



# Cognitive control in the cocktail party: Preparing selective attention to dichotically presented voices supports distractor suppression

Sophie Nolden<sup>1,2</sup> · Camellia N. Ibrahim<sup>2</sup> · Iring Koch<sup>2</sup>

Published online: 19 December 2018  
© The Psychonomic Society, Inc. 2018

## Abstract

The goal of the present study was to investigate preparatory mechanisms of auditory selective attention. In two experiments, participants performed a classification task on one of two dichotically presented spoken number words, one spoken by a female, one spoken by a male. A cue indicated which gender participants had to attend to in the upcoming trial, so that attention switches and repetitions occurred randomly. The cue-target interval (CTI) was either 400 ms or 1,200 ms. Stimulus onset asynchrony (SOA) between target and distractor word varied; hence, the distractor could be presented before or after the target. In two experiments, we found robust performance costs of attention switches. Like in previous studies using versions of this paradigm, these switch costs were not significantly reduced by prolonged CTI, even though we found substantial general cue-based preparation effects. The most important finding refers to the influence of SOA, showing that the general preparation effect was greater in the condition with the distractor presented first than in the condition with the target presented first. Thus, increased time to prepare for the attention focus of the upcoming trials seems to benefit distractor suppression more than target enhancement. This occurred in switch trials and repetition trials alike, suggesting that it is a general feature of auditory attention.

**Keywords** Auditory selective attention · Cognitive control · Task-switching · Preparation · Dichotic listening

## Introduction

Selecting the most relevant of simultaneously available sources of acoustic information requires attentional control of auditory selective attention. This is often the case when we are surrounded by multiple speakers and we want to listen to only one of them. We then have to select a particular speaker and/or ignore the irrelevant speakers (for reviews, see Schneider, Li, & Daneman, 2007; Shinn-Cunningham, 2008). The dichotic-listening paradigm has been used extensively to investigate mechanisms of auditory selective attention in these situations (see, for example, Broadbent, 1953; Cherry, 1953; for reviews, see Bronkhorst, 2015; Hugdahl, 2011).

In the dichotic-listening paradigm, two auditory stimuli are presented simultaneously, one to each ear. Participants have to attend to one of them in order to do a certain task, for example in order to repeat what they have heard of the attended message (“shadowing”). Early studies focused on maintenance of auditory selective attention to one of the messages and on the information that participants reported of the irrelevant message. For example, in a study by Cherry (1953), participants listened to dichotically presented speech and shadowed one of the messages. Afterwards, they were asked what they remembered of the unattended message. They were able to identify the acoustic information from the irrelevant message as speech and noticed changes such as the gender of the speaker, without being able to recall much of the actual content. These results have been interpreted as evidence for an early filter mechanism in auditory attention. Results from the dichotic-listening paradigm such as the one from Cherry (1953) have had a great impact on the development of filter theories of attention (e.g., Broadbent, 1958; see also Deutsch & Deutsch, 1963; Treisman, 1960, 1969). There was, however, one important methodological challenge when investigating processing of to-be-unattended information while maintaining auditory attention to a different stream. Participants might have briefly shifted attention from one message to the other.

---

✉ Sophie Nolden  
nolden@psych.uni-frankfurt.de

<sup>1</sup> Department of Psychology, Goethe-University Frankfurt am Main, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany

<sup>2</sup> Institute of Psychology, RWTH Aachen University, Aachen, Germany

It is thus unclear if information from the irrelevant message was not attended to at all, or if an involuntary attention shift took place, which made the interpretation of certain effects difficult (see Lachter, Forster, & Ruthruff, 2004, for a review).

More recently, the dichotic-listening paradigm has been used to investigate another important aspect of auditory selective attention: voluntary, intentional shifting between different acoustic sources (e.g., Koch, Lawo, Feld, & Vorländer, 2011). Thereby, the focus was not on sustained attention and involuntary attention capture, but on the flexibility of allocating auditory attention between different messages. The methodology was adapted from the task-switching literature (see Kiesel et al., 2010; Koch, Gade, Schuch, & Philipp, 2010; Koch, Poljac, Müller & Kiesel, 2018; Monsell, 2003, for reviews).

In the study by Koch et al. (2011), two number words were presented dichotically. One of them was spoken by a female speaker, the other one by a male speaker. A visual cue that was presented before the number words indicated if participants had to attend to the female or the male speaker in order to perform a number classification task (i.e., is the target number smaller or greater than 5) on the attended number word (Meiran, 1996; for reviews see, e.g., Jost, DeBaene, Koch, & Brass, 2013; Kiesel et al., 2010). There could be so-called auditory attention repetition trials in which the gender of the to-be-attended speaker remained constant, for example when participants had to attend to a female voice in two subsequent trials, or so-called auditory attention switch trials in which the gender of the to-be-attended speaker could change, for example when participants had to attend to a female voice after having attended to a male voice in the previous trial. Participants responded more slowly in switch trials than in repetition trials, they thus showed so-called switch costs. This has been interpreted as an effect of resolving additional interference that occurred in switch trials compared to repetition trials. There are methodological similarities between this approach to study auditory attention switches and in the task-switching literature. Note that the task in the study by Koch and collaborators (2011), the classification task actually remained constant while the attention focus varied. The main focus of the study was thus on *auditory attention shifting*.

The paradigm used by Koch et al. (2011) also allowed investigating congruency effects. Target and the distractor number could belong to the same numerical category, which is smaller or greater than 5, or belong to a different category. In the first case, target and distractor were mapped to the same response (congruent trials), and in the second case, target and distractor were mapped to different responses (incongruent trials). Performance impairments in incongruent trials compared to congruent trials are called congruency effects. In the study by Koch et al. (2011), congruency effects were quite pronounced in the errors rates. In incorrect incongruent trials,

it is possible that participants actually responded to the irrelevant stimuli instead of selecting and responding to the relevant stimulus. The fact that the congruency effects were rather pronounced in the error rates and less evident in the response times (RTs) further corroborates this notion. Once the irrelevant stimulus was incorrectly attended to, participants did not seem to correct for this error, as RTs were much less affected by congruency (see Nolden & Koch, 2017, for further discussion).

In order to explore if shifting auditory attention between stimuli can be prepared before the onset of the stimuli, the interval between the visual cue and the spoken number of words was varied in previous studies (Koch et al., 2011, Experiments 2 and 3; see also Lawo & Koch, 2015). We refer to this interval as the cue-target interval (CTI). Variations of the CTI aim to elucidate the intentional shifting of the attention focus and are therefore informative about cognitive control of auditory attention. If adjustments of the attention settings took place before the presentation of the auditory stimuli, increased time between the cue and the target should result in better performance, especially in switch trials. However, while participants responded generally faster in the long CTI condition than in the short CTI condition, specific effects of CTI on switch costs were not consistently found. The authors suggested that switch costs were thus largely attributable to “attention inertia,” and that active processes of attention shifting would rather take place after the onset of the acoustic stimuli. In addition, CTI had no effect on the congruency effects in the error rates, suggesting that participants did not attend to the relevant stimulus more successfully when they had more time to prepare for the next trial. This again supports the notion of attention inertia. The fact that participants did not seem to prepare actively for shifting auditory attention might be to some degree modality-specific (Koch et al., 2011; Koch & Lawo, 2014; Lawo, Fels, Oberem, & Koch, 2014; Lawo & Koch, 2014, 2015; but see Seibold, Nolden, Oberem, Fels, & Koch, 2018, for predictable switches). In contrast, in visual tasks, task-switching costs can be reduced with increased CTI (e.g., Meiran, Chorev, & Sapir, 2000; see Kiesel et al., 2010; Koch et al., 2018, for reviews).

While a specific effect of CTI on switch costs had not been systematically found, there are clear and substantial general preparation benefits of CTI on RTs. For example, Koch et al. (2011) found in two experiments that participants responded much faster when the CTI was 1,000 ms than when it was 100 ms (RTs of 1,112 ms vs. 1,182 ms in Experiment 1, RTs of 1,063 ms vs. 1,203 ms). Since this effect is not specific to preparing an auditory attention switch, it may be due to general preparatory mechanisms. Specifically, attending to an auditory stimulus while ignoring the distractor requires in general two important processes (Bronkhorst, 2015).

First, acoustic information must be segregated and grouped in order to represent auditory objects (see Bregman, 1990).

Increased preparation time could most likely not be efficiently used for these processes since there was always one out of three female and one out of three male speakers in every trial, and there was no specific information about the individual voices available before their onset. Secondly, participants needed to direct auditory attention to one of the speakers. Here, the preparatory interval could have been used efficiently, because the relevant gender varied and participants knew before the onset of the number words which gender had to be attended to. We assume that the auditory stimuli are represented gradually, i.e., not filtered in an all-or-none fashion (see, e.g., Meiran, Kessler, & Adi-Japha, 2008), even though there might be a strong bias in favor for the attended stimulus. It is unclear if enhancing the target representation, or suppressing the distractor representation, or both, are related to this process. Recent studies using visual stimuli aimed to disentangle the role of target enhancement and distractor suppression (see Noonan, Adamian, Pike, Printzlau, Crittenden, & Stokes, 2016, for visual stimuli; Samson & Johnsrude, 2016, for auditory stimuli).

In the present study, we investigated if increased preparation time before the presentation of an auditory target and an auditory distractor had a systematic influence on target enhancement or distractor suppression. We therefore modified the experimental paradigm developed by Koch and collaborators (2011). In addition to using a short and a long CTI (manipulation of preparation time), we introduced variable delays between target and distractor words to disentangle the effect of CTI on target and distractor processing (see Holmes, Kitterick, & Summerfield, 2018). As the stimulus that is presented first is also the one that enters the perceptual stream first, effects in the target first condition would predominantly reflect target processing, and effects in the distractor first condition would predominantly reflect distractor processing. With the simultaneous condition as the baseline, we investigated if target enhancement and distractor suppression could be prepared before the onset of the target stimulus. If preparation time improved target enhancement, there should be an additional preparation benefit when the target is presented first, thus shorter RTs in the target-first condition after a long CTI than after a short CTI, compared to the simultaneous presentation of target and distractor. If preparation time enhances distractor suppression, there should be an additional preparation benefit when the distractor is presented first, thus shorter RTs in the distractor-first condition after a long CTI than after a short CTI, compared to the simultaneous presentation of target and distractor. We thus aimed to reveal specific effects of CTI on target and distractor processing in order to better understand the cognitive mechanisms underlying a general preparation benefit on RTs.

In addition, we investigated congruency effects as a measure for involuntary attention capture. Whereas the SOA manipulation aimed at specific effects of target and distractor

processing such as target enhancement or distractor suppression (once the relevant stimulus was attended to), congruency effects, especially in the error rates, reflect the capacity to direct attention to the relevant stimulus. We thus investigated preparatory aspects related to two different variables related to target and distractor processing highlighting specific aspects. These were, on the one hand, directing attention to the to-be-attended stimulus (congruency effects) and, on the other hand, processing of target and distractor when attention is directed to the target (SOA effects). We conducted two experiments: In Experiment 1, we used three SOA levels including a simultaneous presentation of target and distractor, and in Experiment 2, we used a simplified version with two SOA levels.

## Experiment 1

In Experiment 1, we used a variation of the auditory attention-shifting paradigm (Koch et al., 2011). We added delays between target and distractor onset, such that the target either preceded the distractor, or the distractor preceded the target. The simultaneous presentation of target and distractor was used as a baseline.

## Method

**Participants** Twenty-five participants participated in Experiment 1. One participant who indicated after a quarter of the experiment that he had forgotten the instructions was replaced by a new participant. The remaining 24 participants had a mean age of 24 years ( $SD = 4$  years, range: 19–34 years), 16 were female, and 18 were right-handed. None of them reported any hearing problems. All participants gave informed consent and participated voluntarily.

**Stimuli and task** Visual stimuli were presented at the center of a 17-in. monitor with white background. The participants' distance to the screen was about 60 cm. Visual stimuli were a black fixation cross and red feedback words (“Fehler!”, German for “error,” or “Schneller!”, German for “faster”). The fixation cross was presented at all times, except when feedback or instructions were presented.

Auditory stimuli were cue tones and auditory target and distractor words. Cue tones cued the to-be-attended gender, such that a high-pitch tone (800 Hz) cued the female speaker and a low-pitch tone (200 Hz) cued the male speaker. Cue tones were 50 ms in duration with onset and offset ramps of 5-ms each. Cue tones consisted of three harmonics with decreasing intensity ( $1/\text{number of harmonic}$ ). Target and distractor words were eight spoken digits (1–9 without 5). Three female speakers and three male speakers were recorded in an anechoic chamber at the Institute of Technical Acoustics of RWTH Aachen University. A subjective loudness

calibration was carried out for each individual digit and for all different speakers. The duration for all different speakers was adjusted across the set of eight digits to be subjectively the same, whereby a time-stretching algorithm was used to shorten the long samples while maintaining pitch. Duration was about 700 ms for each spoken digit (see Koch et al., 2011). Auditory stimuli were presented dichotically via headphones (Grundig 38629 DJ Headphones). All stimuli were presented with E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA, USA).

Participants had to attend to only one of the two dichotically presented stimuli. They were asked to indicate if the target digit (with relevant speaker gender indicated by the prior cue) was smaller than 5 (i.e., 1, 2, 3, or 4) by pressing “y” (left index finger). They were asked to indicate if the target digit was greater than 5 (i.e., 6, 7, 8, or 9) by pressing “.” (right index finger) on the computer keyboard (QWERTZ). This mapping was constant and the same for all participants. The irrelevant stimulus had to be ignored.

**Procedure** Each trial started with the cue tone (duration: 50 ms). After a cue-target interval (CTI, i.e., the time between the onset of the cue tone and the onset of the target) of either 400 ms or 1,200 ms, the target digit, spoken by the to-be-attended speaker, was presented to one ear. The distractor digit, spoken by the to-be-ignored speaker, was presented to the other ear. The mapping of ear and speaker sex varied randomly from trial to trial, so that the gender cue did not convey advance information about the spatial location of the target. One of the three speakers per speaker sex was chosen randomly in each trial. The stimulus onset asynchrony (SOA) between target and distractor digit was one of -200 ms, 0 ms, or 200 ms, hence the distractor onset could be before, simultaneously, or after the target onset. SOA varied randomly within blocks. Participants had maximally 3,800 ms from target onset to indicate if the target digit was smaller or greater than 5. In case of an error, the word “Fehler!” (German for “error”) was displayed in red color on the center of the screen for 500 ms. In case of no response after 3,800 ms, the word “Schneller!” (German for “faster”) was displayed in red color on the center of the screen for 500 ms. After a silent interval (response-cue interval, RCI) of either 1,200 ms, if the CTI was 400 ms, or after an RCI of 400 ms, if the CTI was 1,200 ms, the next trial started (see Fig. 1). CTI and RCI thus varied inversely. This is a common procedure in task-switching research, from which we borrowed our methodology (e.g., Kiesel et al., 2010), and allows us to study preparation effects independently from passive dissipation of the attention setting of the previous trial (which would be indicated by effects of variations of total trial duration [response-target interval] if CTI is kept constant; see Koch & Lawo, 2014).

Participants completed four experimental blocks of 144 trials each. CTI was varied block-wise, whereby half of the

participants started with a CTI of 400 ms, and the other half started with a CTI of 1,200 ms. In sum, the relevant speaker sex, the ear and speaker sex mapping, SOA, congruency, and the actual speaking voice varied randomly within a block, and CTI varied between blocks. Before the experimental blocks, participants completed two practice blocks of 24 trials each. In one practice block, the CTI was 400 ms, in the other practice block, the CTI was 1,200 ms. Practice block order was counterbalanced over participants.

Participants reported demographic data before the experiment and were asked about strategies after the experiment. Participants were instructed orally and with written instructions on the computer screen. The total experiment lasted about 40 min.

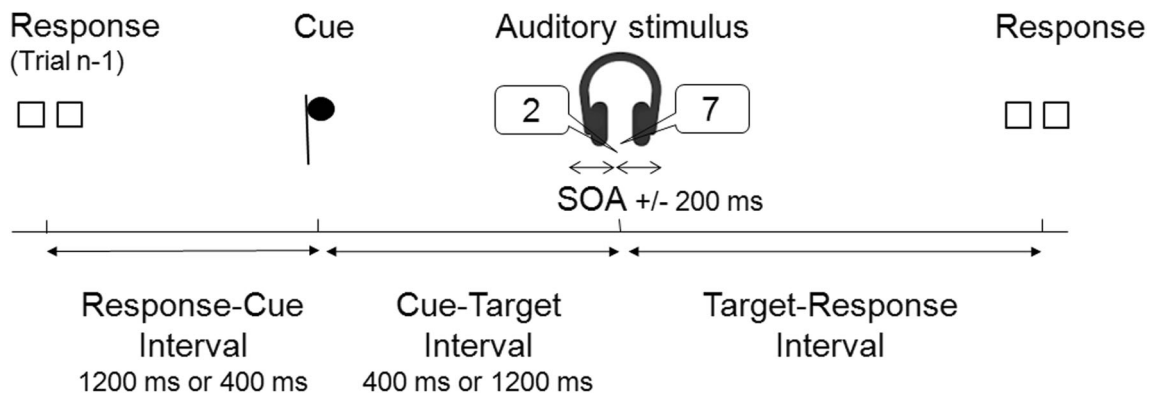
**Design** Independent variables were attention transition (repetition, switch), CTI (400 ms, 1,200 ms), SOA (-200 ms, 0 ms, 200 ms), and congruency (congruent, incongruent). Dependent variables were RTs, measured from target onset, and error rate.

## Results and discussion

Practice trials, the first trial of each block, and error trials were excluded from the analysis of the RTs, as well as outliers ( $RT \pm 3$  SD from the mean of each condition) and responses after the end of the response window. Practice trials and the first trial of each block were excluded from the analysis of the error rates.

**Response times** We conducted a  $2 \times 2 \times 3 \times 2$  ANOVA with the within-subject variables attention transition (repetition, switch), CTI (400 ms, 1,200 ms), SOA (-200 ms, 0 ms, 200 ms) and congruency (congruent, incongruent) on RTs (see Fig. 2). When sphericity was violated, the Huynh-Feldt correction was applied. The ANOVA revealed a main effect of transition,  $F(1, 23) = 22.43, p < .001, \eta_p^2 = .49$ , indicating that participants responded faster in repetition trials than in switch trials, 1,038 ms versus 1,088 ms, thus attention switch costs of 50 ms.

There was a significant main effect of CTI,  $F(1, 23) = 16.66, p < .001, \eta_p^2 = .42$ , indicating that participants responded faster in the CTI 1,200-ms condition than in the CTI 400-ms condition, 1,017 ms versus 1,109 ms, thus a general preparation benefit of 92 ms. There was also a significant main effect of SOA,  $F(2, 46) = 23.64, p < .001, \eta_p^2 = .51$ , indicating that participants responded more slowly when target and distractor were presented simultaneously than when either the target was presented before the distractor, 1,112 ms versus 1,032 ms,  $t(23) = -7.60, p < .001$ , or when the distractor was presented before the target, 1,112 ms versus 1,044 ms,  $t(23) = -5.64, p < .001$ . There was no significant difference in SOA -200 ms and SOA 200 ms,  $t(23) = -0.83, p > .41$ . It was



**Fig. 1** Trial Procedure. A high pitch or low pitch tone instructed participants to either attend to the female or the male voice. After a cue-stimulus interval of 400 ms or 1200 ms, the target number was presented. The distractor number was presented simultaneously with the target number (only in Experiment 1), or 200 ms before or after the target

number (Experiment 1 and Experiment 2). Participants indicated if the attended number was smaller or greater than 5 by pressing one of two buttons. After a response-cue interval of 1200 ms or 400 ms, the cue of the next trial was presented

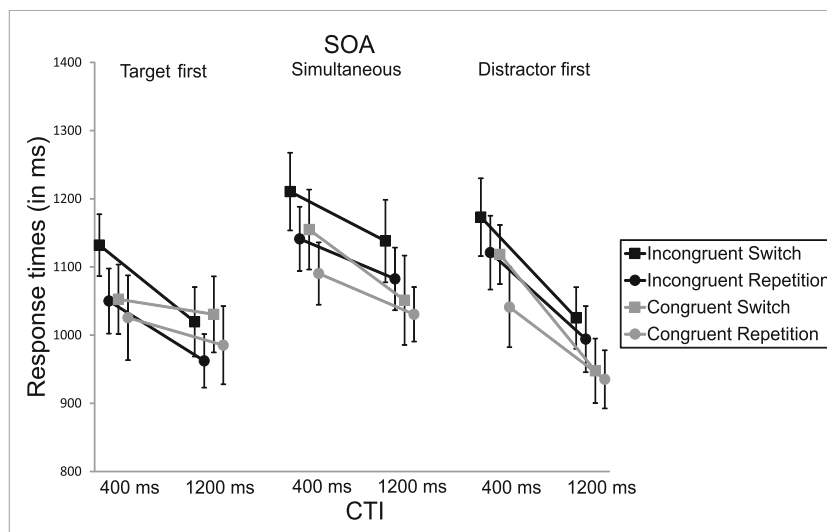
thus advantageous if target and distractor onset varied, which might be related to more effective mutual masking with simultaneous onset of target and distractor. Stream segregation may be facilitated when the onset of two stimuli is temporally separated (Bregman, 1990; Darwin & Carlyon, 1995). The main effect of SOA was further qualified by a significant interaction of CTI and SOA,  $F(2, 46) = 6.19, p < .01, \eta_p^2 = .21$ .

To decompose the interaction of CTI and SOA, we calculated the CTI effect for each SOA condition and conducted three *t*-tests to compare, pairwise, the CTI effect of the three SOA conditions (see Fig. 3). Whereas there was no significant difference in the CTI effects of SOA 0 ms and SOA 200 ms (73 ms vs. 66 ms),  $t(23) = -0.36, p > .72$ , the CTI effect of SOA -200 ms was greater than the CTI effect of SOA 0 ms (137 ms vs. 73 ms),  $t(23) = -2.91, p < .01$ , and also greater than the CTI effect of SOA

200 ms (137 ms vs. 66 ms),  $t(23) = -3.17, p < .01$ . Participants thus benefited most from prolonged preparation time when the distractor was presented before the target, suggesting that preparation was most successful for distractor processing.

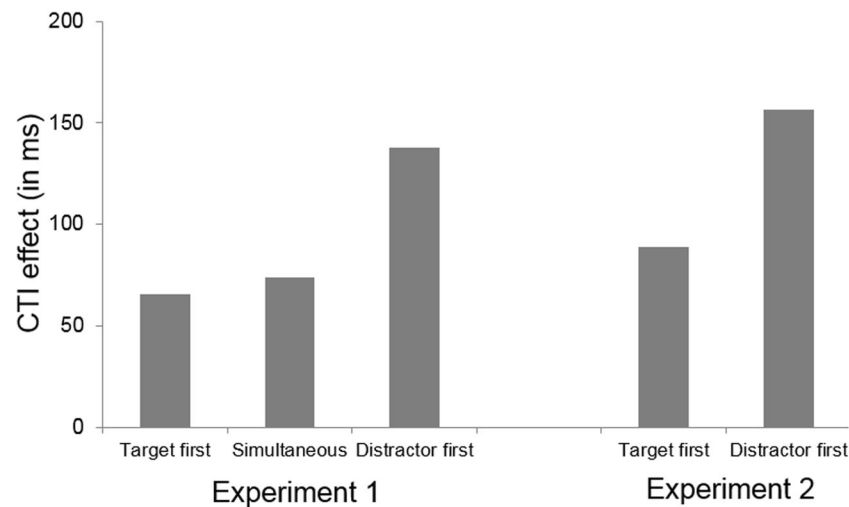
The main effect of congruency was also significant,  $F(1, 23) = 14.63, p < .001, \eta_p^2 = .39$ . This indicated that participants responded faster in congruent trials than in incongruent trials, 1,039 ms versus 1,087 ms, thus there was a congruency effect of 48 ms.

All other effects were not significant, interaction of transition and CTI:  $F(1, 23) = 1.92, p > .17, \eta_p^2 = .08$ ; interaction of transition and congruency:  $F(1, 23) = 1.14, p > .29, \eta_p^2 = .05$ ; interaction of SOA and congruency:  $F(2, 46) = 2.86, p > .06, \eta_p^2 = .11$ ; interaction of CTI, SOA, and congruency:  $F(2, 46) = 1.47, p > .24, \eta_p^2 = .06$ . All other *F*s < 1.



**Fig. 2** Experiment 1. Response times. Participants responded faster in repetition than in switch trials. Response times were faster when the CTI was 1200 ms than when it was 400 ms. This decrease in response

times with increased CTI was greatest when the distractor was presented first. The error bars represent confidence intervals (Cousineau 2005; Morey, 2008)



**Fig. 3** Preparation effects. Difference of response times between the CTI 1200 ms and CTI 400 ms for all three SOA conditions of both experiments

**Errors** We conducted the same ANOVA on error rates (see Fig. 4). The ANOVA revealed a main effect of transition,  $F(1, 23) = 21.28, p < .001, \eta_p^2 = .48$ , corroborating the results from the response times. The main effect of congruency was also significant,  $F(1, 23) = 27.38, p < .001, \eta_p^2 = .54$ , indicating that participants made less errors in congruent trials than in incongruent trials, 2.9% versus 8.4%, thus a general congruency effect of 5.5%. This effect was further modulated by a significant interaction of transition and congruency,  $F(1, 23) = 13.51, p < .01, \eta_p^2 = .37$ , indicating that the congruency effect was smaller in repetition trials than in switch trials, 4.0% versus 7.0%.

In addition, there was a significant interaction of SOA and congruency,  $F(2, 46) = 5.78, p < .01, \eta_p^2 = .20$ . To decompose this interaction, we calculated the congruency effect for each SOA condition and conducted three *t*-tests to compare the congruency effect of the three SOA conditions. Whereas there was no significant difference in the congruency effects of SOA 0 ms and SOA -200 ms (6.8% vs. 6.2%),  $t(23) = 0.58, p > .56$ , the congruency effect of SOA 200 ms was smaller than the congruency effect of SOA 0 ms (3.4% vs. 6.8%),  $t(23) = -2.91, p < .01$ , and also smaller than the congruency effect of SOA -200 ms (3.4% vs. 6.2%),  $t(23) = -2.79, p < .02$ . Congruency effects were thus smallest when the target was presented first (see Fig. 5). In general, congruency effects in the error rates may be related to participants selecting the wrong stimulus (see Nolden & Koch, 2017, for further discussion).

All other effects were not significant, main effect of CTI:  $F(1, 23) = 3.29, p > .08, \eta_p^2 = .13$ ; main effect of SOA:  $F(1.5, 34.56) = 3.48, p > .05, \eta_p^2 = .13$ ; interaction of transition and CTI:  $F(1, 23) = 1.19, p > .28, \eta_p^2 = .05$ ; interaction of CTI, SOA, and congruency:  $F(1.69, 38.91) = 1.58, p > .22, \eta_p^2 = .06$ . All other  $F$ s < 1.

To summarize, preparation effects in RTs are strongest if the distractor is presented first, suggesting that preparation helps to attenuate the distractor influence on the response to

the target. In ER, this effect was not significant, suggesting that preparation relates more to speed than to accuracy.

In comparison, in the error rates, there is no consistent benefit of preparation, but SOA still has an effect, which is mainly focused on the congruency effect. The congruency effect is smaller if the target precedes the distractor, suggesting that a head start of target processing generally reduces the impact of distractor-based interference. Notably, this general effect was independent of the CTI variation and thus did not depend on preparation.

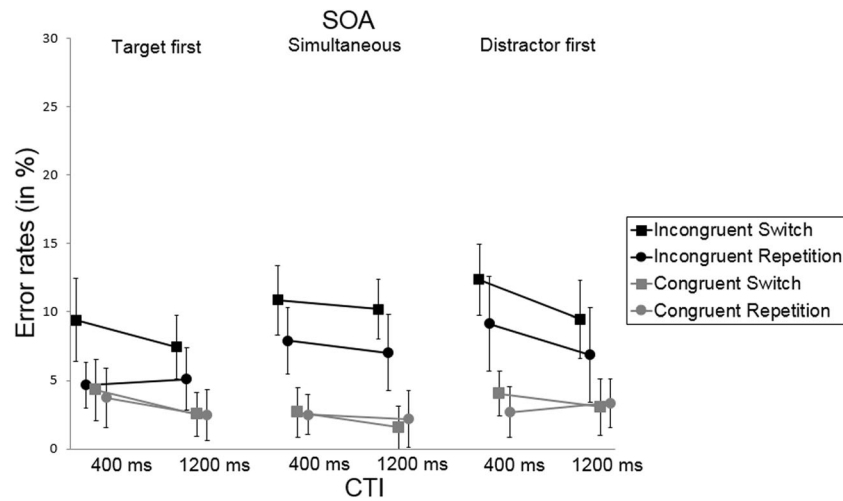
## Experiment 2

We aimed to replicate the core results of Experiment 1 in a simplified version. Instead of three SOA levels, we only used two SOA levels in Experiment 2 by leaving out the simultaneous presentation (SOA = 0).

## Method

**Participants** Twenty-four new participants participated in Experiment 2. The participants had a mean age of 22 years ( $SD = 7$  years, range: 18–54 years), 21 were female, and 23 were right-handed. None of them reported any hearing problems. All participants gave informed consent and participated voluntarily.

**Stimuli and procedure** All stimuli were identical to Experiment 1. The procedure was also identical to Experiment 1, with the exception that target and distractor were never presented simultaneously. The total number of trials in practice blocks and experimental blocks was identical to Experiment 1. This increased the total number of trials per condition, which should also increase reliability.



**Fig. 4** Experiment 1. Error rates. Participants made less errors in congruent trials than in incongruent trials, and even more so in switch trials than in repetition trials. In addition, the congruency effect was smallest when the target was presented first. The error bars represent confidence intervals

**Design** Independent variables were attention transition (repetition, switch), CTI (400 ms, 1,200 ms), SOA (-200 ms, 200 ms), and congruency (congruent, incongruent). Dependent variables were RTs and errors.

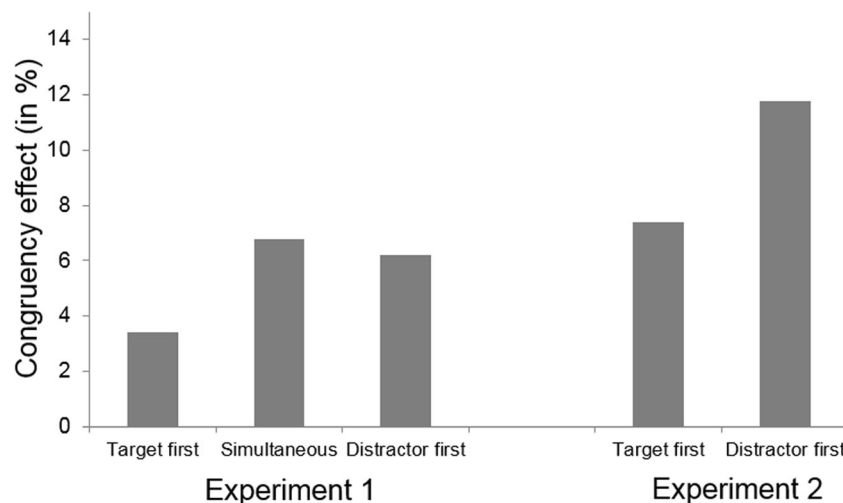
**Results and discussion**

Practice trials, the first trial of each block, and error trials were excluded from the analysis of the RTs, as well as outliers (RT ± 3 SD from the mean of each condition). Practice trials and the first trial of each block were excluded from the analysis of the error rates.

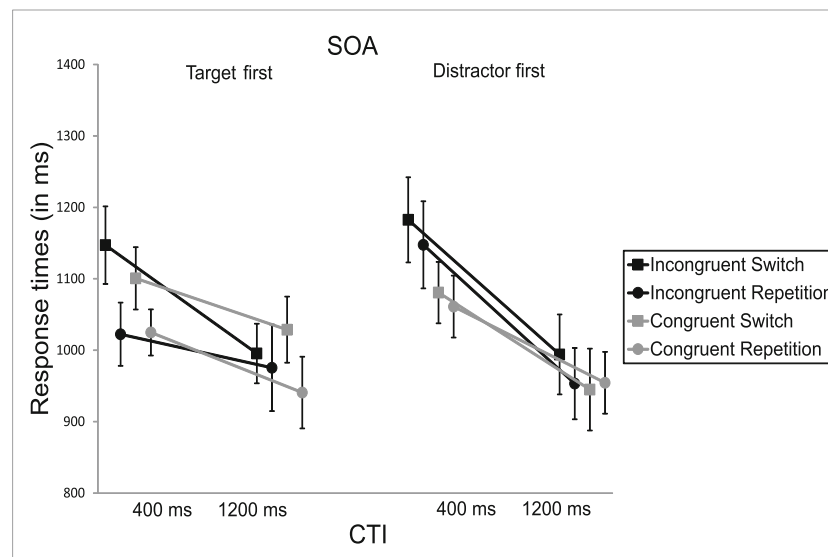
**Response times** We conducted a 2 × 2 × 2 × 2 ANOVA with the within-subject variables attention transition (repetition, switch), CTI (400 ms, 1,200 ms), SOA (-200 ms, 200 ms), and congruency (congruent, incongruent) on RTs (see Fig. 6).

The ANOVA revealed a main effect of transition,  $F(1, 23) = 15.16, p < .001, \eta_p^2 = .40$ , indicating that participants responded faster in repetition trials than in switch trials, 1,010 ms versus 1,059 ms, thus switch costs of 49 ms. In addition, there was a significant interaction of transition and SOA,  $F(1, 23) = 6.39, p < .02, \eta_p^2 = .22$ , indicating that switch costs were greater when the target was presented before the distractor (SOA = 200 ms) than when the distractor was presented before the target (SOA = -200 ms), 77 ms versus 22 ms.

There was a significant main effect of CTI,  $F(1, 23) = 28.19, p < .001, \eta_p^2 = .55$ , indicating that participants responded faster in the CTI 1,200-ms condition than in the CTI 400-ms condition, 973 ms versus 1,096 ms, thus a general preparation benefit of 123 ms. As in Experiment 1, the main effect of CTI was further qualified by a significant interaction of CTI and SOA,  $F(1, 23) = 10.16, p < .01, \eta_p^2 = .31$ , indicating that the CTI effect was greater when the distractor was presented first than



**Fig. 5** Congruency effects. Difference of error rates times between incongruent and congruent trials for all three SOA conditions of both experiments



**Fig. 6** Experiment 2. Response times. Participants responded faster in repetition than in switch trials. Response times were faster when the CTI was 1200 ms than when it was 400 ms. This decrease in response

times with increased CTI was greatest when the distractor was presented first. The error bars represent confidence intervals

when the target was presented first, 156 ms versus 89 ms. Participants thus benefited most from prolonged preparation time when the distractor was presented before the target.

The main effect of congruency was also significant,  $F(1, 23) = 4.31, p < .05, \eta_p^2 = .16$ . This indicated that participants responded faster in congruent trials than in incongruent trials, 1,017 ms versus 1,052 ms, thus there was a congruency effect of 35 ms. The main effect was further qualified by a significant interaction of CTI and congruency,  $F(1, 23) = 6.83, p < .02, \eta_p^2 = .23$ , indicating smaller congruency effects when the CTI was long than when the CTI was short, 13 ms versus 58 ms. In addition, there was a significant interaction of SOA and congruency,  $F(1, 23) = 4.73, p < .05, \eta_p^2 = .17$ , indicating smaller congruency effects when the target was presented before the distractor than when the distractor was presented before the target, 11 ms versus 59 ms.

All other effects were not significant, interaction of transition and CTI:  $F(1, 23) = 2.57, p > .12, \eta_p^2 = .10$ ; interaction of transition, CTI, and SOA:  $F(1, 23) = 1.05, p > .31, \eta_p^2 = .04$ ; interaction of transition, CTI, and congruency:  $F(1, 23) = 1.12, p > .30, \eta_p^2 = .05$ ; interaction of transition, SOA, and congruency:  $F(1, 23) = 3.16, p > .08, \eta_p^2 = .12$ ; interaction of CTI, SOA, and congruency:  $F(1, 23) = 1.17, p > .29, \eta_p^2 = .05$ ; and the four-way interaction,  $F(1, 23) = 3.62, p > .06, \eta_p^2 = .14$ . All other  $F$ s < 1.

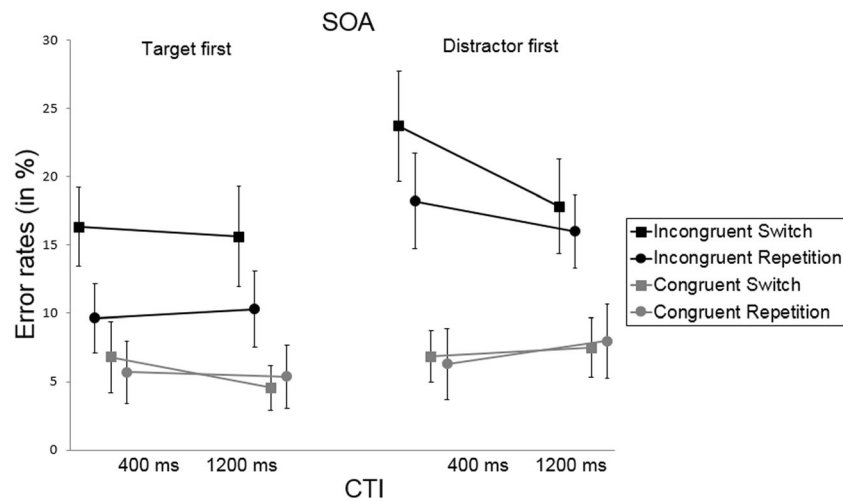
**Errors** We conducted the same ANOVA on error rates (see Fig. 7). The ANOVA revealed a main effect of transition,  $F(1, 23) = 26.00, p < .001, \eta_p^2 = .53$ , indicating switch costs of 2.9%. There was a significant main effect of CTI,  $F(1, 23) = 6.86, p < .02, \eta_p^2 = .23$ , indicating more errors for the short CTI than for the long CTI, 11.5% versus 10.3%. There was a significant

main effect of SOA,  $F(1, 23) = 26.44, p < .001, \eta_p^2 = .54$ , indicating more errors when the distractor was presented before the target than when the target was presented before the target, 12.7% versus 9.0%. There was also a significant interaction of transition and CTI,  $F(1, 23) = 6.72, p < .02, \eta_p^2 = .23$ , indicating smaller switch costs when the CTI was long than when the CTI was short, 1.6% versus 3.8%, and thus a switch-specific preparation benefit.

The main effect of congruency was also significant,  $F(1, 23) = 49.36, p < .001, \eta_p^2 = .68$ , indicating that participants made less errors in congruent trials than in incongruent trials, 5.9% versus 15.8%, thus a general congruency effect of 9.9%. This effect was further qualified by a significant interaction of transition and congruency,  $F(1, 23) = 19.13, p < .01, \eta_p^2 = .45$ , indicating that the congruency effect was smaller in repetition trials than in switch trials, 7.5% versus 12.3%.

In addition, there was a significant interaction of SOA and congruency,  $F(1, 23) = 8.67, p < .01, \eta_p^2 = .27$ , and a significant interaction of CTI, SOA, and congruency,  $F(1, 23) = 11.54, p < .01, \eta_p^2 = .33$ . To decompose the three-way interaction, we separated the two SOA conditions and calculated separate  $2 \times 2$  ANOVAS with the within-subject variables CTI (400 ms, 1,200 ms) and congruency (congruent, incongruent) for each SOA level. When the target was presented before the distractor, the interaction of CTI and congruency was not significant,  $F < 1$ . When the distractor was presented before the target, the interaction of CTI and congruency was significant,  $F(1, 23) = 11.60, p < .01, \eta_p^2 = .36$ , indicating greater congruency effects when the CTI was short than when the CTI was long, 15.5% versus 9.6%, suggesting that the impact of incongruent response information coming from distractor processing is attenuated when there is more preparation time.





**Fig. 7** Experiment 2. Error rates. Participants made less errors in congruent trials than in incongruent trials, and even more so in switch trials than in repetition trials. In addition, the congruency effect was smallest when the target was presented first. The error bars represent confidence intervals

All other effects were not significant, interaction of transition and SOA:  $F(1, 23) = 2.46, p > .13, \eta_p^2 = .10$ ; interaction of CTI, and congruency:  $F(1, 23) = 4.10, p > .05, \eta_p^2 = .15$ ; interaction of transition, SOA, and congruency:  $F(1, 23) = 1.35, p > .26, \eta_p^2 = .06$ . All other  $F$ s < 1.

To summarize, as in Experiment 1, preparation effects in RTs are strongest if the distractor is presented first. In the error rates, as in Experiment 1, the SOA effect is focused on the congruency effect. The congruency effect is smaller if the target precedes the distractor, suggesting that a head start of target processing generally reduces the impact of distractor-based interference. Hence, the main findings of Experiment 1 were replicated in Experiment 2.

## General discussion

The goal of the present study was to investigate target enhancement and distractor suppression in auditory selective attention. More specifically, we focused on dissociating preparation of target enhancement and distractor suppression. Participants performed a classification task on one of two dichotically presented spoken number words, one spoken by a female speaker, one spoken by a male speaker. A cue indicated which gender participants had to attend to in the upcoming trial, so that attention switches and repetitions occurred randomly. We used a short and a long CTI to manipulate preparation time, and we introduced variable delays between target and distractor word to disentangle the effect of CTI on target enhancement and distractor suppression.

Our results revealed a general preparation benefit, and we replicated switch costs. Preparation time did not, in accordance with previous studies, systematically modulate switch costs in the RTs of either experiment. There was a modulation

of switch costs in the error rates of Experiment 2, but not of Experiment 1. We thus found some evidence for switch-specific preparation, but this is restricted to the error rates and is apparently not a very robust finding, which is consistent with previous studies on switch-specific preparation (e.g., Koch et al., 2011; Lawo & Koch, 2015). Importantly, in both experiments, prolonged preparation time decreased RTs more efficiently when the distractor was presented first, compared to the condition with simultaneous presentation of target and distractor (Experiment 1) or when the target was presented first (Experiments 1 and 2). We interpret this result as suggesting that advance preparation facilitates suppressing processing of the distractor. This preparatory process must take place before the onset of target and distractor. In contrast, prolonged preparation time did not lead to additionally reduced RTs when the target was presented first, compared to the baseline condition with simultaneous onset of target and distractor in Experiment 1. Hence, when participants had more time before attending to one of the dichotically presented stimuli, they seemed to prepare more efficiently for distractor suppression than for target enhancement.

In the error rates, congruency effects, especially in switch trials, were replicated. It thus seems as if participants selected and responded to the irrelevant stimuli in some of these trials. This seems to be case especially when participants are supposed to shift the attention focus, thereby presenting further evidence of inert auditory attention settings (see also Nolden & Koch, 2017). Congruency effects were smallest when the target was presented first, compared to when the distractor was presented first or when distractor and target were presented simultaneously. This may indicate that participants select the relevant stimulus more successfully in this condition. Importantly, reduced congruency effects in the target-first condition were not modulated further by transition or preparation time. Participants thus did not seem to use preparation time efficiently to select the

relevant stimulus, so that preparation is more relevant for speed of responding than for accuracy of attention selection.

The current experiment helped us to better understand preparatory mechanisms of auditory selective attention. First of all, our data suggest that target enhancement and distractor suppression in auditory attention might be different processes, as has recently been shown for visual attention as well (Noonan et al., 2016). This fits very well with the idea of a gradual representation of (auditory) stimuli (see, e.g., Meiran et al., 2008), even though there might be a strong bias towards the attended object in auditory attention. It is unclear how the theoretical notion of “biased competition” (e.g., Desimone & Duncan, 1995) could be adapted to the current auditory attention paradigm as the current results suggest that target enhancement and distractor suppression can be, at least partly, dissociated. Even though our study does not provide direct evidence of the relative importance of the two processes in auditory attention, it disentangled the role of the two processes.

We now turn to the differential effects that preparation may have on target versus distractor processing. For a stimulus to stand out in the perceptual stream, the representation of this stimulus can be enhanced, the representation of distracting stimuli can be suppressed, or a combination of both. Our results suggest that target enhancement and distractor suppression benefited differently from preparation. More specifically, this could potentially be put into effect by the strengthening of a template of the target category (e.g., the female voice). When the distractor precedes the target, the distractor could be filtered more efficiently than when the target precedes the distractor. Hence, auditory objects, or some of their properties, not corresponding to the template would be suppressed. In case of the target preceding the distractor, on the other hand, response selection process may be interrupted by distractor processing, and preparation effects would be less beneficial than the other way around (see also Chao, 2010; Ruff & Driver, 2006, for effects of pre-cued distractor locations in vision). Alternatively, a template of the distractor could be built and objects corresponding to it would be suppressed. The irrelevant stimulus would in this case be encoded (in contrast to an early effective filter) and shortly after, disengagement of attention would follow (see also Moher & Egeth, 2012). Even though we cannot specify the exact mechanisms supporting the preparation of distractor suppression, we assume that processes such as segregation and grouping of auditory information, probably based on pitch and timbre of the voices (Rivenez, Guillaume, Bourgeon, & Darwin, 2008), is largely unaffected by these preparatory processes and that they might have the most important effect once target and distractor are represented and one of them is identified as the relevant one (see also Holmes et al., 2018).

Interestingly, the current experiments revealed that auditory attention shifting does not specifically benefit from preparation, nor from specific effects of preparation and target/distractor processing. This finding is in line with previous

research on auditory attention shifting (e.g., Lawo & Koch, 2015; Lawo et al., 2014). We assume that auditory attention settings are quite inert and selecting a new auditory stimulus category for further processing (presumably categorizing, selecting a response etc.) can only be effectuated to a minor extent before stimulus onset. That does, however, not mean that auditory stimulus processing cannot be prepared at all before stimulus onset, as suggested by general preparation benefits and the specific effects of the current study.

On a larger scale, our results seem to suggest that we are indeed able to prepare to a certain degree how to process upcoming auditory targets and distractors. This implies that relevant and irrelevant stimuli are processed to a certain degree and that the suppression of irrelevant stimuli is a process we can actively prepare for. However, preparation does not seem to prevent us from erroneously attending to the irrelevant stimulus in a certain number of trials, since there was neither a reduction of error rates with increased preparation time, nor a reduction of errors in incongruent trials compared to congruent trials with increased preparation time. In addition, since auditory attention ultimately seems to be strongly biased towards the attended object, the preparatory effects of target enhancement and distractor suppression might have augmented the bias in favor of the attended object, but only after the relevant stimulus was actually selected. Our study is therefore in line with previous studies that emphasized the limited flexibility of the auditory attention system (e.g., Koch et al., 2011), and it provides evidence that preparation is especially beneficial for subsequent distractor processing.

Note that our results were obtained in a dichotic listening variant. The interpretation of our results is therefore limited to this specific case of auditory selective attention. Target and distractor were distinct in space and speaker category and should hence have been perceived as distinct auditory objects. Attending to a stimulus when competing stimuli are presented is very common in daily life, for example when we attend to one speaker while other people talk in the background. Other forms of auditory selective attention are also possible, for example when we attend to a temporally limited part of one and the same auditory object. In these settings, auditory attention seems to be much more flexible, probably because streaming is a little less complex (see, e.g., Nolden & Koch, 2017). The present study, however, is based on auditory attention to targets among distractors and our conclusions refer to this special form of auditory selective attention only.

To sum up, we identified processes related to auditory target and distractor representation that can be prepared before the onset of the stimuli. We could show that target enhancement and distractor suppression are different processes and that distractor suppression can be better prepared than target enhancement. We further confirmed results from previous studies, such as that there are switch costs when we switch the selection criterion (e.g., Koch et al., 2011). Preparing for an attention

switch seems to be rather inefficient, pointing to limited flexibility of the auditory attention system. Further research is needed to better specify the mechanisms of target enhancement and distractor suppression as well as their relative importance in auditory attention. Additionally, more research is needed to investigate the flexibility of different variants of auditory attention.

**Acknowledgements** We thank our participants for their participation and our reviewers. Camellia N. Ibrahim is now at Heinrich-Heine-University Düsseldorf. This study was funded by the Excellence Initiative of the German federal and state governments by a grant awarded to Sophie Nolden.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press.
- Broadbent, D. E. (1958). Perception and communication. Oxford: Pergamon Press.
- Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, and Psychophysics*, *77*, 1465–1487.
- Chao, H. F. (2010). Top-down attentional control for distractor locations: The benefit of precuing distractor locations on target localization and discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 303–3016.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*, 975–979.
- Cousineau, D. (2005). Confidence intervals in within-subject design: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45.
- Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). San Diego, CA: Academic Press.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90.
- Holmes, E., Kitterick, P. T., & Summerfield, A. Q. (2018). Cueing listeners to attend to a target talker progressively improves word report as the duration of the cue-target interval lengthens to 2000 ms. *Attention, Perception, and Psychophysics*, *80*, 1520–1538.
- Hugdahl, K. (2011). Fifty years of dichotic listening research- Still going and going and... *Brain and Cognition [Dichotic Listening Anniversary Special Issue]*, *76*, 211–213.
- Jost, K., De Baene, W., Koch, I., & Brass, M. (2013). A review of the role of cue processing in task switching. *Zeitschrift fuer Psychologie*, *221*, 5–14.
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching - A review. *Psychological Bulletin*, *136*, 849–874.
- Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of task inhibition in task switching - a review. *Psychonomic Bulletin and Review*, *17*, 1–14.
- Koch, I., & Lawo, V. (2014). Exploring temporal dissipation of attention settings in auditory attention. *Attention, Perception & Psychophysics*, *76*, 73–80.
- Koch, I., Lawo, V., Fels, J., & Vorländer, M. (2011). Switching in the cocktail party: Exploring intentional control of auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1140–1147.
- Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking – An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*, 557–583.
- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*, *111*, 880–913.
- Lawo, V., Fels, J., Oberem, J., & Koch, I. (2014). Intentional attention switching in dichotic listening: Exploring the efficiency of nonspatial and spatial selection. *The Quarterly Journal of Experimental Psychology*, *67*, 2010–2024.
- Lawo, V., & Koch, I. (2014). Dissociable effects of auditory attention switching and stimulus-response compatibility. *Psychological Research [Special Issue Auditory Attention]*, *78*, 379–386.
- Lawo, V. & Koch, I. (2015). Attention and action: The role of response mappings in auditory attention switching. *Journal of Cognitive Psychology*, *27*, 194–206.
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1423–1442.
- Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, *41*, 211–253.
- Meiran, N., Kessler, Y., & Adi-Japha, E. (2008). Control by action representation and input selection (CARIS): A theoretical framework for task switching. *Psychological Research*, *72*, 473–500.
- Moher, J., & Egeth, H. E. (2012). The ignoring paradox: cueing distractor features leads first to selection, then to inhibition of to-be-ignored items. *Attention, Perception, and Psychophysics*, *74*, 1590–1605.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*, 134–140.
- Morey, R. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64.
- Nolden, S., & Koch, I. (2017). Cognitive control of auditory selective attention to long or short melodic patterns. *Attention, Perception & Psychophysics*, *79*, 1132–1146.
- Noonan, M. P., Adamian, N., Pike, A., Printzlau, F., Crittenden, B. M., & Stokes, M. G. (2016). Distinct mechanisms for distractor suppression and target facilitation. *The Journal of Neuroscience*, *36*, 1797–1807.
- Rivenez, M., Guillaume, A., Bourgeon, L., & Darwin, C. J. (2008). Effect of voice characteristics on the attended and unattended processing of two concurrent messages. *European Journal of Cognitive Psychology*, *20*, 967–993.
- Ruff, C. C., & Driver, J. (2006) Attentional preparation for a lateralized visual distractor: behavioral and fMRI evidence. *Journal of Cognitive Neuroscience*, *18*, 522–538.
- Samson, F., & Johnsrude I. S. (2016). Effects of a consistent target or masker voice on target speech intelligibility in two- and three-talker mixtures. *The Journal of the Acoustical Society of America*, *139*, 1037–1046.
- Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, *18*, 559–572.
- Seibold, J. C., Nolden, S., Oberem, J., Fels, J., & Koch, I. (2018). Intentional preparation of auditory attention-switches: Explicit cueing and sequential switch-predictability. *The Quarterly Journal of Experimental Psychology*, *71*, 1382–1395.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*, 182–186.
- Treisman, A. (1960). Contextual cues in selective listening. *The Quarterly Journal of Experimental Psychology*, *12*, 242–248.
- Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, *76*, 282–299.