al., 2014; De la Rosa & Bausenhart, 2013; Klink et al., 2011; Walker & Scott, 1981). The present experiment suggests that this, however, does not necessarily change the duration for which the internal perceptual representation is available to extract and combine complex feature information, as required for letter identification. This notion will be more directly investigated in Experiment 5.

Taken together, Experiments 1 and 2 provide mixed evidence concerning the effects of concurrent sound presentation on identification of masked letters. A tentative conclusion might be that such a sound-induced benefit indeed can be observed, as long as the visual signal is comparably weak, and thus does not enable asymptotic performance levels (i.e., a ceiling effect) by itself. To consolidate our findings, we conducted a third experiment which employed a within-subject manipulation of letter contrast.

Experiment 3

Previously, we found that the presentation of concurrent sounds clearly facilitated identification of letters presented at a low contrast (Experiment 2), but not (or at most marginally) at a high contrast (Experiment 1), in which overall performance was rather high and thus a ceiling effect might have prevented the occurrence of a sound-induced benefit. In Experiment 3, we aimed at providing further evidence regarding the effects of sound on the identification letters under different contrast conditions. Thus, the same masked letter identification task was used as in the previous experiments, but the letter was now randomly presented either at a low or high contrast letter against a uniform dark background. In contrast to the previous experiments, only two sound conditions were employed: either the target stimulus was presented unimodally (no sound) or a synchronous sound accompanied the target (sound present).

Method

Participants. Forty-eight participants (34 female; age range: 18-33 years; mean age: 22.42 years) from the University of Tübingen voluntarily took place in Experiment 3 in exchange of either course credit or money (8€/hour). Four participants with an average percentage of correct responses below 25% were discarded from data analyses. Thus, 44 participants remained for data analyses.

Apparatus, Stimuli, and Procedure. Apparatus, stimuli und procedure were identical to those employed in Experiment 2 except for the luminance of target and background and the sound duration. In Experiment 3, the target stimulus was presented with a luminance of 90 cd/m^2 (high contrast) or 40 cd/m^2 (low contrast) on a background of 6 cd/m^2, resulting in a Michelson contrast of 0.88 and 0.74, respectively. Moreover, only two different sound conditions were employed: sound and no sound. In the no-sound condition, the target stimulus was presented unimodally for 20 ms. In the sound condition, a 20-ms sound was presented synchronously with the target stimulus. The combination of six ISI durations, two sound conditions and two contrast levels resulted in 24 different experimental conditions. Each of these conditions was repeated 40 times, and all of the resulting 960 experimental trials were presented randomly intermixed and subdivided into blocks of 40 trials each. At the beginning of the experiment, participants performed 20 additional practice trials picked at random from all experimental trials.

Results

The practice block and trials without responses (0.17 %) were discarded from data analyses. A repeated-measures analysis of variance (ANOVA) was conducted with the factors Sound Condition (sound and no sound), Contrast (high and low contrast), and ISI (0, 7, 13, 20, 27, 33 ms) for the mean percentage of correct responses. For the following analyses, when appropriate, *p* values were adjusted for violations of the sphericity assumption using the Greenhouse-Geisser correction.

In Experiment 3, mean percentage of correct responses for the final sample was 63.38 %, ranging from 38.96 % to 81.77 % for individual participants. As expected, participants responded more accurately to letters presented at high contrast (*M* = 73.35%, *SD* = 11.40%) than to letter presented at low contrast (*M* = 53.42%, *SD* = 11.91%), *F*(1, 43) = 730.20, $p < .001$, $\eta^2_p = .94$ (cf. Fig. 4). Letter identification accuracy increased monotonically with increasing ISI, *F*(5, 215) = 784.67, *p* <. 001, $\eta^2_p = .95$, from 19.12% (*SD* = 6.79%) at the shortest ISI to 91.57% (*SD* = 9.72%) at the longest ISI. The interaction of Contrast and ISI was significant, *F*(5, 215) = 46.02, $p < .001$, $\eta^2_p = .52$, reflecting that the largest effects of contrast emerged at intermediate ISIs (12.04, 30.91, 37.32, 21.86, 11.35, and 6.12% difference between the high and the low contrast condition at the 0, 7, 13, 20, 27, and 33 ms ISIs, respectively).

Most importantly, performance was also influenced by Sound Condition, *F*(1, 43) = 10.41, $p < .01$, $\eta^2_p = .19$, with higher accuracy in the sound condition (*M* = 63.94%, *SD* = 11.34%) than in the no-sound condition (*M* = 62.83%, *SD* = 11.58%). Even though the sound-induced benefit in the high contrast condition was numerically smaller (0.75%) than in the low-contrast condition (1.47%), there was no significant interaction of Sound Condition and Contrast, *F*(1, 43) = 1.28, $p = .26$, $\eta^2_p = .03$. Also, none of the other interactions involving Sound Condition was significant (Sound Condition × ISI: *F*(5, 215) = 1.64, $p = .15$, $\eta^2_p = .04$; Sound Condition × ISI × Contrast: *F*(5, 215) = 0.63, $p = .65$, $\eta^2_p = .01$).

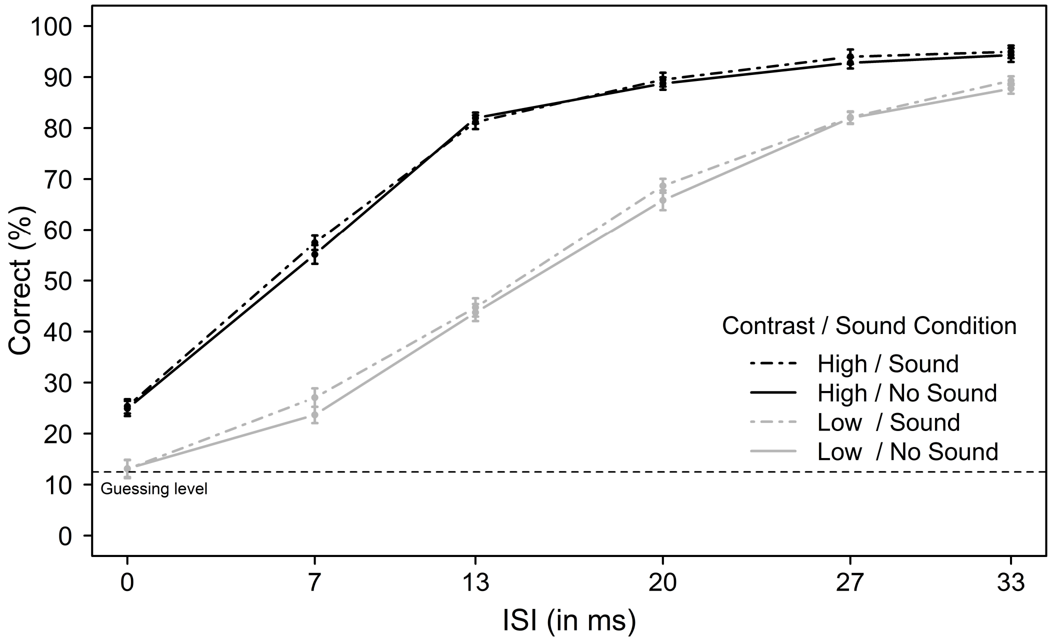

Figure 4. Mean percentage of correct responses in Experiment 3 as a function of Contrast, Sound Condition and ISI. Error bars represent ±1 within-subject standard error of the mean with a correction for between-subjects variability (Morey, 2008).

Discussion

In Experiment 3, the within-subjects-variation of stimulus contrast produced, as expected, a pronounced difference in letter-identification performance, with high target-to-background contrast yielding better performance than low target-to-background contrast. Numerically, the performance level for the high contrast level (around 73%) was quite comparable to the one obtained in Experiment 1 (77%), while the performance level achieved for the low-contrast letters was even slightly lower (53%) than in Experiment 2 (62%).

Therefore, if the explanation of our previous results in terms of inverse effectiveness holds, one should expect a sound-induced benefit on masked target letter identification for the low-contrast condition, but not, or to a reduced extent, for the high contrast condition. This is not clearly supported by our results, however. Statistically, the presence of a sound-induced benefit confirmed the results of

Experiment 2, but even though this benefit was numerically about twice as large in the low- than in the high contrast condition, the corresponding interaction effect (Sound Condition × Contrast) failed to reach significance. Therefore, in this within-subjects experiment and despite the increased sample size compared to Experiments 1 and 2, the sound-induced benefit on letter identification was not reliably reduced under conditions of high target contrast and consequently, high performance levels. However, the lack of interaction should also be viewed in the light of the overall rather small sound-induced benefit on letter identification in Experiment 3 (on average around 1.1% improvement), and thus might have failed significance due to a lack of experimental power. Unfortunately, our results therefore remain somewhat inconclusive regarding the role of target contrast and thus, the principle of inverse effectiveness.

Yet, returning to the main research question of the present study, Experiments 2 and 3 provide quite clear evidence for a small, but reliable, sound-induced benefit on letter discrimination with centrally presented letters followed by a noise-pattern mask. Interestingly, even though the ISI manipulation produced a large span of performance levels from near-chance performance to about 90% accuracy in letter identification, so far, none of the previous experiments showed a reliable variation of the sound-induced benefit with ISI duration, as would be expected based on previous results (Y. C. Chen & Spence, 2011; Takeshima & Gyoba, 2014). Therefore, in the next experiment, we aimed at further broadening the database regarding the sound-induced facilitation effect by employing two even lower contrast levels for the target letter, again presented across varying ISIs.

Experiment 4

So far, with the exception of Experiment 1, the present study provides some evidence for a sound-induced benefit on masked letter identification. However, this effect is rather small compared to the sound-induced benefit observed in previous studies (Y. C. Chen & Spence, 2011; Takeshima & Gyoba, 2014). So far, we observed the largest sound-induced benefit (~ 3 % improvement) in Experiment 2, in which the target stimuli were presented with a consistently low contrast. Therefore, in Experiment 4, we aimed at further corroborating the evidence for sound-induced facilitation in the current masking task by testing two conditions with a rather low contrast, and by keeping contrast constant across subsequent trials (as in Experiment 1 and 2).

Furthermore, the data from the previous experiments did neither show any clear relation between the sound-induced facilitation effect and contrast, nor between the sound-induced facilitation effect and ISI, even though especially an interaction with ISI (i.e., largest facilitation for intermediate ISIs) would be expected based on the existing evidence from previous studies (Y. C. Chen & Spence, 2011; Takeshima & Gyoba, 2014). As the data from the previous experiments showed, there was an especially large interindividual variation in overall accuracy for the task employed in our study. Unfortunately, such interindividual differences might complicate data analysis and the respectively drawn conclusions, especially regarding the expected interaction between ISI and sound condition. Specifically, if the sound-induced facilitation effect is not associated with a specific ISI duration, but rather is most pronounced at a certain level of performance (e.g., for the specific task employed by Chen and Spence, 2011, cf. their Figure 1, one might suggest around 60-70% percent of accurate responses), large interindividual differences in the level of accuracy might, in turn, produce the largest sound-induced facilitation effects at very different ISI durations. Consequently, on average, the sound-induced facilitation effect would “smear” across the whole ISI-masking function. This then might result in the observed statistical invariance of the sound-induced facilitation effect across ISIs. Therefore, in Experiment 4, we a priori planned an additional analysis of the sound-induced facilitation effect to explore its relation to individual performance levels, rather than to fixed ISIs and contrast levels.[3]

---

[3] We thank an anonymous reviewer of a previous version of this manuscript (including Experiments 1-3) for pointing us to the role of interindividual variability and possible alternative forms of data analysis, such as regression/trend analysis.

Method

Participants. Forty-eight participants from the University of Tübingen voluntarily took place in Experiment 4 in exchange of either course credits or money (8€/hour). Seven participants with a percentage of correct responses below 25% in the low-contrast condition were discarded from data analyses, and replaced by additionally tested participants in order to maintain an adequate sample size. The final sample consisted of 45 women and 3 men; with age ranging between 18-35 years (mean age: 21.67 years, *SD* = 3.26).

Apparatus, Stimuli, and Procedure. This experiment was similar to Experiment 3 except for the following changes: Due to technical reasons, all stimuli were displayed on a Samsung SyncMaster 1100 MB monitor running at 150 Hz, but stimulus sizes were identical as in Experiment 3. The target stimulus was presented with a luminance of 70 cd/m^2 (high contrast) or 55 cd/m^2 (low contrast) on a background of 25 cd/m^2, resulting in Michelson contrasts of 0.47 and 0.38, respectively. Frame, mask, fixation point and feedback were presented at 90 cd/m^2. Unlike in Experiment 3, the two target contrast conditions were presented in separate blocks of the experiment (and thus, as in Experiments 1 and 2, target contrast was constant across subsequent trials). Each block consisted of 480 trials (6 ISI durations × 2 sound conditions × 8 target letters × 5 repetitions). The order of the two blocks was counterbalanced across participants, and each block was preceded by 20 practice trials picked at random from all trials of the subsequent block. Again, a self-terminated break and blockwise feedback was administered after each 40 trials of the experimental blocks.

Results

Practice blocks and trials without recorded responses (0.04%) were discarded from data analyses. Two main analyses were conducted in this experiment. First, as in Experiment 3, a repeated-measures analysis of variance (ANOVA) was conducted with the factors Sound Condition (sound and no sound), Contrast (high and low contrast), and ISI (0, 7, 13, 20, 27, 33 ms) for the mean percentage of correct responses. For this analysis, when appropriate, *p* values were adjusted for violations of the sphericity assumption using the Greenhouse-Geisser correction. Second, we conducted a regression analysis of the magnitude of the sound effect depending on the underlying baseline performance (for details, see below).

ANOVA analysis. On average, mean percentage of correct responses for the final sample was 58.63 %, ranging from 35.62 % to 82.60 % for individual participants. Again, participants´ mean percentage of correct responses was higher for letters presented at high contrast (*M* = 66.14%, *SD* = 11.83%) than for letter presented at low contrast (*M* = 51.11%, *SD* = 12.03%), *F*(1, 47) = 294.79, $p < .001$, $\eta^2_p = .86$ (cf. Fig. 5). Letter identification accuracy increased monotonically with increasing ISI, *F*(5, 235) = 869.49, $p <. 001$, $\eta^2_p = .95$, from 16.69% (*SD* = 7.71%) at the shortest ISI to 90.74% (*SD* = 8.64%) at the longest ISI. As in Experiment 3, the interaction of Contrast and ISI was significant, *F*(5, 235) = 29.40, $p < .001$, $\eta^2_p = .38$, reflecting that the largest effects of contrast emerged at intermediate ISIs (5.74, 20.91, 25.02, 19.46, 11.48, and 7.59 % difference between the high and the low contrast condition at the 0, 7, 13, 20, 27, and 33 ms ISIs, respectively). Most importantly, performance was again influenced by Sound Condition, *F*(1, 47) = 15.61, $p < .001$, $\eta^2_p = .25$, with higher accuracy in the sound condition (*M* = 59.38%, *SD* = 11.67%) than in the no-sound condition (*M* = 57.87%, *SD* = 11.57%). Again, none of the interactions involving Sound Condition was significant (all *Fs* < 1).

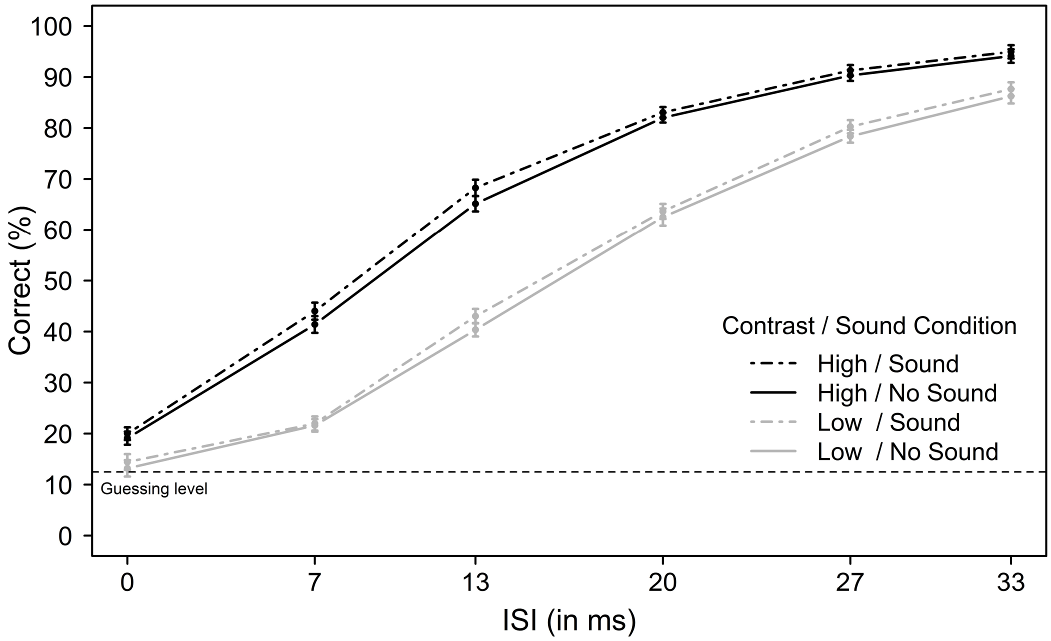

Figure 5. Mean percentage of correct responses in Experiment 4 as a function of Contrast, Sound Condition and ISI. Error bars represent ±1 within-subject standard error of the mean with a correction for between-subjects variability (Morey, 2008).

Regression analysis. For this analysis, we first computed the magnitude of the sound-induced facilitation effect (% correct $_{\text{Sound}}$ - % correct $_{\text{No sound}}$) separately for each participant and combination of ISI and Contrast condition. These data, related to baseline performance (i.e., % correct $_{\text{No sound}}$) are depicted in Figure 6. Several aspects are evident from these figures: Clearly, there is a large overall variability in the magnitude of the sound effect across performance levels. ISI duration is encoded in greyscale, with longer ISIs corresponding to darker colors. The effect of ISI thus gets evident in increasing number of dark colors towards higher performance levels along the x-axis. For convenience, contrast levels are depicted in separate panels. The overall effect of contrast on performance gets especially evident in the relatively higher point density at the left-hand side of the x-axis for low- contrast than for high-contrast trials, and vice versa at the right-hand side. In both figures, a LOESS (locally weighted scatterplot smoothing, Cleveland, 1979) line fitted to the respective data indicates a non-linear relation between baseline correctness and the magnitude of the sound effect. The sound effect seems generally present for small performance levels at or around chance performance followed by a decrease of the sound effect with increasing baseline performance[4]. This initial decrease cannot be unequivocally attributed to an influence of sound, however, since it might be an artifact caused by the low performance levels in the baseline condition (e.g., the sound effect can, by definition, only be zero or positive, if the baseline performance is zero). More notably, however, the LOESS fit indicates a second increase with a peak in the magnitude of the sound effects around a performance level of around 50-60 % correct responses. Despite overall differences in performance level, this curvilinear trend is quite similar for the two different target contrasts. Thus, it seems that the sound-induced facilitation effect observed in the present experiment indeed depends to some extent on the specific level of baseline performance, being especially pronounced at intermediate performance levels, rather than depending on a specific target contrast level or on specific ISI durations.

To investigate this observed curvilinear relationship statistically, we performed a linear regression analysis with the sound effect as criterion and the baseline correctness as predictor. In fact, there were small but significant linear ($\beta = -1.38$, $SE = 0.12$, $t(572) = 11.84$, $p < .001$), quadratic ($\beta = 0.03$, $SE = 0.003$, $t(572) = 11.06$, $p < .001$), and cubic ($\beta = -0.0002$, $SE = 0.00002$, $t(572) = 10.55$, $p < .001$) trends, $F(3, 572) = 58.4$, $p < .001$, thus corroborating the observed curvilinear relationship between performance level and sound effect magnitude described above. It must be acknowledged that this relationship, however, only accounts for a small proportion of variability in the sound effect, as indicated by an R^2 of .23.

---

[4] Note that similar results were obtained in an alternative analysis where data points indicating below chance performance, presumably due to sampling noise, were fixed to the chance performance level of 12.5 %.

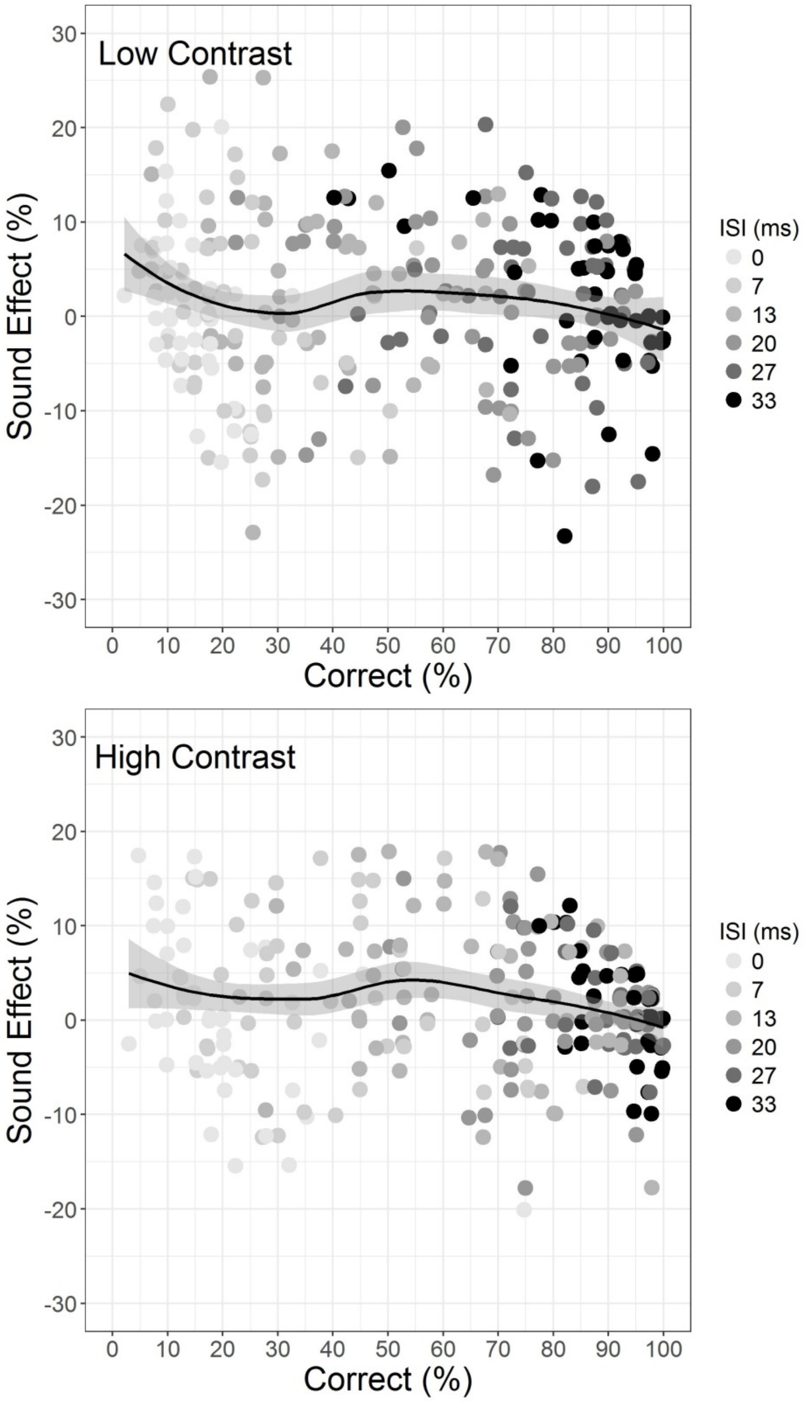

Figure 6. Sound effects in Experiment 4 for each combination of participant, ISI and contrast, plotted against baseline performance (i.e., % correct in the no-sound condition). The upper panel depicts data from the low contrast condition, the lower panel from the high contrast condition. The continuous lines depict the corresponding LOESS (locally weighted scatterplot smoothing, Cleveland, 1979) fits along with their confidence intervals. Data points are slightly jittered to give a better impression of their distribution in case of overlap.

Discussion

As in our previous experiments, the manipulation of contrast and ISI produced also in Experiment 4 pronounced modulations of letter identification performance, with performance improving with prolonging ISI and with higher contrast. Most importantly, we again observed a sound-induced facilitation of letter identification. As in Experiment 3, no direct evidence for the principle of inverse effectiveness was observed in this experiment, since again, this facilitation effect was not modulated reliably by target contrast. Likewise, ISI duration did also not modulate the sound-induced benefit. Therefore, following the reasoning from Chen and Spence (2011), one might attribute the presently observed sound effect to mere preparedness enhancement rather than to signal enhancement or object enhancement. Additional pre-planned analyses taking interindividual variability in accuracy into account, however, redeem the object-enhancement account to some extent. Hence, these analyses provide some evidence that the sound-induced facilitation effect is especially pronounced for targets which, for whatever reasons, are processed with intermediate (around 50-60 % of correct responses) accuracy. We will return to this issue in the General Discussion.

Experiment 5

In this experiment, we finally return to the second original question of the present study: whether sound duration modulates the sound-induced benefit. In summary, so far, a reliable sound-induced benefit was observed in every experiment except Experiment 1. In Experiment 2, the manipulation of the duration of the sounds, however, did not produce any additional modulation of accuracy. The interpretation of this result is, of course, very speculative, since there are several possibilities to account for such a null effect: First, in principle, it might be that the sounds were not processed at all and therefore varying their duration did not affect performance. Second, it might be that the sounds were processed and their presence affected target identification performance, but the manipulation of sound duration was not pronounced enough to alter the duration of the multimodal percept. In this case, manipulation of sound duration should not even affect perceived duration of the visual targets. Third, it might be that the sounds were processed and the manipulation of their duration indeed effectively altered perception of the multimodal compound stimulus, but that this did not have any beneficial effects on the identification of the target letter (i.e., the sound duration might have affected the perceived duration of the multimodal compound stimuli, but this did not at the same time alter the persistence or stability of the visual object representation containing the features enabling identification of letter identity).

Clearly, the first possibility can be discarded due to the presence of the overall sound-induced facilitation effect. Distinguishing between the second and third possibility is more difficult. On the one hand, there is ample existing empirical evidence that sound duration indeed affects perceived visual duration (Asaoka & Gyoba, 2016; Bausenhart et al., 2014; K. M. Chen & Yeh, 2009; de Haas et al., 2013; De la Rosa & Bausenhart, 2013; Hartcher-O'Brien et al., 2014; Klink et al., 2011). However, all these studies employed stimuli lasting at least hundreds of milliseconds. Thus, it remains unclear whether a similar multimodal bias of perceived duration would be also observed for especially brief stimuli as employed in the present study. Therefore, in Experiment 5, we tackle this question by requiring participants to perform two different tasks on the same stimuli. Specifically, participants were again presented with masked letters accompanied by sounds of varying duration or presented unimodally. As in our previous experiments, in Experiment 5a, participants were asked to identify visual letter identity. In Experiment 5b, however, they were asked to identify the duration of the visual letters.

In both experiments, we also varied visual target duration in order to verify how a physical increase in visual target duration would affect identification performance and duration perception. Importantly, it has been suggested that not only the target duration, but the stimulus-termination-asynchrony (STA) between target and mask determines the efficiency and time course of the masking function (Macknik & Livingstone, 1998). Accordingly, if, for example, a "longer" sound would increase persistence of the target representation by means of a temporal ventriloquism effect (Bertelson & Aschersleben, 2003; Morein-Zamir et al., 2003), as outlined in the Introduction , this might correspond to a decrease in the STA, and thus even hamper target identification. Therefore, the variation of the physical duration of the visual target was crucial in order to establish the effects of target persistence on letter discrimination accuracy within our present masking paradigm.

Method

Participants. Forty-eight participants from the University of Tübingen voluntarily took place in two subsequent Experiments (5a and 5b) in exchange of either course credits or money (8€/hour). Based on the procedure employed in the previous experiments, we adopted as exclusion criterion a value of less than 25 % of correct responses for the 20-ms visual target letters. According to this procedure, six participants of the original sample were replaced by additionally tested participants. [5] The final sample consisted of 36 women and 12 men; with age ranging between 18-37 years (mean age: 24.04 years, *SD* = 4.33).

Stimuli and Apparatus. The apparatus was identical to the one employed in Experiment 4. Target and background luminance were chosen according to the high contrast condition of Experiment 4, that is, light grey targets were presented with 70 cd/m² on 25 cd/m² dark grey background.

Procedure. In *Experiment 5a,* the time course of the trials was similar to Experiment 4, with the following exceptions. First, now also visual target duration varied randomly from trial to trial (7, 20, 33 ms), while the onset asynchrony between target letter and mask was constantly 33 ms (thus, for 20-ms targets, this corresponds to the 13-ms ISI condition of Experiment 4). Second, the visual targets were either presented alone (no-sound condition) or orthogonally paired with auditory intervals of different duration (7, 20, 33 ms). As in Experiment 4, the participants' task was to report the identity of the target letters by a pressing the corresponding key on the keyboard.

Each combination of 3 Target Durations, 4 Sound Conditions and 8 Target Letters was presented for 5 times, resulting in 480 experimental trials, presented in random order. These were subdivided into 12 blocks of 40 trials, with feedback about the percentage of correct responses achieved in the preceding block. As in the previous experiments, one practice block of 20 trials preceded the experiment.

---

[5] It should be noted that additional five participants aborted the experiment after practice of Experiment 5a, because they did not achieve over-chance performance and/or reported not "seeing" the presented letters, and thus felt uncomfortable continuing the task.

Participants were allowed to repeat this practice block in case of low performance levels or whenever they felt they needed more practice.

The Procedure in *Experiment 5b* was identical to the one in Experiment 5a except for the participants' task: Rather than reporting target identity, they were required to report the relative duration of the target letter in each trial. Specifically, participants were informed that the target letter would be presented for one of three different durations ("short", "medium", or "long"), and they should indicate which of these target durations they had perceived by pressing "1", "2", or "3" on the numerical keypad of the keyboard, respectively.

All participants performed Experiment 5a prior to Experiment 5b, in order to avoid drawing attention to the duration manipulation of the visual targets and sounds in Experiment 5a. As in the previous Experiments, participants were again informed that the interspersed sounds were completely task-irrelevant.

Results

Experiment 5a

Practice blocks and trials with no recorded response (0.29 %) were discarded from data analysis. A repeated-measures analysis of variance (ANOVA) was conducted with the factors Sound Condition (no sound, 7, 20, and 33 ms) and Target Duration (7, 20, and 33 ms) for the mean percentage of correct responses, computed separately for each participant and condition. For the following analyses, when appropriate, *p* values were adjusted for violations of the sphericity assumption using the Greenhouse-Geisser correction.

Averaged across all conditions, participants´ mean percentage of correct responses was 49.23 (*SD* = 24.17 %), with a wide performance range from 25.21 to 81.04 % between participants. The ANOVA on correct responses showed that accuracy increased with visual target duration, $F(2, 94) = 376.27$, $p < .001$, $\eta^2_p = .89$. Specifically, participants achieved on average 26.28 (*SD* = 12.66), 57.35 (*SD* = 19.32), and 64.44 (*SD* = 18.30) % of correct responses for target durations of 7, 20, and 33, ms respectively. The main effect of Sound Condition was not significant, $F(3, 141) = 1.50, p = .216, \eta^2_p = .03$, but there was an interaction of Sound Condition and Target Duration, $F(6, 282) = 2.48, p = .024, \eta^2_p = .05$ (cf. Figure 7, left panel).

To investigate this interaction in more detail, separate ANOVAs with the factor Sound Condition for each level of Target Duration revealed an effect of sound duration only for the longest (33 ms) visual target, $F(3, 141) = 4.11, p = .011, \eta^2_p = .08$, but not for the shorter ones (all $p > .24$). To evaluate the effect of *sound presence* for the 33-ms target, a paired-samples *t*-test was conducted to compare accuracy between the no-sound and the 33-ms sound condition (i.e., the "same"-duration sound). Participants achieved a higher percentage of correct responses for 33-ms targets accompanied by sounds of the same duration ($M = 65.63$, $SD = 18.09$) than for 33-ms targets presented unimodally ($M = 61.70$, $SD = 19.27$), $t(47) = 3.30, p = .002$. To evaluate the effects of *sound duration* for the 33-ms target, a repeated-measures ANOVA on accuracy in the three bimodal conditions was conducted. This analysis yielded no significant difference in the percentage of correct responses depending on sound duration, $F(2, 94) = 0.87, p = .41, \eta^2_p = .02$.

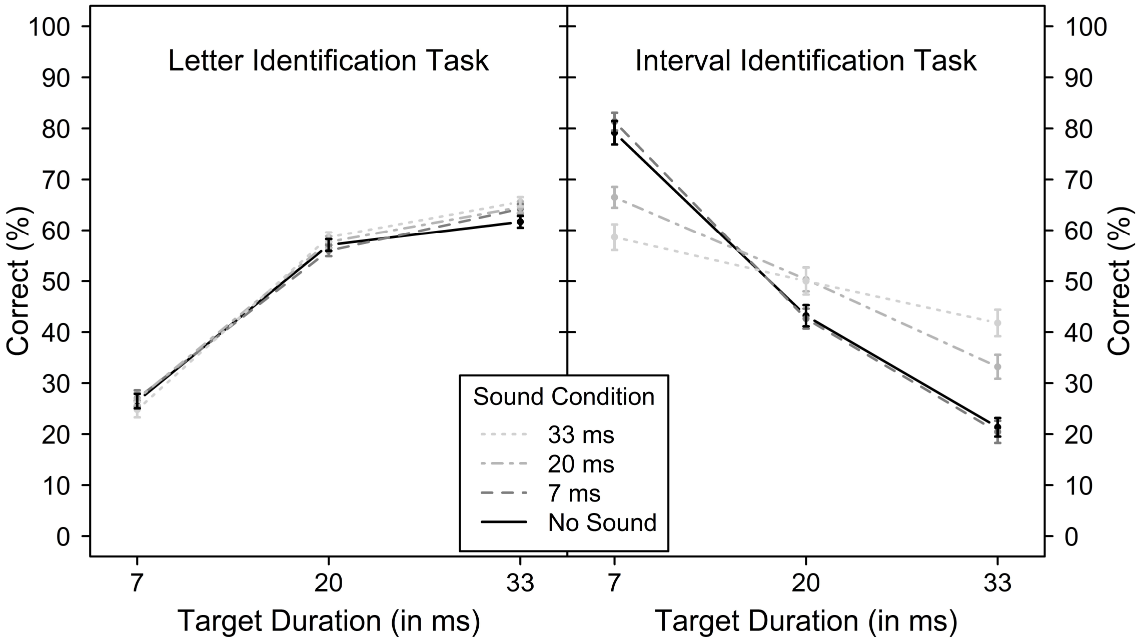

Figure 7. Mean percentage of correct responses in Experiment 5a (letter identification task) and Experiment 5b (interval identification task) as a function of Target Duration and Sound Condition. Error bars represent ±1 within-subject standard error of the mean with a correction for between-subjects variability (Morey, 2008).

Experiment 5b

Practice blocks and trials with no recorded response (0.31 %) were discarded from data analysis. A repeated-measures analysis of variance (ANOVA) was conducted with the factors Sound Condition (no sound, 7, 20, and 33 ms) and Target Duration (7, 20, and 33 ms) for the mean percentage of correct responses in the duration identification task, computed separately for each participant and condition. For the following analyses, when appropriate, *p* values were adjusted for violations of the sphericity assumption using the Greenhouse-Geisser correction.

The ANOVA yielded a main effect of target duration, $F(2,94) = 167.06$, $p < .001$, $\eta^2_p = .78$, reflecting that participants were most accurate in identifying the shortest target duration (M = 71.41%, SD = 11.23), followed by the medium (M = 46.55%, SD = 11.49) and the long duration (M = 29.20%, SD = 14.06). Sound Condition alone did not influence accuracy, $F(3, 141) = 1.19$, $p = .32$, $\eta^2_p = .02$. Most importantly, however, there was a pronounced interaction of target duration and sound condition, $F(6,282) = 30.79$, $p < .001$, $\eta^2_p = .40$. Specifically, for the 7-ms target, participants' achieved the highest duration identification accuracy when it was presented unimodally (M = 79.16%, SD = 17.40) or paired with a 7-ms sound (M = 81.33, SD = 11.23), intermediate accuracy when it was paired with a 20-ms sound (M = 66.50%, SD = 14.54), and lowest accuracy when it was paired with a 33-ms sound (M = 58.64%, SD = 19.31). For the 33-ms target, this pattern reversed: participants' achieved the highest duration identification accuracy when the 33-ms target was presented with a 33-ms sound (M = 41.80%, SD = 20.47), intermediate accuracy when it was paired with a 20-ms sound (M = 33.20%, SD = 18.65), and lowest accuracy when it was paired with a 7-ms sound (M = 20.44%, SD = 16.45) or presented unimodally (M = 21.35%, SD = 15.57, cf. Figure 7, right panel)

Discussion

In Experiments 5a and 5b, several important findings could be established. Experiment 5a was designed to establish the effects of a physical duration manipulation of the visual target letter on identification performance, while reassessing the potential effects of sound presence and sound duration on this measure. The results show that a physical manipulation of target duration in the present masked letter identification paradigm indeed leads to a pronounced change in letter identification accuracy. On average, participants achieved 38% more correct responses when target duration was increased from 7 ms to 33 ms.

A sound-induced facilitation of letter identification was also observed, however, unexpectedly, only for visual targets of 33 ms duration. Identification of unimodally presented letters was worse than identification of the same letters accompanied by sounds of congruent duration. At the same time, varying the duration of the sounds did not modulate performance levels. Therefore, and consistent with the results of Experiment 2, the presence of sound, but not the duration of the sound, altered identification of the visual target letters.

We expected a similar effect to occur at least also for the 20-ms visual target, since this condition attempted to replicate the high-contrast/ 13-ms ISI condition of Experiment 4. We can only speculate that the different results may be due to contextual changes introduced from Experiment 4 to Experiment 5a. For example, in Experiment 4, target duration was consistently 20 ms, and sound duration in bimodal trials was always congruent with visual target duration. In Experiment 5, different target durations and sound durations were intermixed, and some combinations of visual and auditory duration now differed quite substantially (e.g. 7-ms visual targets were sometimes presented with 33 ms sounds and vice versa, resulting in discrepancies of up to 26 ms). When comparing the corresponding conditions of Figure 4 (13 ms ISI, high contrast) and Figure 5 (20 ms visual target), it gets evident that the random intermixing of different target durations and auditory durations in Experiment 5 was quite detrimental to performance: participants achieved approximately 10 % less correct responses than in Experiment 4 despite these trials were identical except for their experimental context. One crucial factor in this result might be the inclusion of the visual 7-ms target, which produced very low identification performance, and might have led to a high experienced task difficulty (which is also evident in the relatively high drop-out rate of participants), which in turn may have affected motivation and thus performance also for the other conditions of stimulation.

Clearly, more research is needed to investigate how trial-by-trial variation in target duration, sound duration, and ISI duration, affects overall performance and multisensory benefits in the case of masked letter identification. Regarding the effect of multisensory stimulation on perceived duration, for example, it is quite well established that the congruency between the inputs of different modalities may affect the efficiency of integration (Bausenhart et al., 2014; Klink et al., 2011; Sarmiento, Shore, Milliken, & Sanabria, 2012). It is therefore conceivable that similar congruency effects may be observed concerning the identification of letters accompanied by uninformative sounds, as in the present study.

Nonetheless, in the present study, a sound-induced benefit was reliably observed at least for the longest visual target duration. Still, the duration of the sounds did not affect identification performance. Relating this finding to the pronounced effect that a *physical* manipulation of visual duration exerts on performance, is seems likely that the manipulation of sound duration did not alter the duration of the internal representation containing the features defining the target letter. To recapitulate, there might be several explanations for a null effect of sound duration: a) the sounds might not have been processed, or not integrated with the visual target. We can refute this finding based on the effects of sound presence on identification performance described above. b) The sounds may have been processed and integrated with the visual targets, but the manipulation of sound duration was not strong enough to elicit any effect on the duration of multimodal percept – and consequently also not on letter identification. c) In principle, the manipulation of sound duration was strong enough to affect judgments regarding the duration of the visual stimulus, but this does not change the processing of information about letter identity.

Experiment 5b was designed to shed some light on the potential explanations b) and c). The results are clear-cut: If participants are asked to identify the visual target duration instead of letter identity, the sound duration modulated performance in a very pronounced manner. 7-ms visual targets accompanied by sounds of longer duration were substantially often misidentified as being of longer duration, and 33-ms targets accompanied by shorter durations as being of shorter duration. Thus, we extended the results of previous studies on the auditory influence on perceived visual duration (Asaoka & Gyoba, 2016; Bausenhart et al., 2014; K. M. Chen & Yeh, 2009; de Haas et al., 2013; De la Rosa & Bausenhart, 2013; Hartcher-O'Brien et al., 2014; Klink et al., 2011), which typically employed much longer interval durations, and correspondingly, larger duration discrepancies between the modalities, to a range of much briefer visual and auditory durations (7 – 33 ms). Accordingly, it is evident that explanation b) from above does not hold – duration judgments for the visual targets were indeed substantially affected by the duration of the accompanying sounds. Yet, this did not elicit any changes in letter identification performance. It should be considered that it is possible that the effect of sound duration on visual duration judgments did not originate as a consequence of a genuine sensory integration, but rather resulted from a cue combination at later (i.e. decisional) stages (but see, Bausenhart et al., 2014; Hartcher-O'Brien et al., 2014; Klink et al., 2011, for relevant evidence against a purely decisional effect). Yet, in any case, this control experiment indicates that the lack of effect of sound duration on audiovisual letter identification cannot be attributed to the possibilities that sound durations were not processed, could not be discriminated, or that the discrepancy between visual and auditory duration was not large enough to affect judgments of visual duration.

Taken together, the results from Experiments 5a and 5b basically support and extend the results of Experiment 2, suggesting that sound presence may indeed facilitate letter identification performance, but that multimodally induced changes in perceived duration are not to be equated with altered persistence of the visual representation of the letter percept.

General Discussion

The two major aims of the present study were, to replicate the sound-induced facilitation of performance in a higher-level perceptual task as masked letter identification (Y. C. Chen & Spence, 2011) and to assess whether varying the duration of the concurrently presented sounds would modulate the sound-induced benefit in this task, similar to what has been observed recently for a low-level contrast detection task (de Haas et al., 2013).

In the present study, with exception of Experiment 1, we showed that the mere presence of a sound can indeed enhance letter identification performance compared to unimodal letter presentation, and thus we have replicated and extended this basic finding from Y. C. Chen & Spence (2011), to identification of centrally presented letters followed by random-dot noise masks. In contrast to these previous results however, this sound-induced facilitation of masked letter identification seemed quite invariant across a wide range of target-to-mask ISIs. In addition, it was also largely resistant to the manipulation of contrast (perhaps with the exception of very high contrast levels leading to high performance levels as in Experiment 1). A further difference is that the facilitation effects in the present study were generally smaller in magnitude than those observed by Y. C. Chen & Spence (2011). As an anonymous reviewer pointed out, spatial alignment of auditoy and visual stimuli may determine early audiovisual interactions (e.g., Frassinetti et al., 2002; Meredith & Stein, 1986; Meyer, Wuerger, Röhrbein, & Zetzsche, 2005) and thus our use of headphones for sound presentation, as opposed to speakers positioned above and below the monitor, might have compromised multisensory integration. However, spatial coincidence of audiovisual stimulation seems to be most important in tasks that require spatial judgements, but only of subordinate importance for tasks requiring temporal judgments or target identification (see also Experiment 5 in Y. C. Chen & Spence, 2011; for an overview, see Spence, 2013). The sound-induced enhancement of letter identification as well as the sound-induced changes in duration judgments observed in the present study add evidence to the notion that spatial alignment is not a crucial prerequisite for multisensory integration effects to occur.

Basically, three different mechanisms have been proposed for such sound-induced enhancements of visual perception (Y. C. Chen & Spence, 2011). First, according to a preparation enhancement mechanism, a nonspecific processing benefit should emerge through arousal evoked by the concurrent sound presentation. Such an enhancement should be present across a wide range of ISIs. Second, according to a signal-enhancement mechanism, the neural signals of visual and auditory stimuli would interact at an early perceptual stage of processing, resulting in a boost of the visual signal especially when this signal is rather weak (see also, Frassinetti et al., 2002; Meredith & Stein, 1983; Noesselt et al., 2010, 2008; Senkowski et al., 2011; Stein & Meredith, 1993). Thus, this account would predict sound-induced facilitation especially at low performance levels associated with short ISIs. Interestingly, pattern masking by noise masks, as employed in the present experiments, has been recently attributed to a masking-related increase in perceptual noise, resulting in a reduced signal-to-noise ratio for target processing (Agaoglu et al., 2015). Thus, in turn, such a masking-related performance impairment might be counteracted by a sound-induced enhancement of the target signal. Third, according to an object-enhancement account, the concurrent sound would facilitate letter identification at a later stage of perceptual processing, by fostering the consolidation of a perceptual object representing the visual target, whenever this representation can be perceptually segregated from the object representation of the mask. Consequently, the masking stimulus would less likely interfere with or substitute the perceptual representation of the target letter, thus reducing the masking effect on the target letter. This account was supported by the results of Y. C. Chen & Spence (2011), since sound-induced facilitation was only observed at intermediate ISIs but not very brief ones and basically independent of low-level factors as the relative luminance of target and mask. Given the different mask types employed in our vs. Chen and Spence's study (i.e., random dot masking vs. structure masking by letters), which typically lead to ISI-masking functions with different time courses and potentially different underlying forms of masking interference (e.g., Agaoglu et al., 2015), it is quite possible that different mechanisms (as preparedness enhancement, signal enhancement, or enhanced object consolidation) may have contributed to the effects of concurrent sound presentation in both studies.

Specifically, based on the invariance of the sound-induced facilitation effect across ISIs and contrasts observed in the present study, one may conclude that it originates from simple preparedness enhancement through the concurrent sounds. Thus, rather than a genuine multisensory interaction, one might conceive the sound effect as mediated by attentional enhancement. Specifically, the abrupt presentation of a sound might produce an alerting effect that would draw attention towards the time point of letter presentation, thus reducing temporal uncertainty. This would in turn facilitate letter identification. Indeed, many previous studies on the effects of warning and accessory signals have shown that visual or auditory warning signals presented before or in close temporal proximity with a target can enhance target processing at early perceptual stages (e.g., Bausenhart, Rolke, & Ulrich, 2007; Correa, Lupiáñez, & Tudela, 2005; Hackley & Valle-Inclán, 2003; Lange, Rösler, & Röder, 2003; Rolke, 2008; Rolke & Hofmann, 2007). Hence, it is possible that the concurrent sound facilitated target processing through an improved allocation of attentional resources rather than multisensory binding. However, these two possible mechanisms are not necessarily incompatible or mutually exclusive. For example, Y. C. Chen & Spence (2011) suggested a perceptual processing system underlying the sound-induced enhancement on letter identification that includes both audiovisual binding and attention at an object processing level. Specifically, when the presentation of the letter and sound coincide, both stimuli might first be bound to form a multisensory object representation (for an overview, see Alais, Newell, & Mamassian, 2010). This multisensory object representation is supposed to be especially salient and thus attracts attention, which in turn would facilitate visual target processing.

The interpretation of our results in terms of unspecific preparedness enhancement, however, may be reconsidered when interindividual differences are taken into account. As outlined and discussed in Experiment 4, we observed especially large interindividual differences in baseline performance in the employed backwards masking task. This in turn, might result in a smearing of the sound-induced benefit along the ISI-masking function and thus obliterate any evidence for accounts that predict more specific patterns of performance enhancement through sound, as signal- or object-enhancement. In fact, when the sound effect was explored depending on individual baseline accuracy rather than depending on fixed ISI or contrast levels, we observed some evidence for a nonlinear relationship of both measures. While the first peak in the accuracy-effect function at low baseline performance levels (which would imply a signal-enhancement account) cannot be unequivocally interpreted due to potential baseline effects, the second peak of this function might be theoretically more informative. Specifically, the sound-induced enhancement appears to be especially pronounced for stimulation conditions evoking intermediate performance levels. This would be consistent with an object enhancement account, considering that object formation might not follow a fixed time-course but rather depends on individual differences in processing speed. This alternative interpretation must be taken with care, certainly, since the observed relationship was not especially pronounced. Yet, we would like to suggest for future research that it might be informative to examine sound-induced benefits in terms of such accuracy-effect functions rather than depending on fixed experimental variations, especially in the presence of large interindividual performance differences. A second potential methodological implication is that sound-induced facilitation effects might be more reliably and efficiently investigated by employing adaptive testing schemes. Specifically, our analysis suggests that sound-induced facilitation would be most reliably observed if the specific parameters of stimulation, such as ISI or contrast, were adjusted individually for each participant, such that comparable performance levels can be obtained (see, e.g., de Haas et al., 2013; Noesselt et al., 2010, for similar procedures).

In any case, however, the observed sound-induced enhancement on masked letter identification again demonstrates auditory influences on visual perceptual processing, which result in enhanced letter identification performance through the mere presence of sound. Importantly, the employed sounds did not provide any information about the response dimension (i.e., letter identity). Therefore, an interpretation of the effect of sound presence in terms of later, purely decisional effects or biases (as responding to the information provided in the auditory modality) can be safely ruled out (cf. Y. C. Chen & Spence, 2011, p. 1785).

In contrast to the effect of sound presence, the duration of the sounds did not differentially affect letter identification in Experiments 1, 2, and 5, even though two important prerequisites for such an effect could be established in Experiment 5. First, manipulating the duration of the visual targets directly, by employing the same durations as for the sounds in the bimodal conditions, affected letter identification severely (Exp. 5a). Second, manipulating the duration of the sounds altered duration judgements for the visual targets severely (Exp. 5b). Thus, even when concurrent sound duration is processed and the multimodal discrepancy is large enough to affect judged visual duration, there is no evidence that this goes along with changes in the persistence of the internal target representation from which letter features are extracted and bound into more complex perceptual objects. This contrasts with a previous study which demonstrated a modulation of contrast detection through the duration of accompanying sounds (de Haas et al., 2013). Since our study is not a direct replication of this prior work, our results of course do not dispute these previous findings. Yet, one may speculate that the mechanisms underlying low-level perception such as contrast detection may be different from those involved in higher-level perceptual tasks as masked letter discrimination. Also, it is of course possible that an effect of sound duration on letter identification might emerge for different set of temporal parameters of stimulation, for example, if numerically longer durations or larger discrepancies between sound and letters were employed (e.g., in the study of de Haas et al., 2013, significant enhancement was observed for a sound 48 ms longer than the 24 ms visual target). Yet, severe discrepancies may likely also hamper the assumption of unity, which is often conceived as a prerequisite of multisensory integration (Y. C. Chen & Spence, 2017; Welch & Warren, 1980). Moreover, in the present masking paradigm, we avoided larger discrepancies between visual and sound duration, since these would inevitably result in temporal overlap of the sound not only with the target but also the masking stimulus. This, in turn might hinder perceptual segregation of the two stimuli and thus counteract any benefit introduced by the sound (cf., Y.-C. Chen & Spence, 2011). Nonetheless, given the pronounced dissociation between the effects of sound duration on identification of (a) letter duration and (b) letter identity, is safe to conclude that at least for the present conditions of stimulation, both types of judgment did not draw on a common (inseparable) and unified object representation (e.g., as if perceived duration and letter identity were both inferred from the clarity or saliency of the same multimodal perceptual object). In other words, duration and identity of a visual perceptual object presumably are separable aspects of perception, which can be differentially modulated by concurrent presentation of sounds of different duration.

In summary, we showed that the mere presence of a sound facilitated identification of noise-masked, centrally presented letters in comparison to unimodally presented letters. Manipulating sound duration did not alter letter identification, while eliciting a pronounced effect on judgments of the duration of the letters. A further interesting aspect of the present results is that the magnitude of the sound-induced facilitation effect appeared to be invariant across ISIs and contrast levels, but rather varied depending on the specific performance level. Accordingly, in order to distinguish between potential mechanisms of multimodal facilitation, as preparedness enhancement, signal enhancement, and object enhancement, it may be necessary to take interindividual performance differences into account.

Acknowledgments

This study was supported by DFG BA 4110/3-2. We thank Laura Schneider and Sarah-Helena Lutz for their assistance in data acquisition.

References

Agaoglu, S., Agaoglu, M. N., Breitmeyer, B., & Ogmen, H. (2015). A statistical perspective to visual masking. *Vision Research*, *115*, 23–39. http://doi.org/10.1016/j.visres.2015.07.003

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*. http://doi.org/10.1163/187847510X488603

Allman, B. L., & Meredith, M. A. (2007). Multisensory Processing in "Unimodal" Neurons: Cross-Modal Subthreshold Auditory Effects in Cat Extrastriate Visual Cortex. *Journal of Neurophysiology*, *98*(1), 545–549. http://doi.org/10.1152/jn.00173.2007

Asaoka, R., & Gyoba, J. (2016). Sounds modulate the perceived duration of visual stimuli via crossmodal integration. *Multisensory Research*, *29*(4–5), 319–335. http://doi.org/10.1163/22134808-00002518

Bausenhart, K. M., De la Rosa, M. D., & Ulrich, R. (2014). Multimodal integration of time: Visual and auditory contributions to perceived duration and sensitivity. *Experimental Psychology*, *61*(4), 310–322. http://doi.org/10.1027/1618-3169/a000249

Bausenhart, K. M., Rolke, B., & Ulrich, R. (2007). Knowing when to hear aids what to hear. *Quarterly Journal of Experimental Psychology*, *60*(12), 1610–1615. http://doi.org/10.1080/17470210701536419

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension---1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, *50*(1--2), 147–155. http://doi.org/10.1016/S0167-8760(03)00130-2

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*. http://doi.org/10.1038/226177a0

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. http://doi.org/10.1163/156856897X00357

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision*, *2*, 467–479. http://doi.org/10.1167/2.6.4

Chen, K. M., & Yeh, S. L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychologica*, *130*(3), 225–234. http://doi.org/10.1016/j.actpsy.2008.12.008

Chen, Y.-C., & Spence, C. (2011). The crossmodal facilitation of visual object representations by sound: Evidence from the backward masking paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1784–1802. http://doi.org/10.1037/a0025638

Chen, Y. C., Huang, P. C., Yeh, S. L., & Spence, C. (2011). Synchronous sounds enhance visual sensitivity without reducing target uncertainty. *Seeing and Perceiving*, *24*(6), 623–638. http://doi.org/10.1163/187847611X603765

Chen, Y. C., & Spence, C. (2011). The crossmodal facilitation of visual object representations by sound: Evidence from the backward masking paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1784–1802. http://doi.org/10.1037/a0025638

Chen, Y. C., & Spence, C. (2017). Assessing the role of the “unity assumption” on multisensory integration: A review. *Frontiers in Psychology*, *8*(MAR), 1–22. http://doi.org/10.3389/fpsyg.2017.00445

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. http://doi.org/10.1080/01621459.1979.10481038

Correa, Á., Lupiáñez, J., & Tudela, P. (2005). Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychonomic Bulletin and Review*, *12*(2), 328–334. http://doi.org/10.3758/BF03196380

de Haas, B., Cecere, R., Cullen, H., Driver, J., & Romei, V. (2013). The duration of a co-occurring sound modulates visual detection performance in humans. *PLoS ONE*, *8*(1).

http://doi.org/10.1371/journal.pone.0054789

De la Rosa, M. D., & Bausenhart, K. M. (2013). Multimodal integration of interval duration: Temporal ventriloquism or changes in pacemaker rate? *Timing and Time Perception*, *1*(2), 189–215. http://doi.org/10.1163/22134468-00002015

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332–343. http://doi.org/10.1007/s00221-002-1262-y

Hackley, S. A., & Valle-Inclán, F. (2003). Which stages of processing are speeded by a warning signal? *Biological Psychology*. http://doi.org/10.1016/S0301-0511(03)00101-7

Hartcher-O'Brien, J., Di Luca, M., & Ernst, M. O. (2014). The duration of uncertain times: Audiovisual information about intervals is integrated in a statistically optimal fashion. *PLoS ONE*, *9*(3). http://doi.org/10.1371/journal.pone.0089339

Henderson, J. M. (1991). Stimulus discrimination following covert attentional orienting to an exogenous cue. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 91–106. http://doi.org/10.1037/0096-1523.17.1.91

Kinsbourne, M., & Warrington, E. K. (1962). The effect of an after-coming random pattern on the perception of brief visual stimuli. *Quarterly Journal of Experimental Psychology*, *14*(4), 223–234. http://doi.org/10.1080/17470216208416540

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, S14. http://doi.org/10.1068/v070821

Klink, P. C., Montijn, J. S., & van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention, Perception, and Psychophysics*, *73*(1), 219–236. http://doi.org/10.3758/s13414-010-0010-9

Kranczioch, C., & Thorne, J. D. (2013). Simultaneous and preceding sounds enhance rapid visual targets: Evidence from the attentional blink. *Advances in Cognitive Psychology*, *9*(3), 130–142. http://doi.org/10.2478/v10053-008-0139-4

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. In *Psychophysiology* (Vol. 40, pp. 806–817). http://doi.org/10.1111/1469-8986.00081

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, *1173*(1), 102–109. http://doi.org/10.1016/j.brainres.2007.07.050

Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience*, *1*(2), 144–149. http://doi.org/10.1038/393

Madec, S., Rey, A., Dufau, S., Klein, M., & Grainger, J. (2011). The time course of visual letter perception. *Journal of Cognitive Neuroscience*, *24*(7), 1645–1655. http://doi.org/10.1162/jocn_a_00178

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389–391. http://doi.org/10.1126/science.6867718

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.

Meyer, G. F., Wuerger, S. M., Röhrbein, F., & Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. In *Experimental Brain Research* (Vol. 166, pp. 538–547). http://doi.org/10.1007/s00221-005-2394-7

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154–163. http://doi.org/10.1016/S0926-6410(03)00089-2

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. http://doi.org/10.20982/tqmp.04.2.p061

Ngo, M., & Spence, C. (2010). Crossmodal facilitation of masked visual target identification. *Attention, Perception, & Psychophysics*, *72*(7), 1938–1947. http://doi.org/10.3758/APP.72.7.1938

Noesselt, T., Bergmann, D., Hake, M., Heinze, H. J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, *1220*, 157–163. http://doi.org/10.1016/j.brainres.2007.12.060

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H. J., & Driver, J. (2010). Sound-induced enhancement of low-intensity vision: Multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *The Journal of Neuroscience*, *30*(41), 13609–13623. http://doi.org/10.1523/JNEUROSCI.4524-09.2010

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. http://doi.org/10.1163/156856897X00366

Pérez-Bellido, A., Soto-Faraco, S., & López-Moliner, J. (2013). Sound-driven enhancement of vision: Disentangling detection-level from decision-level contributions. *Journal of Neurophysiology*, *109*(4), 1065–1077. http://doi.org/10.1152/jn.00226.2012

Pestilli, F., & Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research*, *45*(14), 1867–1875. http://doi.org/10.1016/j.visres.2005.01.019

Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research*, *29*(9), 1133–1151. http://doi.org/10.1016/0042-6989(89)90061-8

Rijsdijk, J. P., Kroon, J. N., & van der Wildt, G. J. (1980). Contrast sensitivity as a function of position on the retina. *Vision Research*, *20*(3), 235–241. http://doi.org/10.1016/0042-6989(80)90108-X

Rolke, B. (2008). Temporal preparation facilitates perceptual identification of letters. *Perception & Psychophysics*, *70*(7), 1305–1313. http://doi.org/10.3758/PP.70.7.1305

Rolke, B., & Hofmann, P. (2007). Temporal uncertainty degrades perceptual processing. *Psychonomic Bulletin & Review*, *14*, 522–526.

Sarmiento, B. R., Shore, D. I., Milliken, B., & Sanabria, D. (2012). Audiovisual interactions depend on context of congruency. *Attention, Perception, and Psychophysics*, *74*(3), 563–574. http://doi.org/10.3758/s13414-011-0249-9

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*(4), 2200–2208. http://doi.org/10.1016/j.neuroimage.2011.03.075

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*(1), 31–49. http://doi.org/10.1111/nyas.12121

Stein, B. E., London, N., Wilkinson, L., & Price, D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*(6), 497–506. http://doi.org/10.1162/jocn.1996.8.6.497

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. *Cognitive neuroscience*. Cambridge, MA, US: The MIT Press.

Takeshima, Y., & Gyoba, J. (2014). Pattern dot quantity affects auditory facilitation effects on visual object representations. *Perception*, *43*(2–3), 107–116. http://doi.org/10.1068/p7540

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive*

Psychology, *12*(1), 97–136. http://doi.org/10.1016/0010-0285(80)90005-5

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, *80*(1), 1. http://doi.org/10.1037/h0033872

Vroomen, J., & Gelder, B. de. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1583–1590. http://doi.org/10.1037/0096-1523.26.5.1583

Vroomen, J., & Keetels, M. (2009). Sounds change four-dot masking. *Acta Psychologica*, *130*(1), 58–63. http://doi.org/10.1016/j.actpsy.2008.10.001

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(6), 1327–1339. http://doi.org/10.1037/0096-1523.7.6.1327

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. http://doi.org/10.1037/0033-2909.88.3.638

APPENDIX D – Study 4

De la Rosa, M. D., & Bausenhart, K. M. (2018). Still no evidence for sustained effects of multisensory integration of duration. *Multisensory Research*, *31*, 601–622. https://doi.org/10.1163/22134808-18001296

Still no evidence for sustained effects of multisensory integration of duration

Maria D. de la Rosa & Karin M. Bausenhart

University of Tübingen

Address editorial correspondence to Maria Dolores de la Rosa Gamiz, Fachbereich Psychologie, University of Tübingen, Schleichstrasse 4, D-72076 Tübingen, Germany.

Phone: +49 (0)7071 29-76423, FAX: +49 (0)7071 29-2410, email: maria-dolores.de-la-rosa-gamiz@uni-tuebingen.de

Abstract

In studies on temporal order perception, immediate as well as sustained effects of multisensory integration have been demonstrated repeatedly. Regarding duration perception, the corresponding literature reports clear immediate effects of multisensory integration, but evidence on sustained effects of multisensory duration integration is scarce. In fact, a single study [Heron, J., Hotchkiss, J., Aaen-Stockdale, C., Roach, N. W., & Whitaker, D. (2013). A neural hierarchy for illusions of time: Duration adaptation precedes multisensory integration. *Journal of Vision, 13*, 1–12.] investigated adaptation to multisensory conflicting intervals, and found no sustained effects of the audiovisual conflict on perceived duration of subsequently presented unimodal visual intervals. In two experiments, we provide independent evidence in support of this finding. In Experiment 1, we demonstrate that adaptation to audiovisual conflict does not alter perceived duration of subsequently presented visual test intervals. Thus, replicating the results of Heron et al. (2013), we observed no sustained effect of multisensory duration integration. However, one might argue that the prolonged exposure to consistent multisensory conflict might have prevented or hampered multisensory integration per se. In Experiment 2, we rule out this alternative explanation by showing that multisensory integration of audiovisual conflicting intervals is still effective after adaptation to audiovisual conflict. This further strengthens the conclusion that multisensory integration of interval duration affects perception in an immediate, but not in a sustained manner.

Keywords: Time perception, multisensory integration, temporal ventriloquism, duration adaptation, recalibration.

Still no evidence for sustained effects of multisensory integration of duration

Human temporal perception is highly susceptible to distortion, such that the perceived temporal characteristics of a stimulus often do not correspond to its physical duration. Such perceptual distortions can be triggered, for example, by the context of stimulus presentation (e.g., Bausenhart, Bratzke, & Ulrich, 2016; Shi, Chen, & Müller, 2010). Specifically, temporal perception of a target stimulus can be altered by accompanying (e.g., Bausenhart, De la Rosa, & Ulrich, 2014; K. M. Chen & Yeh, 2009; De la Rosa & Bausenhart, 2013; Klink, Montijn, & van Wezel, 2011; Romei, De Haas, Mok, & Driver, 2011; Walker & Scott, 1981) or even preceding (Dyjas, Bausenhart, & Ulrich, 2014; Heron et al., 2012; Penton-Voak, Edwards, Percival, & Wearden, 1996) stimulation. One example for such context effects evoked by preceding stimulation is the phenomenon of *duration adaptation,* in which the perceived duration of a target interval is prolonged (shortened) after repeated exposure to an adaptation interval of shorter (longer) duration (Heron et al., 2012). Interestingly, this repulsive aftereffect is only observed when adaptation and test intervals stem from the same modality, that is, duration adaptation is modality-specific (Heron et al., 2012; Walker, Irion, & Gordon, 1981).

Nonetheless, there is also ample evidence for context effects under conditions of multisensory stimulation. In the domain of multisensory integration, such context effects are addressed in situations where the target and the irrelevant stimulation emerge from different modalities. In the following, we will distinguish between immediate effects and sustained effects (i.e., aftereffects) of multisensory integration on temporal perception (cf. L. Chen & Vroomen, 2013). *Immediate* multisensory effects refer to changes in the perception of a current event caused by multimodal stimulation (e.g., when a visual stimulus is perceived differently depending on whether it is accompanied by an auditory stimulus or not). *Sustained* multisensory effects refer to changes in the perception of a current event that are caused by previous multimodal stimulation (e.g., if repeated presentation of a certain multimodal stimulus induces changes in the perception of a subsequent event).

First, immediate multisensory integration effects become clearly evident when the different modalities provide conflicting information. One prominent example for this is the temporal ventriloquism effect (Bertelson & Aschersleben, 2003; Keetels, Stekelenburg, & Vroomen, 2007; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Vroomen & de Gelder, 2004). Specifically, the time point at which a visual stimulus is perceived can be biased towards the time point of an accompanying auditory stimulus presented slightly before or after the visual one, as indicated by studies on temporal order perception. It has been suggested that this effect emerges in perceptual stages of processing, since even the perception of non-temporal features of the visual target stimulus can be altered by auditory temporal information presented slightly before or after the visual one (Morein-Zamir et al., 2003; Vroomen & Keetels, 2009). Moreover, a related effect also emerges in duration perception (rather than order perception). Specifically, the perceived duration of a visual target interval is biased towards the duration of an accompanying auditory interval with slightly shorter or longer duration (Bausenhart et al., 2014; Heron et al., 2013; Klink et al., 2011; Romei et al., 2011). Typically, the reversed effect, that is, an influence of visual irrelevant information on perceived duration of auditory target intervals is much smaller (Bausenhart et al., 2014) or even absent (Klink et al., 2011) This suggests that auditory temporal information, presumably due to the typically higher temporal resolution of the auditory compared to the visual system (Bertelson & Aschersleben, 2003; Getzmann, 2007; Morein-Zamir et al., 2003), usually takes a predominant role in the multisensory percept. However, this typical auditory predominance is not “hardwired”. In fact, when the reliability of the presented auditory temporal information is degraded, the impact of the visual information on audiovisual duration estimates increases (Hartcher-O’Brien, Di Luca, & Ernst, 2014). This is in line with a Bayesian cue combination account of multisensory integration, which assumes that the modalities’ contributions to the combined multisensory percept are determined on basis of their inverse variances, thus leading to statistically optimal multisensory estimates (Ernst & Banks, 2002).

Additionally, by demonstrating in a reminder task that the auditory bias on perceived visual duration remains constant over a large range of durations of the target interval, De la Rosa et al. (2013) attributed this phenomenon to a temporal ventriloquism effect operating on perceived onset and offset of the interval, and thus, the switch latency of a pacemaker-accumulator mechanism (Gibbon, 1977; Gibbon, Church, & Meck, 1984; see also Hartcher-O'Brien et al., 2014, for a related idea framed within the concept of Bayesian cue integration). Thus, similar cognitive mechanisms might underlie the immediate multisensory effects on the perception of temporal order and of duration.

Second, there is also evidence for sustained changes of temporal perception, which are induced by preceding multisensory stimulation (Di Luca, Machulla, & Ernst, 2009; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Hanson, Heron, & Whitaker, 2008; Harrar & Harris, 2008; Heron, Roach, Whitaker, & Hanson, 2010; Heron, Whitaker, McGraw, & Horoshenkov, 2007; Keetels & Vroomen, 2007; Machulla, Di Luca, Froehlich, & Ernst, 2012; Vatakis, Navarra, Soto-Faraco, & Spence, 2007; Vroomen, Keetels, De Gelder, & Bertelson, 2004). For example, Vroomen et al. (2004) repeatedly presented a brief audiovisual stimulus pair with a slight, but consistent conflict between the modalities (e.g., the auditory stimulus always preceded the visual one by 100 ms). During prolonged exposure to this asynchrony, the perceptual system "recalibrated", which became evident as a shift of the psychometric function in a subsequent test phase. For example, a physically simultaneous audiovisual stimulus pair was then judged as non-simultaneous, such that the visual stimulus appeared to precede the auditory one. Such flexibility of perception to recalibrate to persistent multisensory asynchrony was since then consistently demonstrated in a large number of studies involving temporal order judgments (e.g., Di Luca et al., 2009; Hanson et al., 2008; Harrar & Harris, 2008; Heron et al., 2010, 2007; Keetels & Vroomen, 2007; Machulla et al., 2012; Vatakis et al., 2007). Moreover, there is also ample evidence regarding multisensory recalibration and its underlying mechanisms in the spatial domain (e.g., L. Chen & Vroomen, 2013; Witten & Knudsen, 2005; Zaidel, Ma, & Angelaki, 2013; Zaidel, Turner, & Angelaki, 2011), and thus, recalibration seems to be a general phenomenon in perceptual processing of conflicting multisensory stimulation.

Nonetheless, comparable evidence regarding such sustained multisensory effects in the domain of duration perception is much sparser. In fact, we are aware of only one published study which addressed this issue (Heron et al., 2013). In two main experiments (and a series of clever control experiments), these authors basically replicated the aforementioned immediate multisensory effects, and demonstrated the lack of sustained multisensory effects on perceived duration. Regarding immediate effects, visually marked target intervals of 320 ms duration were presented and their corresponding perceived duration was assessed in a 2AFC task employing the method of constant stimuli. Importantly, these task-relevant target intervals were accompanied by either shorter (200 ms) or longer (510 ms) task-irrelevant auditory intervals. Perceived duration of these visual target intervals was biased towards the duration of the auditory intervals. Specifically, the Point of Subjective Equality (PSE) for a unimodal visual test stimulus with the aforementioned multimodal intervals was 277 ms for shorter and 375 ms for longer auditory accompanying durations.

To test for sustained multisensory effects, the same conflicting audiovisual intervals as described above were presented repeatedly in an adaptation phase. In a subsequent test phase, participants had to reproduce the duration of visual unimodal test intervals of either 160, 320, or 640 ms duration. According to Heron et al. (2013), two potential outcomes were hypothesized, depending on the hierarchical order of duration adaptation and multisensory integration in stimulus processing. On the one hand, the audiovisual stimuli in the adaptation phase might be perceptually integrated before duration adaptation occurs. Consequently and in accordance with the multisensory integration results reported above, perceived duration of the visual adapting intervals would be biased towards the duration of the accompanying auditory ones. Participants would then adapt to these either prolonged or shortened perceptual representations, and consequently, in the test phase, repulsive adaptation aftereffects on reproduced duration should be observed. On the other hand, duration adaptation might precede multisensory integration – such that adaptation would operate on the separate, modality-specific representations of the adaptation stimuli. Consequently, the adapted modality-specific representation of the visual 320 ms adaptation interval would not be biased towards the shorter or longer auditory duration, and the unimodal visual 320 ms test interval could be reproduced without any repulsive aftereffect.

Alternatively, one might consider another potential sustained effect of multisensory integration: participants might not adapt to the bimodal integrated durations, but adjust perception to reduce the multimodal conflict between the interval markers. Consequently, they would adapt a tendency to perceptually expand (condense) any visually marked interval, according to whether it was previously presented along with a longer (shorter) auditory one. Then, the unimodal visual test stimuli in the test phase should also be perceived and consequently reproduced as longer (shorter), respectively. This "perceptual attraction" hypothesis was not considered by Heron et al., and it seems rather unlikely given that the phenomenon of duration adaptation is quite well established and even has been replicated under bimodal stimulation conditions (cf. Figs. 1C and 5B in Heron et al., 2013).

The results of Herons et al.'s (2013) experiment were consistent with their second hypothesis, that is, unimodal reproductions in the test phase were not affected by the multimodal conflicting intervals presented during duration adaptation. Consequently, the authors argued that duration adaptation precedes multisensory integration in the processing chain, and this interpretation received further support in a series of control experiments. Important for the present purposes, the authors could also exclude the possibility that during the adaptation phase, participants simply ignored the biasing auditory stimulation – which otherwise also might explain why no sustained effects of multisensory integration were observed. Specifically, in a control experiment, audiovisual adaptation intervals were repeatedly presented with identical durations of either 640 or 160. Subsequently, test stimuli could be either 320 ms unimodal auditory, bimodal, or unimodal visual durations. Therefore, if participants simply ignored the biasing auditory stimulation during the adaptation phase, then a repulsive duration aftereffect should be only observed when the test stimuli contained visual durations. Likewise, if attention was only directed towards the auditory stimulation, then a repulsive duration aftereffect should be only observed when the test stimuli contained auditory durations. Interestingly, for all three types of test stimuli repulsive duration aftereffects were observed, which indicates that neither auditory nor visual stimulation was ignored during duration adaptation. Therefore, this result rules out the possibility that ignoring biasing auditory stimulation explains the lack of repulsive adaptation aftereffect, as well as that directing attention only towards auditory stimulation (and thus, ignoring visual stimulation) explains the lack of sustained effect of multisensory integration

Heron et al.'s (2013) results are highly interesting for two reasons: First, they close the research gap regarding sustained effects of multisensory integration on duration perception. Second, they inform about the hierarchical order of duration adaptation and multisensory integration – two phenomena that had only been regarded in isolation so far – and thus, provide important and novel insights into the cognitive architecture of temporal processing. The aim of the present study is twofold: In Experiment 1, we aim to provide independent evidence regarding whether or not sustained effects of multisensory integration can be observed[1]. This seems especially important since the respective conclusion is based on a null effect. We chose a set of conditions that are presumably even more effective in producing multisensory integration effects than the ones in the study of Heron et al. (2013), and we employed a larger sample size, thus yielding higher experimental power. In Experiment 2, we aim to rule out an alternative explanation according to which prolonged exposure to multisensory conflict might attenuate or even eliminate multisensory integration. To this end, we introduce a novel and crucial control condition to ensure that multisensory integration mechanisms are potentially effective even after prolonged exposure to multisensory conflict.

---

[1] In fact the present experiments were designed independently and prior to publication of Heron et al.'s (2013) study and thus should not be viewed as a direct replication attempt.

Experiment 1

This experiment is similar to the main adaptation experiment of Heron et al. (2013). Audiovisual interval pairs with a persistent conflict were repeatedly presented during an adaptation phase. In each trial of a subsequent test phase, a unimodal visual test interval was presented after several top-up presentations of the adapting stimuli and participants had to reproduce the duration of the test interval. As hypothesized by Heron et al. (2013), if multisensory integration precedes duration adaptation, perceived duration of the visual adaptation stimuli should be biased towards the auditory conflicting durations, and consequently, repulsive duration aftereffects should be observed in the reproductions of the test phase. If multisensory integration operates after duration adaptation, no repulsive aftereffects would be observed in the test phase, since the adapted duration within the visual modality would not differ from the visual unimodal duration in the test phase. Finally, if participants would recalibrate perceived duration in the sense of a multisensory conflict reduction, but no duration adaptation occurs, sustained multisensory effects in the form of perceptual attraction (rather than repulsion) should be observed.

The specific duration of the visual intervals and the auditory conflict were chosen such that the expected multisensory integration effects would be especially pronounced. Specifically, a 753 ms visual interval was combined with sounds of either 653, 753, or 853 ms in the adaptation phase. That is, the difference between auditory shorter and longer conflicting intervals, and thus the maximal biasing effect due to multisensory integration, was Δ_{max} = 853ms - 653 ms = 200 ms and the proportional conflict (i.e. the auditory bias in relation to visual interval duration) amounted to (853 ms – 753 ms) / 753 ms = 0.13. Under similar conditions, De la Rosa et al. (2013, Experiment 2), found that the observed bias in reproduced duration amounted to approximately 70 % of $\Delta_{max.}$

In Heron et al.'s study, Δ_{max} was 510 ms - 200 ms = 310 ms, and the maximal proportional conflict was (510 ms – 320 ms) / 320 ms = 0.59. The observed multisensory integration effect (i.e., the PSE difference between the longer and shorter auditory condition in the multisensory integration experiment) amounted to 98 ms, and thus to only about 32 % of Δ_{max}. Somewhat counterintuitively, one plausible account for this comparably small integration effect is the pronounced proportional conflict between visual and auditory information in Heron et al.'s study. In fact, multisensory integration is especially effective within a rather small "window of integration" (Asaoka & Gyoba, 2016; Klink et al., 2011). When the conflict between two modalities gets too pronounced, an important prerequisite of multisensory integration, the *assumption of unity* (i.e., the assumption that the input from the different modalities stems from a single object; Welch & Warren, 1980) is violated. As a consequence, the modalities' inputs are no longer or, at least, less effectively integrated to a common multimodal percept. It may well be that the large multimodal conflict in Heron et al.'s study, especially with regard to the relatively brief visual interval duration, might have produced comparably small and less stable integration effects. This in turn, might also reduce the potential for eliciting and observing sustained effects of this integration. Thus, maximizing the effect of multisensory integration during the adaptation phase might increase the potential for observing any sustained effects of multisensory integration in the test phase. [2]

Method

Participants. 18 participants (16 females) between 18 and 30 years (*M* = 22.06, *SD* = 3.34 years) voluntarily took part in Experiment 1 in exchange of either course credit or money. For this sample size, a power of .82, and thus a probability for a Type II error of $\beta = 1 - .82 = .18$ is achieved with $df_n = 2$, $df_d = 34$, and $\alpha = .05$ for the main effect of multisensory bias condition, if we assume a medium effect size ($f = 0.3$ or $\eta^2_p = .08$, computed with pwr.f2.test of the pwr package for R. Note that this effect size estimate is very conservative, compared to the large achieved effect size of Bias

---

[2] It should be noted that in a further control experiment, Heron et al. (2013) tested for "physical" adaptation effects by presenting visual unimodal adaptation stimuli which physically corresponded to the multisensory biased durations that were assessed in the multisensory integration experiment. Such "physical" adaptation effects were clearly observed – thus demonstrating that the evoked multisensory bias was, in principle, large enough to induce observable effects of duration adaptation. Nonetheless, it remains unclear whether the observed effects of multisensory integration in the dedicated multisensory integration experiment are identical to those presumably evoked in the adaptation phase of the adaptation experiment. In Experiment 2 of the present study, we will address this issue more directly.

condition in the present Experiment 2, which amounts to $f = 1.25$ or $\eta^2_p = .61$).

Apparatus and Stimuli. The experiment was run on a Macintosh computer connected to a CRT monitor running at 150 Hz. The presentation of all of stimuli and data recording were controlled by Matlab with the Psychophysics Toolbox extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Participants were seated in a dimly lit room at a viewing distance of approximately 55 cm from the computer screen.

Visual stimuli consisted of a small white dot (1.8 mm diameter, 0.19º visual angle) used as fixation point, a white disc (6.4 mm diameter, 0.66º visual angle) which was used to demarcate the visually defined intervals, and an empty green square (4.5 mm diameter, 0.47 visual angle), which was employed as a target stimulus for a control task. All these visual stimuli were presented at the center of the screen on a black background. The auditory intervals were defined by a pure sinusoidal tone (800 Hz, 79 dB) played binaurally through headphones with ramped 5 ms on- and offset.

Responses were collected with a standard German keyboard. Participants were requested to use the *m* key for the reproduction task, and the numeric keypad for the control task. Correctness of stimulus timing and audiovisual (a)synchronies were verified with an oscilloscope.

Procedure. During the whole experiment, adaptation and test phases alternated (cf. Figure 1). The experiment always started with an adaptation phase. At the beginning of the experiment, participants received written instructions on two different tasks they would have to perform during the experiment. For the counting task (which applied to the adaptation phases), participants had to watch the repeated presentation of a visual stimulus while counting green squares presented randomly among these visual stimuli. Later they were required to enter their count through the keypad. For the timing task (which applied to the test phases), participants were instructed to make temporal reproductions as close as possible to the duration of a visually defined interval. In both tasks, participants were instructed to ignore any sounds that were presented along with the visual stimuli. Prior to each phase, participants were informed about which of two tasks would follow in the next block.

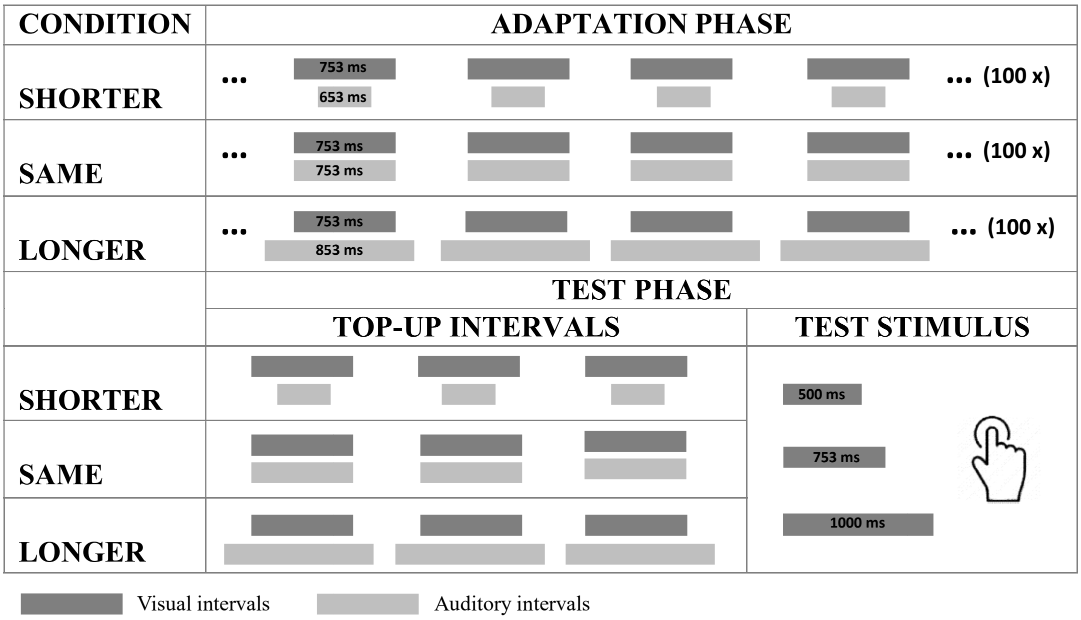

Figure 1. Schematic representation of adaptation and test phases in the different bias conditions for Experiment 1.

Adaptation phase

In each adaptation phase, the white disc was repeatedly presented for 100 times with a constant duration of 753 ms. The interstimulus interval (ISI) between two subsequent visual interval presentations varied randomly within a range of 240 – 747 ms. Concurrently to the presentation of each visual interval an auditory stimulus was presented with a duration according to three different multisensory bias conditions: “longer”, “same”, and “shorter”. In the “longer” condition, the tone started 50 ms before the onset and ended 50 ms after the offset of the visual interval. In the “shorter” condition, the tone started 50 ms after the onset and ended 50 ms before the offset of the visual interval. In the “same” condition, the auditory and visual intervals started and ended simultaneously. The three multisensory bias conditions were presented in separate adaptation phases whose order of presentation was counterbalanced between participants.

In order to ensure that participants attended to the visual stimulation in the adaptation phase, they were instructed to count green squares which were randomly interspersed among the audiovisual intervals. The number of green squares varied randomly between 5 and 9 in each adaptation phase and the squares were presented for a random duration between 120 and 373 ms within randomly picked ISIs.

At the end of the train of audiovisual interval stimuli, the question “Wie viele grüne Quadrate hast du gesehen?” (“How many green squares did you see?”) was displayed at the center of the screen and remained until the participant typed a digit with the keyboard. Afterwards, feedback information was given about the response for 5 sec: the word “Korrekt!” indicated a correct response and the word “Falsch!” indicated a wrong response. Then, participants were informed of the end of the block and that they could start the next block (requiring the timing task) by pressing any key. With this information, each adaptation phase ended and the subsequent test phase began.

Test phase

Each trial of each test phase started with the presentation of a blank screen for 1000 ms. Then, a series of three audiovisual top-up intervals was presented with stimuli, duration, audiovisual bias and (randomly varying) ISIs exactly as in the immediately preceding adaptation phase. After the third audiovisual top-up interval, a small fixation dot was displayed for 1000 ms. Then, a unimodal visual stimulus was presented as test stimulus for an interval duration of either 500, 753, or 1000 ms. These durations were equally likely and randomized across trials.

200 ms after the end of this interval, the word “Jetzt” (“Now”) was displayed at the center of the screen requesting the participant to reproduce the duration of the visual stimulus with a single press of the *m* key. This response prompt disappeared as soon as the key was pressed. Then, the screen remained empty and the next trial started 500 ms after the key was released.

Each test phase consisted of a block of 90 trials. A short break was inserted after each 15 trials, and the mean absolute percentage of deviation of the reproduced durations from the visual test duration in the preceding 15 trials was presented as feedback. This information prevented any hint of whether participants had over- or underestimated the visual stimulus durations. Participants were instructed to keep this score as low as possible during the experiment, since lower values would correspond to more accurate reproductions. Participants could self-terminate the breaks by pressing any key. After all 90 trials of the test phase were performed, participants were informed about the end of the current phase and that they could initiate the next phase (requiring the counting task) by pressing any key.

The experiment consisted of overall 4 cycles of adaptation and test phase. The first cycle constituted practice and consisted of 20 repetitions of the audiovisual interval in the adaptation phase and 15 trials in the test phase. Each of the three experimental cycles consisted of 100 repetitions of the audiovisual interval in the adaptation phase and 90 trials in the test phase (3 test interval durations * 30 repetitions), and the order of the audiovisual bias conditions was counterbalanced between participants. The bias condition employed in the practice cycle corresponded to the one of the first experimental cycle.

Results

Data from the practice block were discarded from analysis. For each participant and condition, mean and standard deviation (SD) of the reproduced duration were computed. Trials with reproduced durations below or above the individual condition means ± 2*SD were discarded from data analysis (4.30 % of trials). A repeated-measures analysis of variance (ANOVA) was conducted with the factors Bias Condition (longer, same, and shorter) and Interval Duration (500, 753, and 1000 ms) on the individual mean reproductions. Greenhouse-Geisser adjusted *p*-values are reported whenever appropriate.

The results show that mean reproduced duration increased with increasing Interval Duration,

$F(2, 34) = 225.64$, $p < .001$, $\eta^2_p = .93$ (Figure 2). Specifically, the 500 ms, 753 ms, and 1,000 ms test intervals were on average reproduced as 531.05 ms (*SD* = 103.02 ms), 750.80 ms (*SD* = 123.73 ms), and 914.63 ms (*SD* = 143.71 ms), respectively. However, neither a significant effect of Bias Condition nor an interaction between both factors was observed (both $F < 1$).

Following the suggestion of an anonymous reviewer, we also conducted an additional analysis of reproduced duration employing the "BayesFactor" package for R. The results of this analysis are consistent with the ANOVA results reported above. Specifically, comparison against a null model yielded the highest Bayes Factor ($BF_{10} = 3.69e^{53} \pm 1.96\%$) for a model only including the factor interval duration. Setting this model as the new null model and comparing it to fuller models including additionally (A) a main effect of bias condition, and (B) a main effect of bias condition and an interaction of interval duration and bias condition, yielded BF_{10} of 0.069 ± 5.09% for (A) and 0.004 ± 1.93% for (B). Restating in terms of BF_{01}, the present data are 14.5 times more likely under the model only including interval duration than under (A), and 277.8 times more likely than under (B).

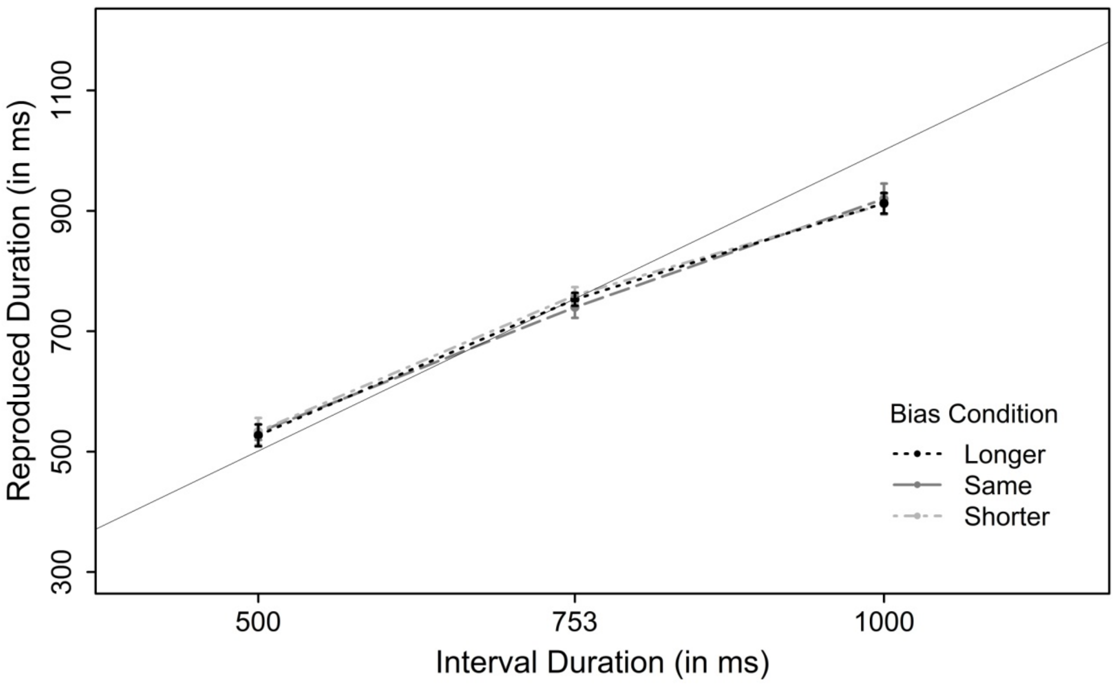

Figure 2. Mean reproduced duration as a function of Bias Condition and Interval Duration in Experiment 1. Error bars represent ±1 within-subject standard error of the mean with a correction for between-subjects variability (Morey, 2008) and the continuous black line depicts the identity line.

An analogous repeated-measures analysis of variance was conducted on the standard deviation (SD) of the individual reproductions to inform about the variability of perception. This analysis showed a main effect of Interval Duration, $F(2, 34) = 23.36$, $p < .001$, $\eta^2_p = .58$, which reflects increasing variability with increasing interval duration. Specifically, SDs amounted to 96.84, 115.26, and 121.23 ms for the 500, 753, and 1000 ms interval durations. Neither the main effect of Bias Condition $F(2, 34) = 0.11$, $p = .896$, $\eta^2_p = .01$, nor the interaction of Bias Condition and Interval Duration, $F(4, 68) = 2.35$, $p = .097$, $\eta^2_p = .12$, were significant.

Discussion

As expected, reproduced duration clearly increased with the duration of the test interval. Most importantly, however, there was no effect of the auditory bias presented in the preceding adaptation block on reproduced duration. Accordingly, no sustained effects of multisensory integration, neither in the form of repulsion nor attraction, were observed. Thus, the present findings are consistent with the respective results of Heron et al. (2013) and also with their interpretation that multisensory integration emerges rather late (i.e., after duration adaptation) in the processing stream. Importantly, the specific set of visual and auditory durations chosen for stimulation elicits especially strong immediate multisensory integration effects (De la Rosa & Bausenhart, 2013, Exp. 2) and the size of the tested sample yielded sufficient power to observe a potential sustained effect of this integration. Therefore, on basis of the present results there is no reason to reject the null hypothesis that multisensory integration of audiovisual intervals has no sustained effects on the perceived duration of subsequently presented unimodal intervals.

However, both Heron et al.'s (2013) adaptation experiment as well as the present Experiment 1 are vulnerable to a crucial alternative explanation. This alternative is based on existing evidence that multisensory integration per se can be hampered or suppressed under certain conditions of stimulation (e.g., Mahani, Sheybani, Bausenhart, Ulrich, & Ahmadabadi, 2017; Powers, Hillock, & Wallace, 2009; Sarmiento, Shore, Milliken, & Sanabria, 2012). Specifically, the repeated presentation of the multisensory biased stimuli during the adaptation phase might actually affect the efficacy of the multisensory integration mechanisms, and thus, even prevent immediate effects of this integration. Consequently, if there was no "immediate" multisensory integration in the adaptation phase, no sustained effects of this integration could be expected as well. Even though the specific audiovisual stimulus combinations evoked clear immediate effects of multisensory stimulation in the dedicated multisensory integration experiments of Heron et al. (2013) and De la Rosa et al. (2013), these experiments were fundamentally different from the experiments that tested for sustained effects. Specifically, in the former experiments, various stimulus durations and multisensory conflicts were presented randomly interleaved. In the latter experiments, participants were exposed to a long and repetitive stimulation with a large and consistent audiovisual conflict during the adaptation phase.

Crucially, such exposure to consistent conflict might actually decrease the likelihood of multisensory integration per se. For example, it has been argued that participants may change their multisensory processing strategy from integration to selective processing of the modalities during prolonged exposure to multisensory conflicting stimulation (Mahani et al., 2017). Moreover, extensive perceptual training of an audiovisual simultaneity judgment task narrows the temporal window of multisensory binding (Powers et al., 2009) and consequently might also affect multisensory integration, even though this relation is not firmly established. Finally, in duration perception it has been shown that the proportion of conflicting versus congruent audiovisual interval pairs affects multisensory integration. Specifically, the more conflicting intervals are presented, the smaller the observed multisensory integration effect - and crucially, this reduction of multisensory integration even aggravates with prolonged exposure to the multisensory stimulation (e.g., when comparing the 2nd to the 1st half of trials) (Sarmiento et al., 2012). Consequently, one might argue that the prolonged exposure to a consistent and pronounced multisensory conflict, as in the adaptation phase of the present Experiment 1 and of Heron et al.’s (2013) adaptation experiment, by itself prevented multisensory integration of interval duration. Consequently, neither immediate nor sustained effects of integration would be observed after duration adaptation. In this case, however, the lack of effect of the multisensory bias after duration adaptation should be attributed to absence of multisensory integration in general, rather than absence of sustained effects of multisensory integration. In Experiment 2, we test whether this alternative explanation holds.

Experiment 2

In this experiment, we assessed whether the prolonged exposure to consistent multisensory conflict, as administered in Experiment 1, prevents multisensory integration in general. To this end, we replicated Experiment 1 with one crucial change. The visual test interval in the test phase of the experiment was now also accompanied by an auditory stimulus. If the prolonged exposure to multisensory conflict during the adaptation phase indeed prevents multisensory integration, then no immediate effects of the multisensory stimulation should be observed during the test phase. On the other hand, if such immediate effects are observed, this would imply that multisensory integration was also effective in Experiment 1, but without evoking any sustained effect on perception of the test intervals.

Method

Participants. 18 participants (15 females; age range: 18 – 41 years; mean age: 21.89 years) voluntarily took part in Experiment 2 in exchange of either course credits or money.

Apparatus and Stimuli. Stimuli, apparatus, and procedure were identical to those employed in Experiment 1, with only one exception: in the test phase of Experiment 2, each visual test interval was presented concurrently with an auditory interval. As in Experiment 1, the visual interval duration could be either 500, 753, or 1000 ms. The duration of the auditory interval varied relatively to this test interval according to the same audiovisual bias condition as employed in the immediately preceding adaptation phase and the top-up pairs. Consequently, the audiovisual bias condition of the visual test stimulus was manipulated blockwise and always coincided with the audiovisual bias condition in the preceding adaptation phase. Specifically, in the longer condition, the auditory intervals were 600, 853, and 1100 ms for the 500, 753, and 1000 ms visual intervals, respectively. In the same condition, the auditory intervals were identical to the visual durations. Finally, in the shorter condition, the auditory intervals were 400, 653, and 900 ms for the 500, 753, and 1000 ms visual intervals, respectively. As in Experiment 1, participants were instructed to reproduce the duration of the visual intervals and ignore any sound.

Results

The practice block was discarded from data analysis. For each participant and condition, mean and standard deviation (SD) of the reproduced duration were computed. Trials with reproduced durations below or above the individual condition means ± 2*SD were discarded from data analysis (4.32 % of trials). A repeated-measures analysis of variance (ANOVA) was conducted with the factors Bias Condition (longer, same, and shorter) and Interval Duration (500, 753, and 1000 ms) on the individual mean reproductions. Greenhouse-Geisser adjusted *p*-values are reported whenever appropriate.

The results showed significantly longer reproduced duration with increasing Interval Duration, $F(2, 34) = 318.18$, $p < .001$, $\eta^2_p = .95$. Specifically, participants reproduced the 500, 753, and 1,000 ms intervals on average as 515.51 ms (SD = 163.76 ms), 829.14 ms (SD = 181.80 ms), and 1026.42 ms (SD = 203.23 ms), respectively. Theoretically most important, reproductions were longer in the longer condition (M = 856.63 ms, SD = 273.38 ms), intermediate in the same condition (M = 802.17 ms, SD = 265.25 ms), and shorter in the shorter condition (M = 712.26 ms, SD = 283.77 ms), $F(2, 34) = 26.65$, $p < .001$, $\eta^2_p = .61$ (Figure 1). Furthermore, there was a marginal interaction of both factors, $F(4, 68) = 2.55$, $p = .078$, $\eta^2_p = .13$. This interaction reflects a slight decrease in the effect of bias condition ($\Delta = M_{longer} – M_{shorter}$) on reproduced duration with increasing interval duration ($\Delta_{500\text{ ms}}$ =188.40 ms, $\Delta_{753\text{ ms}}$ = 127.08 ms, and $\Delta_{1000\text{ ms}}$ = 117.63 ms) cf. Figure 3). On average, the observed multisensory conflict effect amounted to Δ = 144.37 ms, and thus (144.37 ms / 200 ms) * 100 = 72 % of Δ_{max}.

Again, these ANOVA results were basically corroborated by an additional BayesFactor Anova. This analysis yielded the highest Bayes Factor ($BF_{10} = 3.80e^{62} \pm 3.4\%$) for the model with main effects of interval duration and bias condition in comparison to the null model. Additional comparisons of a model only including interval duration with models including additionally (A) a main effect of bias condition, and (B) a main effect of bias condition and an interaction of interval duration and bias condition, yields BF_{10} of 762,701,262,322 ± 1.12% for (A) und 121,265,608,053

± 1.77% for (B). That is, the data are most likely under a model including main effects of interval duration and bias condition (around 762 billion as likely as under a model only including a main effect of interval duration). Restating in terms of BF_{01}, the likelihood of the data given a model only including interval duration is very close to zero (<<1/1000 in both cases) compared to models including effects of bias condition.

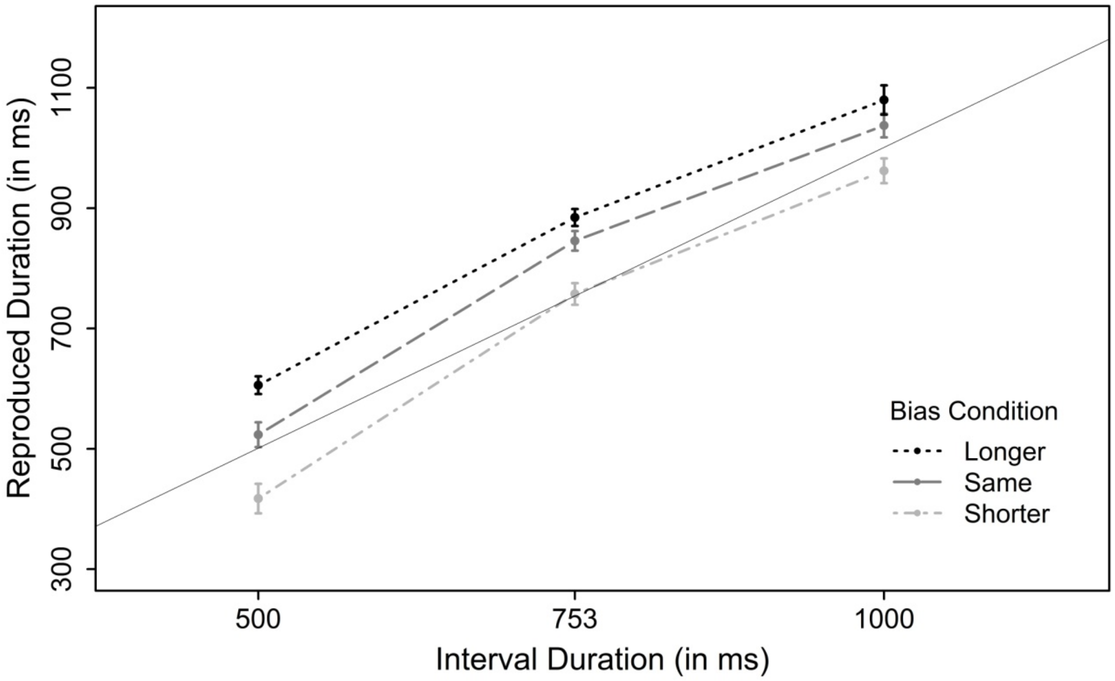

Figure 3. Mean reproduced duration as a function of Bias Condition and Interval Duration in Experiment 2. Error bars represent ±1 within-subject standard error of the mean with a correction for between-subjects variability (Morey, 2008) and the continuous black line depicts the identity line.

An analogous repeated-measures analysis of variance on the standard deviation (SD) of the individual reproductions again showed a main effect of Interval Duration, $F(2, 34) = 6.92$, $p = .008$, $\eta^2_p = .29$, which reflects increasing variability with increasing interval duration. Specifically, SDs amounted to 117.39, 136.98, and 144.14 ms for the 500, 753, and 1000 ms interval durations. In addition, there was a small main effect of Bias Condition $F(2, 34) = 3.82$, $p = .032$, $\eta^2_p = .18$, basically reflecting numerically smaller SD in the shorter (120.48 ms) than in the longer (135.79 ms) and the same (144.14 ms) condition. However, according to Bonferroni-corrected pairwise comparisons, SDs in none of the conditions differed significantly from each other, all $p >$. 50. There was no interaction of Bias Condition and Interval Duration, $F < 1$. [3]

Discussion

The auditory biasing stimulation presented concurrently with the visual test intervals clearly led to a pronounced bias on reproduced durations. This bias amounted to 72 % of the Δ_{max}, and thus is almost identical to the observed bias in a previous reproduction experiment (~70 %) which contained similar interval durations and bias conditions, but no preceding adaptation phase or top-up-stimulation (De la Rosa & Bausenhart, 2013, Exp. 2). Therefore, we conclude that the exposure to persistent conflict during the adaptation phase did not hamper multisensory integration of audiovisual interval duration in the present experimental setup. This also strengthens the interpretation of Experiment 1. Specifically, even though immediate multisensory integration presumably emerged during the adaption phase, no sustained effects of this integration could be observed in the test phase.

General Discussion

---

[3] As an anonymous reviewer pointed out, the provided blockwise feedback about the magnitude (but not the direction) of the reproduction biases (i.e., the absolute deviation from veridical) might have caused participants to adjust their reproductions from block to block in "an unspecified direction" in order to reduce the reproduction bias. Accordingly, in conditions with large reproductions biases, large variability of reproductions should be observed. However, the numerically largest SD of reproductions was observed in the "same" condition, while the perceptual bias associated with this condition was clearly smaller than the one observed in the long condition. Moreover, a very similar pattern regarding the magnitude of perceptual bias and reproduction variability with respect to bias condition was observed in a previous, comparable study (De la Rosa & Bausenhart, 2013, cf. Figs. 2 & 4), in which bias condition varied randomly from trial to trial and thus no differential adjustment of the reproductions for the different bias conditions following blockwise feedback was feasible. In any case, this issue does not hamper our main conclusions of the presence of pronounced immediate multisensory effects in Experiment 2, but the absence of sustained effects in Experiment 1.

The present study addressed immediate and sustained effects of multisensory duration integration. Two conducted interval reproduction experiments provided clear-cut results: First, adaptation to audiovisual conflicting intervals did not alter perceived duration of subsequently presented visual test intervals. Second, adaptation to audiovisual conflicting intervals did not prevent, however, multisensory integration per se. Specifically, for subsequently presented audiovisual conflicting intervals, a strong multisensory bias effect on reproduced duration was observed. Thus, replicating the results of Heron et al. (2013), we observed no sustained effect of multisensory duration integration. In addition, and extending this previous study, we demonstrate clear immediate effects of multisensory integration even after prolonged exposure to consistent multisensory conflict.

Heron et al. (2013) already suggested that multisensory integration follows lower-level perceptual effects as duration adaptation within the processing hierarchy. Thus, the multisensory integration effect might be allocated in higher perceptual stages of processing. On the basis of the present and Heron et al.'s results, one cannot rule out, however, an even later locus within a post-perceptual, decisional processing stage. Specifically, the audiovisual conflict might not have altered perceived duration of the visual intervals, but participants might have just reproduced the auditory temporal information rather than an integrated multimodal percept. This interpretation seems rather unlikely, however, since multisensory duration integration can also be observed under a number of conditions that cannot easily be reconciled with a simple decisional bias.

First, Bausenhart et al. (2014) demonstrated that multisensory duration integration also occurs when a high proportion of unimodal intervals is unpredictably interspersed within the multimodal trials, or when the auditory modality only provides a single pulse-like stimulus which biases either the onset or the offset of the visual interval. Under such conditions, responding to the auditory information alone is hindered or at least requires effortful intermodal attentional shifts between or even within trials, which are typically detrimental to duration discrimination performance (Grondin, Ivry, Franz, Perreault, & Metthé, 1996; Rousseau, Poirier, & Lemyre, 1983; Ulrich, Nitschke, & Rammsayer, 2006). Second, perceived duration of audiovisual intervals does not increase linearly with auditory biasing duration, but remains constant or even decreases again if the audiovisual conflict exceeds a certain value (e.g., Klink et al., 2011) – which is in line with an interpretation in terms of a limited temporal "window of integration" (e.g., Meredith, Nemitz, & Stein, 1987; Spence & Squire, 2003). Finally, there are no auditory bias effects on perceived duration of audiovisual intervals when the auditory stimuli are intra- rather than intermodally grouped (Asaoka & Gyoba, 2016; Klink et al., 2011). Taken together, all these findings indicate that the biasing effect of conflicting auditory interval duration reflects a genuine perceptual integration rather than a simple decisional strategy of responding to the auditory information alone.

Considering the wider context of multisensory integration in the temporal domain, one remaining question is why there are sustained effects in perception of temporal order (Fujisaki et al., 2004; Keetels et al., 2007), but apparently not in duration perception? As outlined in the Introduction, a similar mechanism, that is, temporal ventriloquism, might be responsible for integration of interval duration as well as temporal order (Asaoka & Gyoba, 2016; De la Rosa & Bausenhart, 2013). In temporal order recalibration, participants adapt to a consistent audiovisual conflict between two brief stimuli (e.g., auditory always before visual). For interval duration, there is also a consistent audiovisual conflict (e.g., auditory always longer than visual), but the temporal markers of the interval (i.e., onset and offset) are biased in opposing directions. For example, for longer auditory intervals, the auditory onset is before the visual onset, but the auditory offset is after the visual offset. Thus, multisensory integration might evoke immediate effects on perceived duration, but since the audiovisual conflicts associated with the intervals' on- and the offsets oppose and thus compensate each other, this would prevent sustained effects as temporal order recalibration.

An interesting future research avenue would be to investigate whether the present conclusion regarding absence of sustained multisensory integration effects would still hold, if attention was allocated to the temporal properties of the adapting stimulation. In the present experiments, an additional task was administered to ensure that participants do not ignore the visual input and moreover, attend to the location of the visual adapting stimulation. However, participants were not required to explicitly attend to the duration of the auditory, visual or combined input, and consequently neither to potential conflicts between the modalities' contributions. Even though there is evidence that visual temporal information is automatically processed even in the absence of attention to temporal features (Y. Chen, Huang, Luo, Peng, & Liu, 2010; Heron et al., 2010), allocating attention to temporal features may further enhance the processing of temporal information. Consequently, one might hypothesize that sustained effects of multisensory duration integration might become evident when participants actively process the audiovisual temporal properties of the stimulation during the adaptation phase (even though, as briefly outlined above, awareness of an audiovisual conflict might likewise provide a violation of the unity assumption and thus hinder any multisensory integration after all).

The present data reveal another interesting multimodal effect on reproduced duration, which concerns the difference between reproductions of the "same" bias conditions between Experiments 1 and 2. When comparing Figures 2 and 3, it is evident that the slope of the reproduction function (especially between the 500 and 753 ms interval durations) is much steeper for the reproduction of audiovisual congruent intervals ("same" condition of Exp. 2) than for the reproduction of unimodal visual intervals ("same" condition of Exp. 1). This was confirmed by a significant interaction, $F(2, 68) = 9.56$, $p < .001$, $\eta^2_p = .22$ in additional ANOVA conducted on reproduced duration with interval duration as within- and bias condition (uni- vs. bimodal) as between-subjects factor. This finding replicates the well-known effect of interval modality on duration perception, namely, that auditory (or in this case, audiovisual) perceived duration is perceived as longer than unimodal visual duration, and this effect increases with increasing interval duration (e.g., K. M. Chen & Yeh, 2009; De la Rosa & Bausenhart, 2013; Penton-Voak et al., 1996; Walker & Scott, 1981; Wearden, Edwards, Fakhri, & Percival, 1998). This effect has been attributed to an increase of the pacemaker rate of a pacemaker-accumulator mechanism of duration perception (Wearden et al., 1998), as a consequence of increased arousal caused by the auditory stimulation (Wearden, Norton, Martin, & Montford-Bebb, 2007). The present results demonstrate such an increase even under conditions of repeated preceding auditory stimulation. That indicates that the sound-based increase in pulse rate is (at least partly) rather short-lived and does not spill over from immediately preceding sound presentations (e.g., through the audiovisual top-up pairs) to a unimodal test stimulus. Interestingly, this finding contrasts with related results showing prolonged perceived interval duration caused by immediately preceding click trains, which may be due to the higher temporal frequency of the click trains (Penton-Voak et al., 1996; Wearden et al., 1998).

In conclusion, we demonstrated that multisensory integration of audiovisual intervals of conflicting duration leads to pronounced immediate effects on perceived duration, but does not affect perceived duration in a sustained manner. This finding replicates and extends previous studies (Bausenhart et al., 2014; Heron et al., 2013) on multisensory interval integration. Thus, the present findings once more demonstrate that temporal judgments can be quite sensitive to the experimental context (e.g., concurrent incongruent or congruent multimodal stimulation), but at the same time also purport an important limitation of this context-sensitivity (that is, preceding multimodal stimulation does not affect currently perceived duration).

Acknowledgement

We thank Julia Feyerabend and Laura Schneider for assistance in data collection. This research was supported by the German Research Foundation (DFG grant 4110/3-2).

References

Asaoka, R., & Gyoba, J. (2016). Sounds modulate the perceived duration of visual stimuli via crossmodal integration. *Multisensory Research*, *29*(4–5), 319–335. https://doi.org/10.1163/22134808-00002518

Bausenhart, K. M., Bratzke, D., & Ulrich, R. (2016). Formation and representation of temporal reference information. *Current Opinion in Behavioral Sciences*, *8*, 46–52. https://doi.org/10.1016/j.cobeha.2016.01.007

Bausenhart, K. M., De la Rosa, M. D., & Ulrich, R. (2014). Multimodal integration of time: Visual and auditory contributions to perceived duration and sensitivity. *Experimental Psychology*, *61*(4), 310–322. https://doi.org/10.1027/1618-3169/a000249

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension: 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, *50*(1), 147–155. https://doi.org/10.1016/S0167-8760(03)00130-2

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Chen, K. M., & Yeh, S. L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychologica*, *130*(3), 225–234. https://doi.org/10.1016/j.actpsy.2008.12.008

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

Chen, Y., Huang, X., Luo, Y., Peng, C., & Liu, C. (2010). Differences in the neural basis of automatic auditory and visual time perception: ERP evidence from an across-modal delayed response oddball task. *Brain Research*, *1325*, 100–111. https://doi.org/10.1016/j.brainres.2010.02.040

De la Rosa, M. D., & Bausenhart, K. M. (2013). Multimodal integration of interval duration:

Temporal ventriloquism or changes in pacemaker rate? *Timing and Time Perception*, *1*(2), 189–215. https://doi.org/10.1163/22134468-00002015

Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, *9*(12), 7–7. https://doi.org/10.1167/9.12.7

Dyjas, O., Bausenhart, K. M., & Ulrich, R. (2014). Effects of stimulus order on duration discrimination sensitivity are under attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 292–307. https://doi.org/10.1037/a0033611

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

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778. https://doi.org/10.1167/3.9.34

Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception*, *36*(7), 1089–1103. https://doi.org/10.1068/p5741

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325. https://doi.org/10.1037/0033-295X.84.3.279

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, *423*(1), 52–77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x

Grondin, S., Ivry, R. B., Franz, E., Perreault, L., & Metthé, L. (1996). Markers' influence on the duration discrimination of intermodal intervals. *Perception & Psychophysics*, *58*(3), 424–433. https://doi.org/10.3758/BF03206818

Hanson, J. V. M., Heron, J., & Whitaker, D. (2008). Recalibration of perceived time across sensory modalities. *Experimental Brain Research*, *185*, 347–352. https://doi.org/10.1007/s00221-008-1282-3

Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research*, *186*(4), 517–524. https://doi.org/10.1007/s00221-007-1253-0

Hartcher-O'Brien, J., Di Luca, M., & Ernst, M. O. (2014). The duration of uncertain times: Audiovisual information about intervals is integrated in a statistically optimal fashion. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0089339

Heron, J., Aaen-Stockdale, C., Hotchkiss, J., Roach, N. W., McGraw, P. V., & Whitaker, D. (2012). Duration channels mediate human time perception. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 690–698. https://doi.org/10.1098/rspb.2011.1131

Heron, J., Hotchkiss, J., Aaen-stockdale, C., Roach, N. W., & Whitaker, D. (2013). A neural hierarchy for illusions of time: Duration adaptation precedes multisensory integration. *Journal of Vision*, *13*(14), 1–12. https://doi.org/10.1167/13.14.4.doi

Heron, J., Roach, N. W., Whitaker, D., & Hanson, J. V. M. (2010). Attention regulates the plasticity of multisensory timing. *European Journal of Neuroscience*, *31*(10), 1755–1762. https://doi.org/10.1111/j.1460-9568.2010.07194.x

Heron, J., Whitaker, D., McGraw, P. V., & Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision*, *7*(13), 5.1-8. https://doi.org/10.1167/7.13.5

Keetels, M., Stekelenburg, J., & Vroomen, J. (2007). Auditory grouping occurs prior to intersensory pairing: Evidence from temporal ventriloquism. *Experimental Brain Research*, *180*(3), 449–456. https://doi.org/10.1007/s00221-007-0881-8

Keetels, M., & Vroomen, J. (2007). No effect of auditory-visual spatial disparity on temporal recalibration. *Experimental Brain Research*, *182*(4), 559–565. https://doi.org/10.1007/s00221-007-1012-2

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(14), 1. https://doi.org/10.1068/v070821

Klink, P. C., Montijn, J. S., & van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention,* *Perception & Psychophysics*, *73*(1), 219–236. https://doi.org/10.3758/s13414-010-0010-9

Machulla, T. K., Di Luca, M., Froehlich, E., & Ernst, M. O. (2012). Multisensory simultaneity recalibration: Storage of the aftereffect in the absence of counterevidence. *Experimental Brain Research*, *217*(1), 89–97. https://doi.org/10.1007/s00221-011-2976-5

Mahani, M. A. N., Sheybani, S., Bausenhart, K. M., Ulrich, R., & Ahmadabadi, M. N. (2017). Multisensory perception of contradictory information in an environment of varying reliability: Evidence for conscious perception and optimal causal inference. *Scientific Reports*, *7*. https://doi.org/DOI:10.1038/s41598-017-03521-2

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, *7*(10), 3215–29. https://doi.org/citeulike-article-id:409430

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

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. https://doi.org/10.1163/156856897X00366

Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*(3), 307–320. https://doi.org/10.1037/0097-7403.22.3.307

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

Romei, V., De Haas, B., Mok, R. M., & Driver, J. (2011). Auditory stimulus timing influences perceived duration of co-occurring visual stimuli. *Frontiers in Psychology*, *2*, 1–8. https://doi.org/10.3389/fpsyg.2011.00215

Rousseau, R., Poirier, J., & Lemyre, L. (1983). Duration discrimination of empty time intervals marked by intermodal pulses. *Perception & Psychophysics*, *34*(6), 541–548. https://doi.org/10.3758/BF03205907

Sarmiento, B. R., Shore, D. I., Milliken, B., & Sanabria, D. (2012). Audiovisual interactions depend on context of congruency. *Attention, Perception, & Psychophysics*, *74*(3), 563–574. https://doi.org/10.3758/s13414-011-0249-9

Shi, Z., Chen, L., & Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: The influence of time interval. *Experimental Brain Research*, *203*(4), 723–735. https://doi.org/10.1007/s00221-010-2286-3

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, *13*(3), 519–521. https://doi.org/10.1016/S0960-9822(03)00445-7

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Crossmodal temporal discrimination: Assessing the predictions of a general pacemaker-counter model. *Perception & Psychophysics*, *68*(7), 1140–1152.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*, *181*(1), 173–181. https://doi.org/10.1007/s00221-007-0918-z

Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 513–518. https://doi.org/10.1037/0096-1523.30.3.513

Vroomen, J., & Keetels, M. (2009). Sounds change four-dot masking. *Acta Psychologica*, *130*(1), 58–63. https://doi.org/10.1016/j.actpsy.2008.10.001

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

Walker, J. T., Irion, A. L., & Gordon, D. G. (1981). Simple and contingent aftereffects of perceived duration in vision and audition. *Attention, Perception, & Psychophysics*, *29*(5), 475–486. https://doi.org/10.3758/BF03207361

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(6), 1327–1339. https://doi.org/10.1037/0096-1523.7.6.1327

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why "sounds are judged longer than lights": Application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology*, *51B*(2), 97–120. https://doi.org/10.1080/713932672

Wearden, J. H., Norton, R., Martin, S., & Montford-Bebb, O. (2007). Internal clock processes and the filled-duration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 716–729. https://doi.org/10.1037/0096-1523.33.3.716

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron*, *48*(3), 489–496. https://doi.org/10.1016/j.neuron.2005.10.020

Zaidel, A., Ma, W., & Angelaki, D. E. (2013). Supervised calibration relies on the multisensory percept. *Neuron*, *80*(6), 1544–1557. https://doi.org/10.1016/j.neuron.2013.09.026

Zaidel, A., Turner, A. H., & Angelaki, D. E. (2011). Multisensory calibration is independent of cue reliability. *Journal of Neuroscience*, *31*(39), 13949–13962. https://doi.org/10.1523/JNEUROSCI.2732-11.2011