Acta Psychologica 190 (2018) 38–52

Contents lists available at ScienceDirect

Acta Psychologica

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

Enhancement of letter identification by concurrent auditory stimuli of varying duration

Maria D. de la Rosa*, Karin M. Bausenhart

University of Tübingen, Germany

ARTICLE INFO

Keywords:
Backwards-masking paradigm
Letter identification
Multisensory integration
Duration perception

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.

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 suprathreshold visual stimuli such as Gabor patches, lines, or simple shapes, defined by elementary features as luminance, contrast, color, or orientation (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, Chen and 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 cross-modal facilitation depended on a reliable temporal coincidence of the target letter and sound (Experiment 4) but, otherwise, their spatial

* Corresponding author at: Fachbereich Psychologie, University of Tübingen, Schleichstrasse 4, D-72076 Tübingen, Germany.
E-mail address: maria-dolores.de-la-rosa-gamiz@uni-tuebingen.de (M.D. de la Rosa).

https://doi.org/10.1016/j.actpsy.2018.07.001
Received 23 March 2018; Received in revised form 25 June 2018; Accepted 2 July 2018
Available online 10 July 2018
0001-6918/

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" (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, Chen and 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; Chen & Yeh, 2009; De la Rosa & Bausenhart, 2013; Hartcher-O'Brien, Di Luca, & Ernst, 2014; Klink, Montijn, & 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 Chen and 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 Chen and 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 Chen and 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).

1. 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