Acta Psychologica 134 (2010) 372–384

Contents lists available at ScienceDirect

Acta Psychologica

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

Attention and the multiple stages of multisensory integration: A review of audiovisual studies

Thomas Koelewijn [a,b,*], Adelbert Bronkhorst [a,c], Jan Theeuwes [a]

[a] Cognitive Psychology, Vrije Universiteit, Amsterdam, The Netherlands
[b] ENT Audiology, VU University medical center, Amsterdam, The Netherlands
[c] TNO Human Factors, Soesterberg, The Netherlands

ARTICLE INFO

Article history:
Received 10 December 2009
Received in revised form 23 March 2010
Accepted 27 March 2010
Available online 27 April 2010

Keywords:
Multisensory integration
Crossmodal
Attention
Audiovisual

ABSTRACT

Multisensory integration and crossmodal attention have a large impact on how we perceive the world. Therefore, it is important to know under what circumstances these processes take place and how they affect our performance. So far, no consensus has been reached on whether multisensory integration and crossmodal attention operate independently and whether they represent truly automatic processes. This review describes the constraints under which multisensory integration and crossmodal attention occur and in what brain areas these processes take place. Some studies suggest that multisensory integration and crossmodal attention take place in higher heteromodal brain areas, while others show the involvement of early sensory specific areas. Additionally, the current literature suggests that multisensory integration and attention interact depending on what processing level integration takes place. To shed light on this issue, different frameworks regarding the level at which multisensory interactions takes place are discussed. Finally, this review focuses on the question whether audiovisual interactions and crossmodal attention in particular are automatic processes. Recent studies suggest that this is not always the case. Overall, this review provides evidence for a parallel processing framework suggesting that both multisensory integration and attentional processes take place and can interact at multiple stages in the brain.

1. Introduction

When you are reading a newspaper on a train the sound of loud music to your left or someone talking on the phone to your right can be distracting. You may skip a line, misread a header, or even stop reading when a conversation behind you grasps your attention. Why is it so hard to stay focused on your readings when you hear sounds around you? Why can't you block out these sounds when you know that they are irrelevant? Although distracting when you try to read, these interactions between what we hear and what we see can save your life — for example when the sound of a car coming from your left makes you freeze.

These everyday examples illustrate the strong interactions that exist between our auditory and visual systems. These interactions can occur at the level of 'multisensory integration' (see Stein & Stanford, 2008), as when a voice and a moving mouth are integrated into a single event (e.g., McGurk & MacDonald, 1976). Multisensory integration helps us perceive information better, which might be why it is so tempting to look over our newspaper when eavesdropping on a conversation between two people sitting opposite in the train. Additionally, these interactions can be at an attentional level (see Driver & Spence, 1998) in which, for example, a sound draws our visual attention to a certain location (e.g., Spence & Driver, 1997). This might be why it is so hard to focus our attention on the words in the newspaper in front of us when someone is snapping chewing gum next to us.

Early studies on perceptual and attentional processes primarily investigated sensory modalities in isolation. However, in the last two decades or so more research has addressed the interaction between modalities. This allows us to get a full picture of how these processes work in the brain, but also to relate these outcomes to more realistic situations in which auditory and visual events hardly ever occur in isolation. With current technology developments the question of when to expect audiovisual interactions becomes more pressing than ever. For instance, in-car technologies like navigational systems overflow us with audiovisual information. The impact of sounds on our driving ability, which is primarily a visual task, has become a hot research topic (see Ho & Spence, 2005; Spence & Ho, 2008).

Audiovisual interactions may allow us to focus on relevant information and filter out irrelevant information, or may cause distraction when our attention is captured against our will by audiovisual information that is irrelevant for our task. We speak of

* Corresponding author. ENT Audiology, VU University medical center, Boelelaan 1117, 1081HV Amsterdam, The Netherlands. Tel.: +31 20 44 40900.
E-mail address: t.koelewijn@vumc.nl (T. Koelewijn).

0001-6918/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2010.03.010

attentional capture when spatial attention is drawn to a location in space against our intentions (Theeuwes, Belopolsky, & Olivers, 2009). For example, even though our goal may be to read a book, our attention may get drawn to the location where a person is making a sound. The question that is central in this review is whether visual attention that is voluntarily directed to a specific spatial location can be drawn away automatically from that location towards the location where a sound is coming from. Even though previous studies have shown that attentional capture can occur between the different modalities (e.g., Spence & Driver, 1997), the question remains whether a localizable sound captures visual spatial attention (crossmodal capture) under all circumstances. Recent studies have shown that in some circumstances audiovisual interactions like crossmodal capture do not occur (Koelewijn, Bronkhorst, & Theeuwes, 2009b; Santangelo & Spence, 2007) while other studies have shown that in most circumstances irrelevant sounds do affect our visual system (e.g., Koelewijn, Bronkhorst, & Theeuwes, 2009a; Mazza, Turatto, Rossi, & Umilta, 2007; McDonald, Teder-Salejarvi, & Hillyard, 2000; Spence & Driver, 1997; van der Lubbe & Postma, 2005; Ward, 1994). This review addresses the question under what circumstances crossmodal capture occurs. Additionally, recent research has shown that multisensory integration and (crossmodal) attention interact at certain brain levels (e.g., Fairhall & Macaluso, 2009; Mozolic, Hugenschmidt, Peiffer & Laurienti, 2008; Talsma, Doty, & Woldorff, 2005; Talsma, Doty, & Woldorff, 2007; Talsma & Woldorff, 2005). This review also addresses the levels at which these interactions may occur.

In addition to vision and audition, multimodal interactions are also known to occur between taste, smell, and touch senses (e.g., see Driver & Spence, 1998; Stein & Stanford, 2008). So far, most research has been directed at the interactions between our visual, auditory, and somatosensory systems and has been focused on interactions at an attentional level or at a multisensory integration level. This review will focus on studies mainly discussing interactions between the visual and auditory modality, although sometimes a reference will be made to somatosensory studies to illustrate that effects apply more generally.

Although our perceptual systems seem fully integrated, modality specific features tend not to interact, as shown by Alais, Morrone and Burr (2006) for auditory pitch and visual contrast perception. However, there is a form of interaction called synaesthesia where non-overlapping features between modalities do integrate. For example Baron-Cohen, Wyke and Binnie (1987) have shown that some people see colors when hearing numbers which seems to imply some form of multimodal interaction. However, Rouw and Scholte (2007) have shown that the structure of the brain of those people that experience synaesthesia may be different from those that do not experience synaesthesia, suggesting that the occurrence of synaesthesia and its implied multimodal interaction is not a general phenomenon.

This paper reviews studies that investigated audiovisual interactions in the form of multisensory integration and crossmodal attention. Both types of interactions take place at multiple processing levels within the brain. The first section describes the processing levels at which information from the auditory and visual modalities meet and integrate. This is followed by a review of studies that specifically look at attentional capture across the auditory and visual modalities. The section that follows introduces the idea that multisensory integration and crossmodal attention sometimes act independently, and at other times interact. To shed light on this issue, different frameworks regarding the level at which multisensory interactions take place are discussed. The final section focuses on the question of whether audiovisual interactions and crossmodal attention are automatic processes. The literature shows that crossmodal attention does not always meet the criteria for automaticity. One possibility is that these findings can be explained in terms of parallel processing. Both behavioral and electrophysiological studies will be discussed to provide a full picture of the current status on this topic. The present paper presents a broad overview of studies regarding audiovisual integration and attention.

2. Multisensory integration

We need multisensory integration in order to recognize different types of sensory input as belonging to the same object. Multisensory integration helps to reduce noise within our perceptual system by combining information from different sensory modalities (see Stein, Stanford, Wallace, & Jiang, 2004). Less noisy input allows for an easy separation of events from background noise and division between successive events. For example a sound can boost the detectability of visual events (see Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008). Even though some multimodal behavioral effects and illusions resulting from multisensory integration were reported as early as the 1960′s and 1970′s (e.g., Hershenson, 1962; McGurk & MacDonald, 1976), research on multisensory integration has skyrocketed in the last two decades. Psychophysical studies have demonstrated that the notion that sensory information is processed within each modality separately in a feedforward fashion is incorrect (see Driver & Spence, 2000). In addition, animal physiology (see Stein & Stanford, 2008), human electrophysiology (Talsma et al., 2007) and human imaging studies (Calvert, Campbell, & Brammer, 2000) have provided evidence that multisensory integration is not restricted to higher multisensory (heteromodal) brain areas (see Macaluso & Driver, 2005). This section discusses under what circumstances and where in the brain multisensory integration takes place. First, some multisensory illusions and effects will be discussed to illustrate the strength of multisensory integration.

2.1. Multisensory integration effects and illusions

Although multisensory integration is the process that binds information from different modalities, most of the time you are not aware of its occurrence. Still, there are some multisensory integration effects or illusions of which we can become consciously aware. Ventriloquism (Thurlow & Jack, 1973) is a well-known example. In this illusion, the voice of the puppeteer seems to project from the mouth of the puppet itself. This attribution of voices to congruent sources is generally beneficial and results in improved perception under noisy circumstances (Sumby & Pollack, 1954).

Ventriloquism is most commonly demonstrated in the shift of sound toward the location of a visual event. In the puppet illusion, sound is shifted toward a congruent source, but Slutsky and Recanzone (2001) demonstrated that ventriloquism also occurs with simple auditory and visual onsets that have no semantic value. The same study showed that there are spatial and temporal constraints to the ventriloquism effect. This means that these events should take place not too far apart in space and preferably should co-occur in time. Temporal and spatial restrictions generally apply to multisensory integration and will be discussed in the next section. The ventriloquism effect suggests that the visual system is dominant over the auditory system when it comes to spatial localization. However, other illusions that are discussed below demonstrate that this is not always the case.

Ventriloquism can also pull sensory events together in terms of time, such that the perceived temporal proximity of two successive visual events is affected by auditory input. For example, in Morein-Zamir, Soto-Faraco, and Kingstone (2003) participants performed a temporal order judgment task on the onsets of two LEDs. When a sound was presented before the first onset and after the second onset, compared to a neutral condition in which the sound coincided with the LED onsets, the participants' performance benefitted. It seemed as if the visual onset was pulled in time towards the auditory onsets, which made temporal order judgment of the visual events easier.