RUNNING HEAD: MULTISENSORY FEATURE INTEGRATION

Multisensory feature integration in (and out) of the focus of spatial attention

Charles Spence[1*] & Christian Frings[2*]

[1] University of Oxford, United Kingdom

[2] University of Trier, Germany

[*]shared first authorship

RESUBMITTED TO: *ATTENTION, PERCEPTION & PSYCHOPHYSICS* (TREISMAN SPECIAL ISSUE)

WORD COUNT: 12,050 WORDS

CORRESPONDENCE TO: Prof. Charles Spence, Department of Experimental Psychology, Anna Watts Building, University of Oxford, Oxford, OX2 6GG, UK. E-mail: charles.spence@psy.ox.ac.uk

ABSTRACT

Anne Treisman transformed the way in which we think about visual feature integration. However, that does not mean that she was necessarily right, nor that she looked much beyond vision when considering how features might be bound together into perceptual objects. While such a narrow focus undoubtedly makes sense, given the complexity of human multisensory information processing, it is nevertheless somewhat surprising to find that Treisman herself never extended her feature integration theory outside of the visual modality. After all, she first cut her ‘attentional teeth’ thinking about problems of auditory and audiovisual selective attention. In this article, we review the literature concerning feature integration beyond the visual modality, concentrating, in particular, on the integration of features from different sensory modalities. We highlight a number of the challenges, as far as any straightforward attempt to extend feature integration to the non-visual (i.e., auditory and tactile) and crossmodal (or multisensory) cases, is concerned. These challenges include the problem of how basic features should be defined, the question of whether it even makes sense to talk of objects of perception in the auditory and olfactory modalities, the possibility of integration outside of the focus of spatial attention, and the integration of features from different sensory modalities in the control of action. Nevertheless, despite such limitations, Treisman’s feature integration theory still stands as the standard approach against which alternatives are assessed, be it in the visual case, or increasingly, beyond.

KEYWORDS; ATTENTION; FEATURE INTEGRATION; VISION; AUDITION; SPACE; MULTISENSORY; OBJECT REPRESENTATION.

Introduction

Treisman's Feature Integration Theory (FIT) undoubtedly transformed the way in which we think about the integration of features (e.g., Treisman & Gelade, 1980). Her key suggestion was that attention was required to bind features together into perceptual objects (see Treisman, 1982, 1986, 1988, 1996, and 1998, for reviews). Placed in the context of the early/late selection debate[1] in which her own thinking developed, her theorizing can be seen as a neurophysiologically-inspired second attempt to try to resolve the literature on selective attention. Treisman herself started out her research career in Oxford University's Department of Experimental Psychology publishing in the area of dichotic listening (e.g., see Treisman, 1964, 1969). In her early years as an experimental psychologist, she developed an 'Attenuator Model' of selective listening. Importantly, however, this early cognitive (i.e., box-and-arrow) approach was soon criticized for its lack of neurophysiological plausibility (Styles, 2006; see also Driver, 2001). Her next attempt to resolve the early/late debate in selective attention would be much more firmly grounded in (or at least inspired by) the known visual neurophysiology of the period (see Cowey, 1979, 1985; Livingstone & Hubel, 1985; Zeki, 1978).

Interestingly, though, while often criticizing Treisman's theory on various grounds, subsequent accounts of feature integration have still largely chosen to retain a relatively narrow focus on vision (e.g., see Eckstein, 2011; Quinlan, 2003; Wolfe, 1998, for reviews). That is certainly the case for the most successful successors to Treisman's visual search, such as Wolfe's influential "Guided search" (Wolfe, Cave, & Franzel, 1989) and Müller's dimensional weighting model (Müller, Heller, & Ziegler, 1995; Müller, Krummennacher, & Heller, 2004). While such a narrow focus does perhaps make sense in light of the daunting complexity of human information processing, it nevertheless clearly fails to engage with much of our everyday experience, which involves either non-visual or multisensory information processing (and possibly also object representations). The aim of this review is therefore to provide an up-to-date critical analysis of Treisman's FIT beyond the unimodal (or unisensory) visual case.

---

[1] The early/late debate in cognitive psychology concerns the question of when in information processing attentional selection occurred (i.e., early or late). Part of the reason why this debate has rumbled on for so long relates to the fact that it was often unclear whether researchers were using the terms 'early' and 'late' to refer to time after stimulus onset, or how far along the information processing stream the selection was occurring (see Allport, 1992; Driver, 2001; Shulman, 1990; Styles, 2006, for reviews).

FIT: Key features

Several key (testable) claims were associated with Treisman's original formulation of FIT:

1) All visual features (such as colour, form, and motion) were processed in parallel;

2) A spatial attentional spotlight was required to glue, or bind, visual features together effectively at particular locations within a putative mastermap of locations. Neurophysiological support for the existence of such a mastermap subsequently emerged from data showing that parietal damage (Friedman-Hill, Robertson, & Treisman, 1995) or transcranial magnetic stimulation (TMS) over parietal areas (e.g., Ashbridge, Walsh, & Cowey, 1997) could selectively interfere with feature conjunction (see also Bichot, Rossi, & Desimone, 2005; Shulman, Astafiev, McAvoy, d'Avossa, & Corbetta, 2007).[2]

These two claims were thought to give rise to the apparent dissociation between the parallel search for targets defined by the presence of a unique feature singleton and the serial search for those targets defined by a conjunction of features:

3) Feature binding in the absence of (sufficient) attention was likely to give rise to illusory conjunctions (ICs; e.g., see Treisman & Schmidt, 1982).[3] That is, features from different visual objects might inadvertently be bound together. Indeed, subsequent research revealed that ICs are more common between features that occur closer together spatially (Cohen & Ivry, 1989), or else are otherwise grouped (Prinzmetal, 1981),[4] suggesting some role for spatial segregation, even here.

Treisman's FIT has, in the years since it was originally put forward, been criticized on a number of fronts.[5] Some researchers have questioned whether unique features are necessarily detected in the absence of attention (e.g., Kim & Cave, 1995; Mack & Rock, 1998; see also Braun, 1998). Others have provided examples of feature conjunction targets that are seemingly detected in parallel (e.g., Enns & Rensink, 1990; McLeod, Driver, & Crisp, 1988; Nakayama & Silverman, 1986, for a number of such early examples). Others, meanwhile,

---

[2] According to a somewhat distinct literature, the suggestion has been made that binding may actually be instantiated by synchronized neural firing – this known as the temporal correlation hypothesis (e.g., Gray, 1999; Singer & Gray, 1995; though see also Shadlen & Movshon, 1999).

[3] Note that here we are talking about property binding, one of the seven classes of binding identified by Treisman (1996).

[4] It is an interesting question as to whether ICs are any more common within a modality as compared to between modalities.

[5] This, in fact, being one of the strengths of her theorizing – namely that it offered up a number of testable predictions that galvanized other researchers to try and prove her (right or, more often) wrong.

have questioned whether ICs really are genuinely perceptual in nature, as Treisman would have us believe (e.g., see Virzi & Egeth, 1984, for an early study raising just such concerns). Questions have also been raised about quite how the attentional spotlight (see Tresiman & Gelade, 1980) manages to search efficiently through a scene (Wolfe et al., 1989; see also Klein & MacInnes, 1998);[6] Then, there are those who have questioned just how clear the dichotomy between serial and parallel search really is (e.g., Duncan & Humphreys, 1989); and finally, there are those who have wanted to argue against the very notion of conjunction search as a serial process (see Palmer, 1994).

That all being said, the one aspect of FIT that most researchers working in the area have seemingly not wanted to (or at least have not thought to) question concerns its focus on only a single sense, namely vision. In hindsight, Treisman and her colleagues' theorizing on feature integration was surprisingly narrow, in that it only really engaged with the question of how visual features might be integrated. But what about the integration of features in the other senses such as, for example, audition and touch? Thereafter, one might legitimately also want to know about the integration of features from different sensory modalities (i.e., tapping into questions of multisensory integration; e.g., Kubovy & Schutz, 2010; O'Callaghan, 2016)? After all, object representations, no matter how they are defined (see below for more on this problematic theme), are often signified by cues from multiple distal senses, and not just vision. Beyond the visual dominance that is such a distinctive feature of human information processing (e.g., Posner, Nissen, & Klein, 1976; Spence, Shore, & Klein, 2001), and the fact that the neuroscience underpinnings of visual perception are better worked out than is the case for any of our other senses (e.g., Luck & Beach, 1998),[7] one might think that there is, actually, little reason to prioritize visual object perception over that taking place in the auditory or tactile modalities, say.[8]

---

[6] This leading to a contentious debate concerning the question of whether visual search has a memory or not (see Horowitz & Wolfe, 1994; and the robust reply from Klein, Shore, MacInnes, Matheson, & Christie, 1999; Shore & Klein, 2000).

[7] This no mean feat given the fact that more of the human brain is given over to the processing of visual stimuli than to any of the other senses (see Felleman & Van Essen, 1991, on this theme).

[8] According to some commentators, the technology available to researchers to manipulate/present stimuli might also have played a not insignificant role here too (e.g., see Styles, 2006). The reel-to-reel tape recorder, for instance, facilitating early work on auditory dichotic presentation, and the revolution in visual stimuli technologies with the advent of the personal computer (see Neisser, 1964, for the pre-computer approach to visual search). Interesting in this regard, on her final visit to Oxford, to give a special guest lecture, Treisman reminisced how she had her children draw the stimulus displays for her early experiments involving tachistoscopic presentation.

On the integration of non-visual features

Feature integration surely does take place in other senses, too. For instance, think only of the sound of a musical instrument and how the different features (such as pitch, timbre, and amplitude; see Giard, Lavikainen, Reinikainen, Perrin, Bertrand, Pernier, & Näätänen, 1995) are integrated perceptually. The same presumably must also hold true for touch, where tactile cues are combined into felt objects. That said, online searches reveal little evidence to suggest that researchers have really attempted to extend Treisman's framework into the tactile modality (see Gallace & Spence, 2014, for a review of tactile information processing in humans). Some researchers also talk about flavour objects (see Auvray & Spence, 2008, for a review). However, while it may well be true that mention is made of perceptual objects in the visual, auditory (e.g., Cusack, Carlyon, & Robertson, 2000; Darwin & Hukin, 1999; Kubovy & Van Valkenburg, 2001; O'Callaghan, 2008; Shinn-Cunningham, 2008), tactile, and olfactory (or flavour) modalities (Stevenson, 2014; Stevenson & Wilson, 2007), here it is probably safer (especially given space constraints) to restrict our discussion/critique primarily to the spatial senses of vision, audition, and to a much lesser extent, touch.[9] After all, one of the specific problems that is faced just as soon as one delves into the chemical senses, is that it immediately becomes especially unclear what 'basic features' are (Stevenson, 2014). And, in the case of taste (gustatory) perception, while there are several commonly discussed basic tastes, it is by no means clear that they should be equated with features, as conceptualized in the literature on visual FIT.

However, even restricting ourselves primarily to the integration of features within/between the spatial senses, it is interesting to note how soon one runs into problems/uncertainties/challenges in terms of adapting Treisman's approach. While the neuroscience is reasonably clear in terms of the specific features that are processed in parallel in vision, audition, and touch, what has been taken to constitute a basic feature in each of the senses is somewhat different. So, for example, while spatial location is coded from the retina onwards in vision (and hence constitutes the backdrop against which features are integrated), the analogous dimension in audition is frequency/pitch. This has led some theorists to wonder whether frequency/pitch is to hearing what space is to vision (cf. Kubovy, 1988). Furthermore, in vision, there is anyway a much clearer distinction between features and receptors, while in the case of touch and gustation, say there would appear to be a much

---

[9] Given that the chemical senses are essentially non-spatial, there is presumably no mastermap of locations on which the attentional spotlight can operate, unless integrated with one of the spatial senses. This was presumably the case in Delwiche, Lera, and Breslin's (2000), paper on gustatory search.

closer link. That said, even in the purely visual case, some have raised the concern over a lack of a clear definition of what constitutes a visual feature (see Briand & Klein, 1989, for early discussion of this issue).

Integrating features within audition and touch

Over the years, a number of researchers have tried to extend Treisman's revolutionary ideas around feature integration beyond the confines of the visual modality. For instance, some have attempted to adapt (or extend) her FIT to help explain the constraints on the integration of auditory features (e.g., Woods & Alain, 1993; Woods, Alain, Covarrubias, & Zaidel, 1993; Woods, Alain, Diaz, Rhodes, & Ogawa, 2001; Woods, Alain, & Ogawa, 1998). Certainly, there is widespread talk of auditory objects (Bizley & Cohen, 2013), albeit not without its own controversy/philosophical intrigue (e.g., Griffiths & Warren, 2004; Matthen, 2010; Nudds, 2010; Shinn-Cunningham, 2008). As has already been mentioned, localization comes 'later' in hearing than in vision or touch, and hence one might wonder whether features are perhaps processed prior to their effective localization (although a spatial visual cue has been shown to help with auditory identification; Best, Ozmerla, & Shinn-Cunnigham, 2007). Some have gone even further in suggesting that perhaps space itself should be considered as a perceptual feature, just like pitch, and timbre, say (Woods et al., 2001).

In one study by Woods et al. (year, for example, the participants were required to detect auditory targets from a rapidly presented stream of tone pips (low, medium, or high pitch) that were presented to either ear. In other words, the auditory conjunction search task was, for instance, to respond to high-pitched targets presented from the right ear, say, while the feature search task consisted of reporting whenever a tone of a specific pitch was heard (i.e., regardless of the ear in which it was presented). The results revealed that participants actually detected the conjunction targets more rapidly than the feature targets. Here, though, it should be noted that restricting conjunction targets to one ear may have allowed for mechanisms of spatially selective attention to operate (see Kidd, Arbogast, Mason, & Gallun, 2005; Spence & Driver, 1994; and Shinn-Cunningham, 2008), hence facilitating performance by means of a rather different mechanism. In fact, transposing the experimental design back to the visual modality would presumably also have given the same result – namely faster detection responses when the target location is fixed (see Posner, 1978).

Others, meanwhile, have presented spatial arrays of auditory stimuli (i.e., using a presentation protocol more similar to that seen in vision than the sequential presentation used by Woods et al., 2001). Using such an approach, Hall, Pastore, Acker, and Huang (2000) managed to provide evidence for the role of attention in auditory feature integration. Others have reported the existence of ICs (see Thompson, 1994; Thompson & Hall, 2001).[10] Similarly, other researchers have provided evidence suggestive of pre-attentive feature integration (of timbre and pitch) in audition – specifically for a pair of simultaneously-presented by spatially-distributed sounds using the mismatch negativity response (i.e., an index of auditory deviance; see Takegata, Brattico, Tervaniemi, Varyagina, Näätänen, & Winkler, 2005).

The problematic definition of features

Another challenge that is thrown into sharp relief when one moves outside of the visual modality concerns how, exactly, ‘features’ should be defined? In its original formulation in vision, features were associated with the existence of discrete early visual processing areas (such as for colour, orientation, or motion; e.g., Treisman & Schmidt, 1980). However, the inspiration of features being synonymous with physiologically discrete feature maps in the brain has long since been superseded (see Treisman & Gormican, 1988, p. 16; and Bartels & Zeki, 1998). According to the latter authors, a feature is similar to the concept of a neural channel (see Braddick, Campbell, & Atkinson, 1978). However, outside of the visual modality, it is not so clear that features are necessarily associated with discrete processing areas (although discreet areas have been identified e.g., for frequency, intensity, and duration in audition; see Giard et al., 1995). What is more, contemporary researchers have often questioned whether all features necessarily have a distinct neural processing area attached. That said, there is even some uncertainty about how exactly features should be defined within the visual modality (e.g., Briand & Klein, 1989; Schyns, Goldstone, & Thibaut, 1998). Despite this uncertainty, a number of authors have, over the years, been more than happy to talk about olfactory or tactile objects as feature compounds (e.g., Carvalho, 2014; Keller, 2016; Stevenson & Wilson, 2007; Thomas-Danguin, Sinding, Romagny, El Mountassir, Atanasova, Le Berre, Le Bon, & Coureaud, 2014; Yeshurun & Sobel, 2010). It should be noted here though that, in the case of hearing, the individual identity of features of pure tones say, may blend into harmonies. Furthermore, matters are more complex still in the world of

---

[10] And, intriguingly, Harvey and Treisman (1973) got close to the notion of ICs in audition under conditions of attentional load.

olfactory feature integration where the perceptual consequences of combining discrete odours are still not well understood (Yeshurun & Sobel, 2010).

The problem of spatial alignment

There are undoubtedly a number of important challenges for any account of feature integration as soon as one starts thinking about how to combine features from the different senses. One of the most salient of which concerns the determination of which features come from the same location, and hence should be bound together. Note here only the fact that information encoding is initially retinotopic in vision, tonotopic in hearing, and somatotopic in touch (Spence & Driver, 2004). Hence, for example, the location from which a sound is presented is not given initially, but (as we have seen already) is coded later in information processing. Hence, location is processed – in some sense – late in the auditory modality, while being early in vision (see Shulman, 1990, on this point; see also Kopco, Lin, Shinn-Cunningham, & Groh, 2009).

What is more, there is no obvious immediate means of spatially aligning features in the different senses, given the different frames of reference in which spatial encoding takes place in vision, audition, and touch. The computational problem here being exacerbated by the fact that the various frames of reference will immediately fall out of any kind of spatial alignment once the eyes are moved with respect to the head, say, or the head with respect to the body. Note here that in their earlier work, Spence and Driver (2004) focused on crossmodal links in spatial attention between the auditory, visual, and tactile modalities under just such conditions of receptor misalignment.[11] That said, perhaps we need to stop for a moment to consider whether Treisman's glue really is synonymous with Posner's spotlight, as the preceding discussion would appear to have assumed. For, according to research by Briand and Klein (1987; see also Soetens, Derrost, & Notebaert, 2003) that is by no means necessarily the case: In fact, the answer with respect to endogenous and exogenous attention has been shown to differ. According to Briand and Klein only exogenous attentional orienting behaves equivalently to Treisman's glue.

Multisensory object representations

---

[11] There is also a separate debate on the similarities / differences between attention and integration (see Chen & Spence, 2017a, for one recent review).

While one finds some researchers talking about ‘multisensory objects’ (e.g., Busse, Roberts, Crist, Weissman, & Woldorff, 2005; Turatto, Mazza, & Umiltà, 2005), a closer inspection of the literature soon reveals that the definition of what exactly constitutes a multisensory object is pretty ‘thin’ (see Spence & Bayne, 2015, on this theme). For example, Turatto et al. appear to assume that if an auditory and a visual stimulus are presented from the same location then whoever perceives that combination of cues will *de facto* experience a multisensory object. Much the same can be said in the case of Busse et al.’s study.[12] The definition of what constitutes the necessary and sufficient conditions for positing that an object representation has been formed is, it should be noted, not well defined even within the visual modality (see Feldman, 2003; Scholl, 2001, 2007). Hence, perhaps no wonder that the problem becomes all the more challenging as soon as one considers multisensory object representations (see also Spence & Bayne, 2015). According to Bizley, Maddox, and Lee (2016, p. 74) audiovisual objects can be defined as *“a perceptual construct which occur when a constellation of stimulus features are bound within the brain”.* At the same time, however, while Spence and Bayne acknowledge the widespread evidence for crossmodal interactions (see e.g., Frings & Spence, 2010; Mast, Frings, & Spence, 2014), they question whether any of the evidence that has been published to date convincingly demonstrates the occurrence of multisensory awareness, what some might take as a necessary component of the very existence of multisensory object representations.

Temporal constraints on the integration of signals from different sensory modalities

Separate from the problem of spatial alignment is the challenge of integrating different features that may be processed (i.e., in some sense ‘available’) at different points in time after stimulus onset. A resolution to this problem in the visual modality by has been to assume that the integration of different visual features is achieved as a consequence of the feedforward progression of neuronal responses as they advance through visual areas with progressively more complex receptive fields (Bodelon, Fallah, & Reynolds, 2007). However, the multisensory case is likely to involve greater temporal differences in terms of processing latencies (see also Spence & Squire, 2003).

---

[12] The challenges associated with defining of the minimum conditions necessary for asserting the existence of a multisensory object representation is obviously all the more challenging in young infants (see Bremner, Lewkowicz, & Spence, 2012).

One of the difficulties with ensuring the integration of the appropriate features that becomes potentially more pronounced just as soon as one steps outside of the visual modality relates to the different points in time after stimulus onset at which specific features become available for integration (see Grossberg & Grunewald, 1997). In the case of audiovisual feature integration, for example, Fiebelkorn, Foxe, and Molholm (2010, 2012) have documented the different points in time at which the processing of stimuli is seen in different neural structures. If features from different senses are to be bound appropriately then some means of resynchronizing desynchronized signals may be needed (see Grossberg & Grunewald, 1997, for one potential computational solution to this problem in the visual modality).

Illusory conjunctions

One of the key lines of evidence in support of Treisman's FIT was the existence of ICs under those conditions where, for whatever reason, attention was limited. While evidence in support of the existence of ICs in vision was obtained early on (e.g., Treisman & Schmidt, 1982), one of the enduring controversies in the visual search literature has been whether ICs are genuinely perceptual in nature or whether instead they reflect nothing more than a memory error (see Virzi & Egeth, 1984). Given the theoretical importance of ICs to FIT in vision, one might legitimately ask about the existence of ICs in the other senses, not to mention between them. Researchers have documented the existence of ICs, e.g., between pitch and timbre in audition (Hall & Wieberg, 2003; see also Thompson, 1994; Thompson, Hall, & Pressing, 2001). That said, there would appear to have been little attempt to apply Treisman's FIT to the integration of tactile features. Indeed, we are not aware of any reports documenting ICs within the tactile modality.

But what about ICs in the case of multisensory feature integration?[13] Cinel, Humphreys, and Poli (2002) conducted one of the most thorough (not to mention perhaps the only) studies of crossmodal ICs involving visual and tactile stimuli. These researchers presented visual and

---

[13] It is interesting to consider whether the ventriloquism effect (i.e., when sounds are mislocalized towards the location of simultaneously-presented visual stimuli; e.g., Alais & Burr, 2004; Spence & Driver, 2000) should also be considered as a crossmodal IC or not? The answer here may hinge on whether or not there are reasons to treat the auditory and visual inputs as belonging to the same object or event. Another kind of multisensory IC that occurs on a daily basis is when we mislocalize (and misidentify) odours as tastes in the mouth in the phenomenon known as 'olfactory referral' (see Spence, 2016, for a review). However, according to Spence, the fact that in mislocalizing the olfactory input to the oral cavity, we tend to misidentify the source of the input as gustatory (rather than olfactory), makes this a special, if not unique, case in the world of multisensory perception.

tactile textured shapes to their participants. The latter were required to report the texture of the visual stimuli. Intriguingly, however, in those conditions in which the tactile textures did not match up with the visual textures, tactile-visual-conjunction errors were sometimes observed with participants erroneously reporting the tactile texture as the visual one. As one might have expected, given the tenets of FIT, such ICs were found to be even more common under those conditions where the participant's attention was constrained.

Integration outside of the focus of spatial attention

One of the key claims associated with FIT, as originally proposed by Treisman, is that visual features are essentially only integrated within the focus of (spatial) attention.[14] However, in the crossmodal case, there is now plenty of evidence to suggest that multisensory integration sometimes occurs outside of the focus of (spatial) attention as well. So, for instance, in a series of studies reported by Santangelo and Spence (2007; see also Ho, Santangelo, & Spence, 2009; Santangelo, Ho, & Spence, 2008), audiovisual and audiotactile combinations of spatially co-located peripheral cues were shown to capture participants' spatial attention regardless of the perceptual load of a central attention-demanding rapid serial visual presentation (RSVP) task. Furthermore, multisensory cues captured attention in a way that unisensory auditory, visual, or tactile cues simply failed to do (thus suggesting that multisensory integration had taken place in the absence of, or prior to, spatial attention being allocated to the cued location). Here, though, it should be borne in mind that one might want to separate out the attention-capturing capacity of a certain combination of multisensory cues (when presented from the same location, or direction, at more or less the same time) from the integration of those cues into a coherent whole (that is, a multisensory object of awareness; see Spence & Bayne, 2015, on this theme).[15]

There is, though, a debate here, with Treisman, Sykes, and Gelade (1977) originally suggesting that features come first into perception, and objects identified only later as a result

---

[14] Hence, when attention is stretched and/or stimuli are presented only briefly (Treisman & Schmidt, 1982), or following certain specific kinds of brain damage (Freidman-Hill et al., 1995), ICs are sometimes the result.

[15] In fact, here it is perhaps worth considering the distinction between "what" vs. "where", or "how" pathways that have been widely discussed in vision in recent decades (see Spence, 2013, for a review). More recently, similar distinctions have also been made in the auditory, tactile, and crossmodal cases as well (e.g., see Chan & Newell, 2008; Sestieri, Di Matteo, Ferretti, Del Gratta, Caulo, Tartaro, Olivetti Belardinelli, & Romani, 2006). Hence, it might be relevant to consider whether multisensory integration necessarily always needs to occur in both pathways, or whether instead, 'where-type' integration might drive spatial attention and action without 'what-type' integration occurring, or *vice-versa*.

of focused attention. This view can be contrasted with Wolfe and Cave's (2001, p. 15) suggestion that *"(visual) features of an object are bundled together preattentively but that explicit knowledge of the relationship of one feature to another requires spatial attention"*.

The pip-and-pop effect

One paradigm where this distinction between crossmodal influences and multisensory integration is brought out most clearly is the so-called 'Pip-and-Pop' effect (e.g., Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008; see also Klapetek, Ngo, & Spence, 2012). Van der Burg and his colleagues have conducted numerous studies over the last decade or so showing that the search for a uniquely-oriented line segment (either horizontal or vertical) placed in-amongst an array of diagonally-oriented distractor line segments (i.e., in a complex visual search task) could be made to pop-out (or at least search slopes could be made significantly less steep), simply by presenting a spatially non-predictive auditory tone in synchrony with the sudden change in colour of the visual target (note that both targets and distractors alternated randomly back-and-forth between red-and-green in this experimental paradigm). However, while it may well be tempting to suggest that such results are consistent with the claim that the auditory stimulus and the visual target are integrated in order to create some sort of multisensory object representation, it should be noted that many such crossmodal effects can be accounted for equally well in terms of the crossmodal focusing of temporal attention instead (see Spence & Ngo, 2012, for a review).[16] According to the latter account, note, there really is no need to suggest that any kind of multisensory integration has taken place. What this example therefore helps to illustrate is that just because crosmodal effects are observed in a given experimental paradigm that doesn't necessarily guarantee that any multisensory integration of stimulus features has taken place.

Multisensory integration outside of the focus of attention

For a number of years now, it has been argued that spatio-temporal co-occurrence is key to multisensory integration (see Mast, Frings, & Spence, 2015; Spence, 2007; Stein & Meredith, 1990, 1993; Stein & Stanford, 2008). That said, as highlighted by Spence (2013), closer

---

[16] Here it might be interesting to determine whether the apparent source of the sound is biased toward the location of the visual target or not, as this might be taken as providing evidence of multisensory binding.

inspection of the literature soon reveals that spatial co-occurrence mostly only appears to be necessary in those situations where space is somehow made relevant (either explicitly or implicitly) to a participant's task. By contrast, temporal co-occurrence really does seem to be a prerequisite for any kind of audiovisual (or rather multisensory) integration (e.g., Kolewijn, Bronkhorst, & Theeuwes, 2010; Van der Burg et al., 2008).[17] Notice here only how enhanced multisensory integration is often seen with synchronized sensory signals (e.g., Harrar, Spence, & Harris, 2016). However, it is important to stress that multisensory integration is still often observed for those signals that are slightly desynchronized, providing, that is, that both signals fall within what has been termed the window of multisensory integration (the so-called 'temporal binding window'; e.g., see Colonius & Diederich, 2004; Soto-Faraco & Alsius, 2007, 2009; Wallace & Stevenson, 2014).

The width of this temporal binding window, however, changes as a function of the demands of the participant's task, not to mention the types of stimuli used (e.g., Spence & Squire, 2003; Vatakis, Maragos, Rodomagoulakis, & Spence, 2012; see Vatakis & Spence, 2010, for a review), or various individual differences-related factors (e.g., see Stevenson, Siemann, Schneider, Eberly, Woynaroski, Camarata, & Wallace, 2014). Finally here, it should be noted that there are other factors, such as the correlation between the unisensory signals (Parise, Spence, & Ernst, 2012), crossmodal perceptual grouping (see Spence, 2015, for a review), and higher-order cognitive factors, such as the 'unity assumption' (see Chen & Spence, 2017b, for a review), that have also been shown to contribute to the multisensory integration of audiovisual stimuli.[18]

However, at the same time that the roles of these various factors in multisensory integration are being revealed, the putative role of attention in multisensory integration remains much more ambiguous. While several published studies have demonstrated that attention is needed for successful multisensory integration, a number of other researchers have published findings suggesting that integration is automatic and seemingly independent of attention (e.g., Bertelson, Vroomen, De Gelder, & Driver, 2000; Caclin, Soto-Faraco, Kingstone, & Spence,

---

[17] Intriguingly, according to the results of several audiovisual studies, it turns out that sensory transients may be key to audiovisual binding (e.g., Andersen & Mamassian, 2008; Fujisaki, Koene, Arnold, Johnston, & Nishida, 2006; Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010). At the same time, however, it should also be noted that transients really tell us about events not objects (Meyerhoff, Merz, & Frings, 2018).

[18] One other area of research that may be worthy of further consideration here is how seemingly-unrelated features in different sensory modalities appear, through repeated co-exposure, to become related, such that the presentation of one stimulus primes/evokes the image or representation of the other (e.g., Zangenehpour & Zatorre, 2010; see also Jordan, Clark, & Mitroff, 2010).

2002; Helbig & Ernst, 2008; Santangelo et al., 2008; Santangelo & Spence, 2007; Van der Burg et al., 2008; Vroomen, Bertelson, & De Gelder, 2001). Those working in the field of multisensory perception research have attempted to address the controversy concerning the relationship between attention and multisensory integration by introducing conceptual frameworks that define moderating factors, such as, for example, stimulus complexity, stimulus competition (e.g., Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010), and perceptual load (see Navarra, Alsius, Soto-Faraco, & Spence, 2010, for a review). Furthermore, according to Chen and Spence (2017a), hemispheric asymmetries may also help tease apart the effect of attention from those associated with integration. (Note that hemispheric asymmetry is only expected to influence perception/behaviour when dealing with attentional phenomena.) For instance, according to Talsma and his colleagues, multisensory integration is modulated by top-down attention in those situations in which the competition between stimuli is high, and/or where the stimuli themselves are complex. Under such conditions, the participant’s intentions and goals may well help to determine what is integrated, and attended stimuli will likely be integrated first. On the other hand, when the stimuli are simple, and the competition between them is low, multisensory integration is thought to precede attentional selection and operate in more of a bottom-up manner instead. In this case, preattentive integration may help drive attention to the source of the stimuli (see Spence & Driver, 2000).

Perceptual load moderates the relationship between multisensory integration and attention (see Navarra et al., 2010, for a review). According to Perceptual Load Theory (e.g., Lavie, 1995, 2005, 2010), processing resources are fully used until an individual’s capacity limit is reached. Hence, the suggestion is that under conditions of low load, all stimuli are automatically processed (and hence integrated) because the limit has yet to be reached. With increasing load, however, task-relevant stimuli are processed and integrated first and the integration of task-irrelevant stimuli starts to depend on the remaining processing resources.

Alsius and her colleagues demonstrated a modulation of the McGurk effect (McGurk & MacDonald, 1976) as a function of the perceptual load of a concurrent visual or tactile task (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Alsius, Navarra, & Soto-Faraco, 2007; see also Alsius, Möttönen, Sams, Soto-Faraco, & Tiippana, 2014). Note here also that a similar modulation of the audiovisual ventriloquism effect by perceptual load was reported by Eramudugolla, Kamke, Soto-Faraco, and Mattingley (2011). While such results do fall short of demonstrating that attention is necessary for multisensory integration, they nevertheless do show that attention may modulate it. However, problems with this intuitive perceptual load-based account include the fact that it is difficult to objectively measure capacity, or the perceptual load of a given task, as well as continuing uncertainty over whether or not resources are modality-specific (see Otten, Alain, & Pickton, 2000; Rees, Frith, & Lavie, 2001).

Crucially, however, the studies that have been presented so far have all focused on the integration of target features (i.e., stimulus features that are somehow task-relevant) and the variation in the amount of attention that is devoted to them. The features that have been integrated have always more-or-less been in the focus of attention because the participants have always been tasked with responding to them. At the same time, however, evidence concerning the processing of multisensory distractors – that is, stimuli that are irrelevant to (or may even interfere with) the task at hand – has, until very recently at least, been scarce. One major advantage associated with investigating the multisensory integration of distractor stimuli is that it can be argued that the latter are genuinely processed outside of the focus of attention (excepting, of course, the possibility that they may be actively inhibited; Spence et al., 2001).

In two recent studies, we investigated whether multisensory distractor features are integrated (that is, whether the features are processed independently or not) or whether instead they are only processed on a unisensory level (see Jensen, Merz, Spence, & Frings, 2019; Merz, Jensen, Spence, & Frings, in press). Specifically, multisensory variants of the flanker task were developed using either audiovisual (Jensen et al., 2019) or visuotacile (Merz et al., 2019) stimuli as both the targets and the distractors. Multisensory target stimuli were created by mapping specific combinations of a visual and an auditory (or tactile) feature onto a particular response. In order to respond correctly, the participants in our studies had to process both target features together (i.e. a particular tone together with a particular light colour was assigned to a specific response). Importantly, however, while responding to the multisensory target, the participants had to ignore a multisensory distractor that also comprised visual and auditory (tactile) features. Both of the distractor features could be congruent or incongruent with the target. Crucially, overt spatial attention was manipulated by varying whether the participants fixated on the location from which the distractors were presented (with the targets presented to one side), or *vice versa* (see **Figure 1**).[19]

INSERT FIGURE 1 ABOUT HERE

The results of both studies (Jensen et al., in press; Merz et al., 2019) revealed congruency effects for each distractor feature separately (that is reaction times and error rates were faster or lower when a distractor feature matched the target). Intriguingly, the two modalities only interacted when the multisensory distractor was presented at fixation – i.e., the processing of one modality was not independent of the other one. So, for instance, the effect of a congruent visual distractor feature was more pronounced if the auditory (tactile) feature also happened to be congruent (at a statistical level, this is reflected in significant interaction effects of both congruency effects). By contrast, when the participants' gaze did not fall on the distractor stimuli, each distractor modality produced congruency effects that were independent of the other modality. These results, observed both for audiovisual and visuotactile distractors, can be taken to suggest that overt spatial attention is needed to integrate multisensory distractor features in this demanding selection situation. Without it, the distractor features are likely to be processed independently of each other. Taken together, then, these results concerning the multisensory integration of distractor stimuli fit nicely into the frameworks discussed above that defines moderating factors such as stimulus complexity or the competition between stimuli (Talsma et al., 2010), and perceptual load (Koelewijn et al., 2010; Navarra et al., 2010) as key factors influencing the possible impact of attention on multisensory feature integration. That said, in a way, they also still fit with Treisman's original FIT, as attention here (overt spatial attention, that is) might still be considered the glue that is needed to bind multisensory distractors into multisensory object representations.

Feature integration in action control

Before closing this review, it is worth highlighting the fact that in the decades since Treisman first developed her feature integration framework, the general approach has been extended to various other aspects of cognition (i.e., beyond the purely perceptual). For instance, in the

---

[19] Note that the spatial and temporal rules applied equally in all of the conditions (distractor features were always literally presented at the same location/cube and were presented in synchrony) and cannot be used to explain the different patterns of multisensory integration that were observed in the various conditions of our two studies. Multisensory integration was, however, dependent on overt spatial attention.

field of action control, it is nowadays assumed that stimulus and response features are somehow integrated (e.g., Frings, Koch, Rothermund, Dignath, Giesen, Hommel, et al., in press; Henson, Eckstein, Waszak, Frings, & Horner, 2014; Hommel, 1998) into stimulus-response (S-R) episodes (that can be retrieved later on, and hence may modulate a participant’s behaviour). Interestingly, the role of attention here is even more controversial than in the perceptual literature (e.g., Henson et al., 2014; Hommel, 2004; Moeller & Frings, 2014; Singh, Moeller, & Frings, 2018). Sometimes, it is assumed that attention is needed for integration, sometimes it is not. Part of the problem here is that most current approaches to action control use sequential priming paradigms. This can make it difficult to disentangle integration from retrieval processes. As such, the possible modulation by attention can be hard to pinpoint (see Frings et al., in press; Laub, Moeller, & Frings, 2018, for this argument; see also Töllner, Gramann, Müller, Kiss, & Eimer, 2008; and Zehetleitner, Rangelov, & Müller, 2012, for a related discussion in the literature on visual search).

For present purposes, however, the potential role of attention in S-R feature integration can be neglected. Instead, one can consider these kinds of feature integration as giving rise to multimodal/multisensory feature compounds. Specifically, the perceptual features are integrated with the motor features and the anticipated sensory effects that they will produce (Harless, 1861; James, 1890; Lotze, 1852; for more recent approaches, see Hommel, 2009; and Stock & Stock, 2004, for an overview). In particular, if a participant is instructed to make a keypress (in a standard cognitive experimental paradigm, such as the Stroop task, say) in response to a specific feature, here colour, it is assumed that perceiving the colour will activate the perceptual features, but also the requisite motor features, and then these features will be integrated into an S-R episode. On the next occurrence of the particular stimulus, the previous S-R episode will then be retrieved, including the sensory effects that this episode produced (here, for instance, the tactile sensation of pressing the key) thereby directly facilitating the currently demanded behaviour. Thus, while this kind of feature integration is typically discussed in the context of action control, it can also be seen as an example of multisensory feature integration as – in the example described above – visual stimulus features are integrated with motor features and tactile features produced by pressing the response key. Furthermore, it is worth noting that in many papers on action control that use feature integration and retrieval as basic mechanisms of behaviour, FIT is mentioned, or even discussed, as a relevant precursor (e.g., Frings & Rothermund, 2011, 2017; Henson et al., 2014; Hommel, 2004).

Coming back to vision and visual feature integration

Let us finally reconsider the integration of visual features having discussed auditory, tactile, and multisensory feature integration. The central question of ‘What constitutes a feature?’ is a thorny one. It seems fair to say – as outlined above – even when only looking only at the visual modality, there is still no commonly-agreed definition of feature-hood – given that purely neuronal or physiological definitions seem to be outdated (e.g., Bartels & Zeki, 1998); this becomes even clearer when looking at the other spatial senses (see, for instance, Woods et al.’s, 2001, treating of location, or ear of entry, as a feature in their auditory sequential search study). Furthermore, while location or spatial alignment might be the glue to integrate features in vision, location might just be a feature itself in other senses (e.g., audition) or become even irrelevant (e.g., in the case of olfactory feature binding). In the same vein, one might ask ‘What constitutes an object’? Once again, there is quite some debate about how to define object-hood; yet, when set against the above discussion it becomes abundantly clear that unisensory object-hood, at least in the spatial senses, such as vision, can be assumed to be spatially guided – that is, features belonging to the same object emerge from in the same location. This argument does probably not hold to multisensory objects. Thus, what we can say looking at vision from a multisensory perspective is that if one tries to define feature-hood, the current literature clearly suggests a modality-specific definition of features, perhaps even a modality specific way to integrate these features, and ultimately a modality-specific definition of objects.

Conclusions

Returning to Treisman, and her monumental contribution to the field of cognitive psychology, it is undoubtedly the case that she was, at least in her early research, interested in (not to mention publishing papers on) crossmodal attention (e.g., Treisman & Davies, 1973). It does, therefore, seem a little strange that she never really came back to multisensory issues later on in her career, at least not in the context of FIT.[20] Still, as has hopefully been made clear here, expanding the idea of feature integration beyond the visual modality, and ultimately into the world of multisensory processing, is no easy venture. Problems soon

---

[20] Except, that is, when thinking about the role of crossmodal correspondences in feature binding (Evans & Treisman, 2010) as well as thinking about binding in the case of synaesthesia (Treisman, 2005).

emerge in terms of considering how best to define features, never mind those who question whether it really makes sense to talk of auditory or olfactory object representations. Particularly when processing demands become more complex, as in the case of multisensory integration, or feature integration in action control, the potential role of attention as the glue needed to integrate features becomes increasingly questionable. Nevertheless, no matter what position one chooses to adopt concerning the relationship between attention and integration, it is fair to say that Treisman’s FIT laid the foundations for the modern approach and still, in many contexts, influences current research. Certainly, we have often found ourselves framing our combined research agenda on the theme of multisensory selection in terms of the theoretical framework outlined initially by Treisman some four decades ago.

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, **14**, 257-262.

Allport, D. A. (1992). Selection and control: A critical review of 25 years. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (Vol. 14, pp. 183-218). Hillsdale, NJ: Erlbaum.

Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S., & Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: Evidence from ERPs. *Frontiers in Psychology*, **5**:727.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, **15**, 1-5.

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, **183**, 399-404.

Andersen, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research*, **48**, 2537-2544.

Ashbridge, E., Walsh, V., & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia*, **35**, 1121-1131.

Auvray, M., & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and Cognition*, **17**, 1016-1031.

Bartels, A., & Zeki, S. (1998). The theory of multistage integration in the visual brain. *Philosophical Transactions of the Royal Society, London, Series B. Biological Sciences*, **265**, 2327-2332.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, **62**, 321-332.

Best, V., Ozmerla, E. J., & Shinn-Cunnigham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal of the Association for Research in Otolaryngology*, **8**, 294-2304.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, **308**, 529-534.

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, **14**, 693-707.

Bizley, J. K., Maddox, R. K., & Lee, A. K. C. (2016). Defining auditory-visual objects: Behavioral tests and physiological mechanisms. *Trends in Neuroscience*, **39**, 74-85.

Bodelón, C., Fallah, M., & Reynolds, J. H. (2007). Temporal resolution of the perception of features and conjunctions. *The Journal of Neuroscience*, **27**, 725-730.

Braddick, O., Campbell, F. W., & Atkinson, J. (1978). Channels in vision: Basic aspects. In R. Held, H. L. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. 7* (pp. 3-38). New York, NY: Springer.

Braun, J. (1998). Vision and attention: The role of training. *Nature*, **393**, 424-425.

Bremner, A., Lewkowicz, D., & Spence, C. (Eds.). (2012). *Multisensory development*. Oxford, UK: Oxford University Press.

Briand, K. A., & Klein, R. M. (1987). Is Posner's "beam" the same as Treisman's "glue"?: On the relation between visual orienting and feature integration theory. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 228-241.

Briand, K. A., & Klein, R. M. (1989). Has feature integration theory come unglued? A reply to Tsal. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 401-406.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the USA*, **102**, 18751-18756.

Caclin, A., Soto-Faraco, S., Kingstone, A., & Spence, C. (2002). Tactile “capture” of attention. *Perception & Psychophysics*, **64**, 616-630.

Carvalho, F. (2014). Olfactory objects. *Disputatio*, **6(38)**, 45-66.

Chan, J. S., & Newell, F. N. (2008). Behavioral evidence for task-dependent “what” versus “where” processing within and across modalities. *Perception & Psychophysics*, **70**, 36-49.

Chen, Y.-C., & Spence, C. (2017a). Hemispheric asymmetry: A novel signature of attention’s role in multisensory integration. *Psychonomic Bulletin & Review*, **24**, 690-707.

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

Cinel, C., Humphreys, G. W., & Poli, R. (2002). Cross-modal illusory conjunctions between vision and touch. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 1243-1266.

Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 650-663.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, **16**, 1000-1009.

Cowey, A. (1979). Cortical maps and visual perception. The Grindley Memorial Lecture. *Quarterly Journal of Experimental Psychology*, **31**, 1-17.

Cowey, A. (1985). Aspects of cortical organization related to selective impairments of visual perception: A tutorial review. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance (*Vol. 11, pp. 41-62). Hillsdale, NJ: Erlbaum.

Cusack, R., Carlyon, R. P., & Robertson, I. H. (2000). Neglect between but not within auditory objects. *Journal of Cognitive Neuroscience*, **12**, 1056-1065.

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 617-629.

Delwiche, J. F., Lera, M. F., & Breslin, P. A. S. (2000). Selective removal of a target stimulus localized by taste in humans. *Chemical Senses*, **25**, 181-187.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, **92**, 53-78.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433-458.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, **11**:14. doi:10.1167/11.5.14

Enns, J., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science*, **247**, 721-723.

Eramudugolla, R., Kamke, M., Soto-Faraco, S., & Mattingley, J. B. (2011). Perceptual load influences auditory space perception in the ventriloquist aftereffect. *Cognition*, **118**, 62-74.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, **10(1)**:6, 1-12.

Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences*, **7**, 252-256.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex*, **1**, 1-47.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cerebral Cortex*, **20**, 109-120.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2012). Attention and multisensory feature integration. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 383-394). Cambridge, MA: MIT Press.

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, **269**, 853-855.

Frings, C., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Hommel, B., *et al.* (in press). Merkmalsintegration und Abruf als zentrale Prozesse der Handlungssteuerung – eine Paradigmen-übergreifende Perspektive [Feature binding and retrieval as central processes of action control – an across-paradigm perspective]. *Psychologische Rundschau*.

Frings, C., & Rothermund, K. (2011). To be or not to be...included in an event file: Integration and retrieval of distractors in stimulus-response episodes is influenced by perceptual grouping. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **37**, 1209-1227.

Frings, C., & Rothermund, K. (2017). How perception guides action: Figure-ground segmentation modulates integration of context features into S-R episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **43**, 1720-1729.

Frings, C., & Spence, C. (2010). Crossmodal congruency effects based on stimulus identity. *Brain Research*, **1354**, 113-122.

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society (B)*, **273**, 865-874.

Gallace, A., & Spence, C. (2014). *In touch with the future: The sense of touch from cognitive neuroscience to virtual reality*. Oxford, UK: Oxford University Press.

Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, **7**, 133-143.

Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, **24**, 31-47.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience,* **5(11)**, 887-892.

Grossberg, S., & Grunewald, A. (1997). Cortical synchronization and perceptual framing. *Journal of Cognitive Neuroscience,* **9**, 117-132.

Hall, M. D., Pastore, R. E., Acker, B. E., & Huang, W. (2000). Evidence for auditory feature integration with spatially distributed items. *Perception & Psychophysics,* **62**, 1243-1257.

Hall, M. D., & Wieberg, K. (2003). Illusory conjunctions of musical pitch and timbre. *Acoustics Research Letters Online*, **4**:65; doi: 10.1121/1.1578951

Harless, E. (1861). Der Apparat des Willens [The apparatus of will]. *Zeitschrift für Philosophie und philosophische Kritik*, **38**, 50-73.

Harrar, V., Spence, C., & Harris, L. R. (2017). Multisensory integration is independent of perceived simultaneity. *Experimental Brain Research,* **235**, 763-775.

Helbig, H. B., & Ernst, M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention. *Journal of Vision*, **8**:21.

Henson, R. N., Eckstein, D., Waszak, F., Frings, C., & Horner, A. J. (2014). Stimulus-response bindings in priming. *Trends in Cognitive Sciences,* **18**, 376-384.

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: When spatial correspondence matters. *Experimental Brain Research*, **195**, 261-272.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, **5**, 183-216.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, **8**, 494-500.

Hommel, B. (2009). Action control according to TEC (Theory of Event Coding). *Psychological Research*, **73**, 512-526.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, **394**, 575-577.

James, W. (1890). *The principles of psychology (Vol. 2).* New York, NY: Dover Publications.

Jensen, A., Merz, S., Spence, C., & Frings, C. (2019). Overt spatial attention modulates multisensory selection. *Journal of Experimental Psychology: Human Perception & Performance,* **45**, 174-188.

Jordan, K., Clark, K., & Mitroff, S. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition*, **18**, 492-503.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24**, 175-219.

Keller, A. (2016). Olfactory objects. In *Philosophy of olfactory perception*. Cham, Switzerland: Palgrave Macmillan.

Kidd, G., Jr. Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America*, **118**, 3804-3815.

Kim, M.-S., & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science*, **6**, 376-380.

Klapetek, A., Ngo, M. K., & Spence, C. (2012). Do crossmodal correspondences enhance the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics,* **74**, 1154-1167.

Klein, R. M., & MacInnes, W. J. (1998). Inhibition of return is a foraging facilitator in visual search. *Psychological Science,* **10**, 346-352.

Klein, R. M., Shore, D. I., MacInnes, W. J., Matheson, W. R., & Christie, J. (1999). *Remember that memoryless search theory? Well, forget it! (A Critical Commentary on "Visual search has no memory", by Horowitz & Wolfe, Nature, 394, pp. 575-577).* Unpublished manuscript.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, **134**, 372-384.

Kopco, N., Lin, I.-F., Shinn-Cunningham, B. G., & Groh, J. M. (2009). Reference frame of the ventriloquism aftereffect. *Journal of Neuroscience*, **29**, 13809-13814.

Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance,* **14**, 318-320.

Kubovy, M., & Schutz, M. (2010). Audio-visual objects. *Review of Philosophy & Psychology*, **1**, 41-61.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, **80**, 97-126.

Laub, R., Frings, C., & Moeller, B. (2018). Dissecting stimulus-response binding effects: Grouping by color separately impacts integration and retrieval processes. *Attention, Perception, & Psychophysics*, **80**, 1474-1488.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human, Perception and Performance,* **21**, 451-468.

Lavie, N. (1997). Visual feature integration and focused attention: Response competition from multiple distractor features. *Perception & Psychophysics*, **59**, 543-556.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, **9**, 75-82.

Livingstone, M., & Hubel, D. (1988). Segregation of color, movement, and depth: Anatomy, physiology, and perception. *Science*, **240**, 740-749.

Lotze, R. H. (1852). *Medicinische Psychologie oder die Physiologie der Seele [Medical Psychology or The Physiology of the Soul]*. Leipzig: Weidmann'sche Buchhandlung.

Luck, S. J., & Beach, N. J. (1998). Visual attention and the binding problem: A neurophysiological perspective. In R. D. Wright, *Visual attention* (pp. 455-478). New York, NY: Oxford University Press.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Mast, F., Frings, C., & Spence, C. (2014). Response interference in touch, vision, & crossmodally: Beyond the spatial dimension. *Experimental Brain Research*, **232**, 2325-2336.

Mast, F., Frings, C., & Spence, C. (2015). Multisensory top-down sets: Evidence for contingent crossmodal capture. *Attention, Perception, and Psychophysics,* **77**, 1970-1985.

Matthen, M. (2010). On the diversity of auditory objects. *Review of Philosophy and Psychology*, **1**, 63-89.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, **332**, 154-155.

Merz, S., Jensen, A., Spence, C., & Frings, C. (2019). Multisensory distractor processing is modulated by spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*.

Meyerhoff, H. S., Merz, S., & Frings, C. (in press). Tactile stimulation disambiguates the perception of visual motion paths. *Psychonomic Bulletin & Review*.

Moeller, B., & Frings, C. (2014). Attention meets binding: Only attended distractors are used for the retrieval of event files. *Attention, Perception, & Psychophysics*, **76**, 959-978.

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, **57**, 1-17.

Müller, H. J., Krummenacher, J., & Heller, D. (2004). Dimension-specific inter-trial facilitation in visual search for pop-out targets: Evidence for a top-down modulable visual short-term memory effect. *Visual Cognition*, **11**, 577-602.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, **320**, 264-265.

Navarra, J., Alsius, A., Soto-Faraco, S., & Spence, C. (2010). Assessing the role of attention in the audiovisual integration of speech. *Information Fusion*, **11**, 4-11.

Neisser, U. (1964). Visual search. *Scientific American*, **210(July)**, 94-102.

Nudds, M. (2010). What are auditory objects? *Review of Philosophy and Psychology*, **1**, 105-122.

O’Callaghan, C. (2008). Object perception: Vision and audition. *Philosophy Compass*, **3**, 803-829.

O'Callaghan, C. (2016). Objects for multisensory perception. *Philosophical Studies*, **173**, 1269-1289.

Otten, L. J., Alain, C., & Picton, T. W. (2000). Effects of visual attentional load on auditory processing. *NeuroReport*, **11**, 875-880.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, **34**, 1703-1721.

Parise, C. V., Spence, C., & Ernst, M. (2012). When correlation implies causation in multisensory integration. *Current Biology*, **22**, 46-49.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, **83**, 157-171.

Prinzmetal, W. (1981). Principles of feature integration in visual attention. *Perception & Psychophysics*, **30**, 330-340.

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, **129**, 643-673.

Rees, G., Frith, C., & Lavie, N. (2001). Processing of irrelevant visual motion during performance of an auditory attention task. *Neuropsychologia*, **39**, 937-949.

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human, Perception and Performance*, **33**, 1311-1321.

Santangelo, V., Ho, C., & Spence, C. (2008). Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review*, **15**, 398-403.

Scholl, B. J. (2001). Objects and attention: The state of art. *Cognition*, **80**, 1-46.

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind and Language*, **22**, 563-591.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, **21**, 1-54.

Sestieri, C., Di Matteo, R., Ferretti, A., Del Gratta, C. Caulo, M. Tartaro, A. Olivetti Belardinelli, M., & Romani, G. L. (2006). "What" versus "where" in the audiovisual domain: An fMRI study. *NeuroImage*, **33**, 672-680.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the binding hypothesis. *Neuron*, **24**, 67-77.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, **12**, 182-186.

Shore, D. I., & Klein, R. M. (2000). On the manifestations of memory in visual search. *Spatial Vision*, **14(1)**, 59-75.

Shulman, G. L. (1990). Relating attention to visual mechanisms. *Perception & Psychophysics*, **47**, 199-203.

Shulman, G. L. Astafiev, S. V., McAvoy, M. P., d'Avossa, G., & Corbetta, M. (2007). Right TPJ deactivation during visual search: Functional significance and support for a filter hypothesis. *Cerebral Cortex*, **17**, 2625-2633.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Reviews of Neuroscience*, **18**, 555-586.

Singh, T., Moeller, B., Koch, I., & Frings, C. (2018). May I have your attention please: Binding attended but response irrelevant features. *Attention, Perception, & Psychophysics*, **80**, 1143-1156.

Soetens, E., Derrost, N., & Notebaert, W. (2003). Is Treisman’s ’glue’ related to Posner’s ‘beam’? *Abstracts of the Psychonomic Society*, **8**, 10-11.

Soto-Faraco, S., & Alsius, A. (2007). Conscious access to the unisensory components of a cross-modal illusion. *Neuroreport*, **18**, 347-350.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *Journal of Experimental Psychology: Human Perception & Performance*, **35**, 580-587.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science & Technology*, **28**, 61-70.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, **73**, 971-995.

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.

Spence, C. (2015). Cross-modal perceptual organization. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 649-664). Oxford, UK: Oxford University Press.

Spence, C. (2016). Oral referral: On the mislocalization of odours to the mouth. *Food Quality & Preference,* **50**, 117-128.

Spence, C., & Bayne, T. (2015). Is consciousness multisensory? In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and its modalities* (pp. 95-132). Oxford, UK: Oxford University Press.

Spence, C. [J.], & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 555-574.

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: Reflexive crossmodal orienting and ventriloquism. *NeuroReport,* **11**, 2057-2061.

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

Spence, C., & Ngo, M. K. (2012). Does attention or multisensory integration explain the crossmodal facilitation of masked visual target identification? In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 345-358). Cambridge, MA: MIT Press.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multimodal prior entry. *Journal of Experimental Psychology: General*, **130**, 799-832.

Spence, C., & Squire, S. B. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology,* **13**, R519-R521.

Stein, B. E., & Meredith, M. A. (1990). Multisensory integration. Neural and behavioral solutions for dealing with stimuli from different sensory modalities. *Annals of the New York Academy of Sciences,* **608**, 51-65; discussion 65-70.

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

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, **9**, 255-267.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in Autism Spectrum Disorders. *Journal of Neuroscience*, **34**, 691-697.

Stevenson, R. J. (2014). Object concepts in the chemical senses. *Cognitive Science,* **38(7)**, 1360-1383.

Stevenson, R. J., & Wilson, D. A. (2007). Odour perception: An object-recognition approach. *Perception,* **36**, 1821-1833.

Stock, A., & Stock, C. (2004). A short history of ideomotor action. *Psychological Research,* **68**, 176-188.

Styles, E. A. (2006). *The psychology of attention* (2nd Ed.). Hove, UK: Psychology Press.

Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects. *Cognitive Brain Research*, **25**, 169-179.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, **14**, 400-410.

Thomas-Danguin, T., Sinding, C., Romagny, S., El Mountassir, F., Atanasova, B., Le Berre, E., Le Bon, A.-M., & Coureaud, G. (2014). The perception of odor objects in everyday life: A review on the processing of odor mixtures. *Frontiers in Psychology*, **5**:504. http://doi.org/10.3389/fpsyg.2014.00504.

Thompson, W. F. (1994). Sensitivity to combinations of musical parameters: Pitch with duration and pitch pattern with durational pattern. *Perception & Psychophysics*, **56**, 363-374.

Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, **27**, 128-140.

Töllner, T., Gramann, K., Müller, H. J., Kiss, M., & Eimer, M. (2008). Electrophysiological markers of visual dimension changes and response changes. *Journal of Experimental Psychology: Human Perception and Performance*, **34**, 531-542.

Treisman, A. (1964). The effects of irrelevant material on the efficiency of selective listening. *American Journal of Psychology*, **77**, 533-546.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, **76**, 282-299.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, **8(2)**, 194-214.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, **255**, 106-111.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, **40A**, 201-237.

Treisman, A. (1991). Search, similarity and interpretation of features between and within dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 652-676.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, **6**, 171-178.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society London B*, **353**, 1295-1306.

Treisman, A. (2005). Synesthesia: Implications for attention, binding, and consciousness – A commentary. In L. Robertson & N. Sagiv (Ed.), *Synaesthesia: Perspectives from cognitive neuroscience* (pp. 239-254). Oxford, UK: Oxford University Press.

Treisman, A. M., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and performance* (Vol. 4, pp. 101-117). New York, NY: Academic Press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, **95**, 15-48.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, **14**, 107-141.

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), *Attention and performance VI* (pp. 333-361). Hillsdale, NJ: Lawrence Erlbaum.

Turatto, M., Mazza, V., & Umiltà, C. (2005). Crossmodal object-based attention: Auditory objects affect visual processing. *Cognition*, **96**, B55-B64.

Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE*, **5**:e10664. doi:10.1371/journal.pone.001066

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Non-spatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **34**, 1053-1065.

Vatakis, A., Maragos, P., Rodomagoulakis, I., & Spence, C. (2012). Assessing the effect of physical differences in the articulation of consonants and vowels on audiovisual temporal perception. *Frontiers in Integrative Neuroscience*, **6:71**, 1-18. doi: 10.3389/fnint.2012.00071

Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli. In M. J. Naumer & J. Kaiser (Eds.), *Multisensory object perception in the primate brain* (pp. 95-121). New York, NY: Springer.

Virzi, R. A., & Egeth, H. E. (1984). Is meaning implicated in illusory conjunctions? *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 573-580.

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, **63**, 651-659.

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, **64**, 105-123.

Wolfe, J. M. (1998). *Visual search*. In H. Pashler (Ed.), *Attention* (pp. 13-73). Hove, East Sussex: Psychology Press.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, **24**, 11-17.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 419-433.

Woods, D. L., & Alain, C. (1993). Feature processing during high-rate auditory selective attention. *Perception & Psychophysics*, **53**, 391-402.

Woods, D. L., Alain, C., Covarrubias, D., & Zaidel, O. (1993). Frequency-related differences in the speed of human auditory processing. *Hearing Research*, **66**, 46-52.

Woods, D. L., Alain, C., Diaz, R., Rhodes, D., & Ogawa, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 65-74.

Woods, D. L., Alain, C., & Ogawa, K. H. (1998). Conjoining of auditory and visual features during high-rate serial presentation: Processing and conjoining two features can be faster than processing one. *Perception & Psychophysics*, **60**, 239-249.

Yeshurun, Y., & Sobel, N. (2010). An odor is not worth a thousand words: From multidimensional odors to unidimensional odor objects. *Annual Review of Psychology*, **61**, 219-241.

Zangenehpour, S., & Zatorre, R. J. (2010). Cross-modal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, **48**, 591-600.

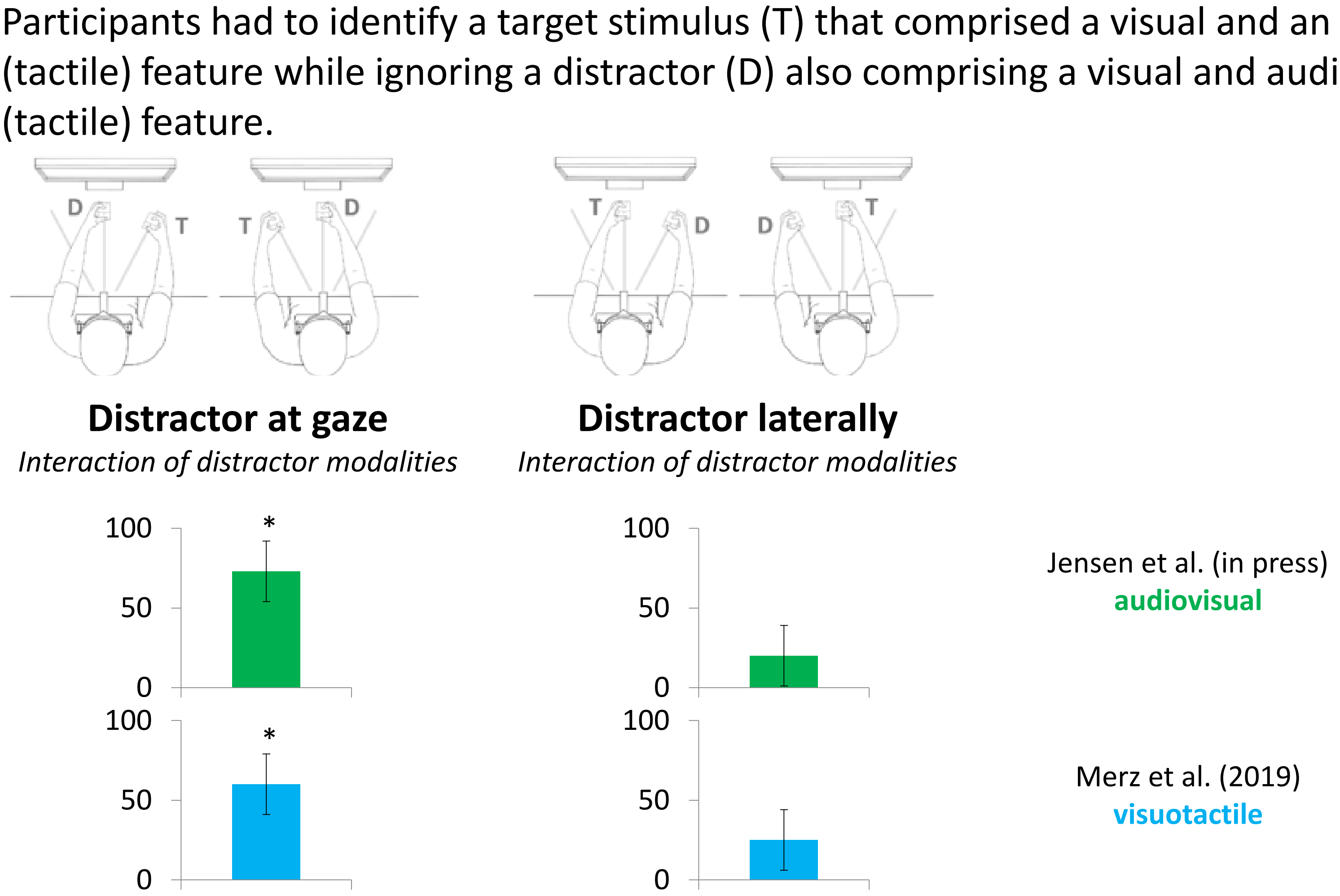
Zehetleitner, M., Rangelov, D., & Müller, H. J. (2012). Partial repetition costs persist in nonsearch compound tasks: Evidence for multiple-weighting-systems hypothesis. *Attention, Perception & Psychophysics*, **74**, 879-890.

Zeki, S. M. (1978). Functional specialization in the visual cortex of the rhesus monkey. *Nature*, **274**, 423-428.

FIGURE LEGENDS

Figure 1. Summary of experimental set-up and results from Jensen et al.'s (2019) and Merz et al.'s (2019) studies. Bird's eye view on the experimental set-up and key results of both studies highlighting dependence or independence of the congruency of distractors in both sensory modalities (the interaction term of visual and auditory [tactile] congruency RT effects in milliseconds; error bars depict standard error of the mean; * $p < .01$).

Figure 1



Participants had to identify a target stimulus (T) that comprised a visual and an auditory (tactile) feature while ignoring a distractor (D) also comprising a visual and auditory (tactile) feature.

Distractor at gaze

Interaction of distractor modalities

Distractor laterally

Interaction of distractor modalities

Jensen et al. (in press)
audiovisual

Merz et al. (2019)
visuotactile