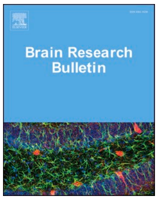
Brain Research Bulletin 210 (2024) 110923

Contents lists available at ScienceDirect

Brain Research Bulletin

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



The role of attention in immersion: The two–competitor model

Daniel J. Strauss [a,*,1], Alexander L. Francis [b,1], Jonas Vibell [c], Farah I. Corona–Strauss [a]

[a] *Systems Neuroscience & Neurotechnology Unit, Faculty of Medicine, Saarland University & School of Engineering, htw saar, Homburg/Saar, Germany*
[b] *Speech Perception & Cognitive Effort Lab, Dept. of Speech, Language & Hearing Sciences, Purdue University, West Lafayette, IN, USA*
[c] *Brain & Behavior Lab, Dept. of Psychology, University of Hawai'i at Manoa, Honolulu, HI, USA*

ARTICLE INFO

Keywords:
Immersion
Virtual Reality
Augmented Reality
Metaverse
Attention
Affect

ABSTRACT

Currently, we face an exponentially increasing interest in immersion, especially sensory–driven immersion, mainly due to the rapid development of ideas and business models centered around a digital virtual universe as well as the increasing availability of affordable immersive technologies for education, communication, and entertainment. However, a clear definition of 'immersion', in terms of established neurocognitive concepts and measurable properties, remains elusive, slowing research on the human side of immersive interfaces.

To address this problem, we propose a conceptual, taxonomic model of attention in immersion. We argue (a) modeling immersion theoretically as well as studying immersion experimentally requires a detailed characterization of the role of attention in immersion, even though (b) attention, while necessary, cannot be a sufficient condition for defining immersion. Our broader goal is to characterize immersion in terms that will be compatible with established psychophysiolgical measures that could then in principle be used for the assessment and eventually the optimization of an immersive experience. We start from the perspective that immersion requires the projection of attention to an induced reality, and build on accepted taxonomies of different modes of attention for the development of our two–competitor model. The two–competitor model allows for a quantitative implementation and has an easy graphical interpretation. It helps to highlight the important link between different modes of attention and affect in studying immersion.

1. Introduction

Sensory–driven immersion is becoming more and more a topic of our everyday life. The reasons for this are manifold: The rapid development and availability of 'immersive technologies'; from virtual and mixed reality headsets to virtual audio and haptic displays, and the concomitant development of business models organized around a digital virtual universe (metaverse) (Cagnina and Poian, 2008; Hopkins, 2022; Panagiotakopoulos et al., 2022) but also, more recently, the greatly increased use of digital environments for communication and education in the COVID–19 pandemic see, e.g., (Cheng et al., 2022). The nature of immersion is addressed in several scientific domains, ranging from computer game research (Ermi and Mäyrä, 2005), education (Makransky and Petersen, 2021) and entertainment (Visch et al., 2010) to tracts in philosophy (Schellenberg, 2013). Even though some categorizations have been proposed such as perceptual (Biocca and Delaney, 1995), sensory (Ermi and Mäyrä, 2005), and narrative immersion (Ryan, 2003; Thon, 2008) and their subcategories of emotional, spatial, and temporal immersion (see Ryan, 2003), science has not really kept pace with the inflationary use of the catch–all term 'immersion' by defining what immersion as generic concept actually is; a problem which has been pointed out by several researchers, e.g., see (McMahan, 2003; Chasid, 2017; Nilsson et al., 2016; Murray, 2017; Agrawal et al., 2020). Defining immersion is not just important from a theoretical perspective. It also has practical implications. Immersion is a state that does not seem to stand up well to overt introspection in the moment. It is not useful to ask someone "are you immersed?" any more than it is useful to ask "are you asleep?". Immersion seems to rely at least in part on a sort of self–delusion, such that the mere conscious awareness of the sense of being immersed may be sufficient to destroy it. Thus, if we are to develop methods to evaluate whether or to what degree a person has become successfully immersed in an induced reality, whether one induced by reading a book or by the latest headset based virtual environment, those methods must necessarily be covert and unobtrusive, and applicable

---

* Correspondence to: Systems Neuroscience & Neurotechnology Unit, Faculty of Medicine, Saarland University & School of Engineering, htw saar, Building 90.5, D-66421 Homburg/Saar, Germany.

E-mail address: daniel.strauss@uni-saarland.de (D.J. Strauss).

[1] These authors contributed equally to the paper.

https://doi.org/10.1016/j.brainresbull.2024.110923
Received 10 May 2023; Received in revised form 19 November 2023; Accepted 6 March 2024

without interfering with the immersive state itself. Note that we use the term ‘induced reality’ to cover virtual, augmented, and mixed realities as well as being simply immersed in reading a book or a ‘daydream’.

Methods of evaluating immersion must focus on attention. In a recent, comprehensive attempt to define immersion in the context of audiovisual experiences, Agrawal et al. (2020) defined immersion as “the state of deep mental involvement in which the subject may experience disassociation from the awareness of the physical world due to a shift in their attentional state”. In addition to Agrawal et al. (2020), several other scientists have highlighted the prominent role of attention in immersion, e.g., see (Murray, 2017; Thon, 2008; Sanderson et al., 2010; Souza and Naves, 2021), linking research on immersion to established concepts in cognitive neuroscience. Here, we start from an attentional emphasis similar to that of Agrawal et al. (2020), but build out details of the model with the goal of using predictions from existing cognitive neuroscience models of attention to guide research on immersion. For example, the assumption that immersion requires the projection of attention to an induced reality would suggest that processing resources must be allocated (see Kahneman, 1973) to information associated with that reality, and that physiological measures associated with attentional allocation could best be employed to assess immersion as well.

2. Different modes of attention in immersion

In order to model the attentional properties important to understand immersion, we similarly begin with the idea that immersion requires being focused on a task in such a way that one loses awareness of events or sensations outside of it Agrawal et al. (2020). The increased awareness of one stimulus (or set of stimuli) at the expense of others is, of course, a hallmark of simple selective attention. Thus this preliminary characterization of immersion could, in principle, be accommodated entirely in terms of models developed to account for selective attention and ‘inattentional blindness’ (Simons, 2000; Mack and Rock, 1998; Lavie, 2005). Nevertheless, there are surely other qualities besides selective attention that must be present when one is truly immersed in an induced reality. Evidence that many researchers are aware of this additional quality can be found in literature addressing the importance of distinguishing immersion from related terms used mainly in virtual reality research, in particular, presence (Slater et al., 1994; Jennett et al., 2008; Tamborini and Skalski, 2006; Michailidis and He, 2018), transportation (Green and Brock, 2000; Van Laer et al., 2013), flow (Michailidis and He, 2018; Swann et al., 2012), and envelopment Berger et al. (2009), see Agrawal et al. (2020) for detailed discussions. It is beyond the scope of this article to evaluate the degree to which immersion does or does not differ from these various other qualities; the authors in (Agrawal et al., 2020) did already an excellent analysis of these differentiations.

However, it is important to note that one factor that is consistently cited as distinguishing immersion from simple selective attention is the importance of self-initiation and intrinsic motivation in immersion, see (Deci and Ryan, 2000; Ryan and Deci, 2000; Richter, 2013; Kurzban, 2016; Herrmann and Johnsrude, 2020). We suggest that the concept of intrinsic motivation is particularly relevant here, such that immersion is strengthened when the motivation to attend to an induced reality is supported specifically by task- or narrative-related properties of the activities being conducted in that reality. That is, there seems to be more to immersion than simply the intentional, self-initiated direction of attention toward the sensory streams of that reality. Immersion requires more than just the will or the intention to attend to some stimuli more than others, because that is simply paying attention. We propose that immersion requires attention to be directed intentionally to specific mental representations (memories, narrative structures, schemas, etc.) that must be associated with a particular set of coordinated stimuli existing within the induced reality because those stimuli *fit together* in some particularly compelling way.

Indeed, if we are to develop a model of attention in immersion that can be applied broadly to a wide variety of immersive situations, we must also consider circumstances in which one becomes immersed in a wholly internal reality, as in “daydreaming.” Even in the absence of any novel sensory input, it is still possible to become immersed in an induced reality that consists entirely of purely self–generated internal objects and that therefore depends to an extreme degree on the deployment of so-called ‘internal attention’ (Chun et al., 2011; Strauss and Francis, 2017). Such cases highlight the first of two important distinctions between modes of attention. Internal attention refers to attention directed toward internally generated mental representations such as schemas and narratives. In such extreme cases, internally directed attention is all there is to support immersion. In the technologically more dominant case of sensory–driven immersion – whether as simple as visual input from reading a electronic book or as advanced as using multisensory virtual, augmented, or mixed reality high–fidelity technologies – the induced reality may be established by immersion–provoking sensory events but we argue that the sense of immersion in that reality still likely requires the development, maintenance, and attention to internal representations as well.

Attention in Sensory Dynamics: Therefore, to cover the sensory dynamics when dealing with a sensory–driven reality as well as the internal processes related to the narrative and schemas of the induced world, e.g., the unfolding of the story (Ryan, 2003) (and potentially also the self-generated/imagined sensory inputs of an internally-generated world, e.g., from daydreaming or reading), our model employs two different but established taxonomies of attention. These attention taxonomies are integrated in a conceptual model that also incorporates affect (as any experience of emotion or feeling), effort, and motivation to pursue goals in an induced reality (in line with (Deci and Ryan, 2000; Ryan and Deci, 2000)) enabling future consideration of the role of narrative engagement (in line with Albrecht and O’Brien, 1993; Green et al., 2004; Bilandzic and Brusselle, 2017).

Although we have already mentioned the importance of internally directed attention in maintaining a sense of immersion, when discussing sensory dynamics it is simpler to start with the traditional taxonomy of exogenous and endogenous of attention in sensory processing, see (Müller and Rabbitt, 1989; Spence and Driver, 1994; Berger et al., 2005; Sani et al., 2020; Jigo et al., 2021; Ren et al., 2021). Exogenous or bottom–up attention to the respective sensory modality is driven by the physical properties and saliency of the stimuli. Such properties might include location, intensity, and abruptness of onset in any domain, as well as more sense-specific properties such as tonality or roughness in the auditory domain and color or brightness in the visual domain. It is important to note that these properties should influence the distribution of attention irrespective of whether the stimulus to which they are attributed is perceived in the actual or induced reality. Exogenous attention is involuntary and comparatively automatic. In contrast, endogenous attention refers to a (goal–driven) top–down attention which is voluntarily allocated to an object or stream of interest. Endogenous attention is closely linked to intention, motivation and thus goal pursuit in a broader sense, e.g., see Deci and Ryan (2000). Which sensory information stream is within the attentional focus thus depends on a combination of exogenous and endogenous factors and might be modeled by a probabilistic stream selection model Trenado et al. (2009). Here different exogenous and endogenous weights are assigned to the individual sensory streams and define their probability for being selected, i.e., being within the attentional focus. This probabilistic selection scheme can be seen as computational approach to the biased competition model which is well-known in visual perception, see Desimone and Duncan (1995), and has also been discussed for the auditory modality by Shinn-Cunningham (2008).

The exogenous and endogenous components of the model reduce the selection of sensory streams in an actual or induced reality to the abstract notation of exogenous and endogenous weights. However, to understand the role of attention in immersion, and, in particular, to understand what drives attention to induced realities, we need to have a closer look at the concepts that define these weights. For this, we apply the taxonomy of Chun et al. (2011) of internal and external attention which is based on the types of information that attention operates over. External attention is associated with the selection and modulation of sensory information, i.e., information originating from outside the observer. It can thus be thought of as similar in some ways to exogenous attention, but, importantly, can also be endogenously directed. For example, external attention may be directed by endogenous systems when directing attention to a particular location in an otherwise static visual scene, e.g., when waiting for a stoplight to change color. In this case, the observer is directing the distribution of attention – this is endogenous attention – but it is being directed toward an external stimulus. Internal attention, in contrast, refers to the selection, modulation, and maintenance of internally generated information such as task rules, schemas, responses, and elements of long–term or working memory Chun et al. (2011). It is this internal attention to narrative structures, schemas, and mental representations related to the coherence of the induced reality that we have already identified as being so crucial for maintaining the state of immersion.

Even though the exogenous/endogenous and internal/external modes of attention are related, see Chun et al. (2011), they must be distinguished if we are to map out the role of different processes when paying attention to an induced reality, especially in a conceptual, let alone a computational implementation. Whereas the exogenous/endogenous taxonomy directly reflects the sensory dynamics of the competition between the information present in each of the two realities, the internal/external taxonomy is more relevant to understanding the importance of narrative and conceptual factors such as the unfolding of a story line while reading a book, the 'willing suspension of disbelief' when interacting in a world with, perhaps, unrealistic physics or events, and even the case of daydreaming in the absence of any sensory input. Thus internal attention is closely linked to 'being in harmony with the induced reality' and goal pursuit, e.g., see Deci and Ryan (2000) and provides, even in sensory–driven immersion, a framework to analyze higher–order, i.e., not just reflexive, interactions with affect as we will see below.

The "what" and "why" of goal pursuits in humans is a long lasting debate, see Deci and Ryan (2000). Why somebody should seek goals in an induced reality is certainly highly individual and context sensitive. The prominent role of the narrative in immersion has been stressed by several researchers, e.g., see (Ryan, 2003; Adams and Rollings, 2006; Ermi and Mäyrä, 2005; Thon, 2008). More accessible than a goal itself in our analysis, is the motivation in pursuing this goal and the acceptance of the narrative. We assume that paying attention to the induced reality has to be rewarding (in a broader sense, i.e., in terms of achieving a goal, whether that is to have fun, gain information, communicate with others, etc.). However, we assume that there is always a competition between the induced and the actual reality as the latter has an inherent personal relevance. From a purely endogenous point of view, if there is no ongoing motivation to pursue a goal in the induced reality, the attentional focus must sooner or later shift back to the actual reality as sensations from that reality begin to intrude (if nothing else, interoceptive sensations related to posture, hunger, thirst, fatigue, etc.). Exogenous processing must also play a role, particularly in terms of reflexive attention shifts. For instance, aversive and abrupt sound stimuli that immediately capture our attention seem likely override voluntary attentional processes and their respective endogenous weights, respectively.

We have previously considered the importance of exogenous attentional capture in the context of distraction and annoyance from distracting sounds (Francis and Love, 2020). Indeed, distraction seems like a reasonable characterization of the involuntary shift of attention from the induced reality stream(s) back to the actual reality under the influence of a powerful real-world stimulus such as the sound of a door slamming or perhaps the interoceptive feeling of hunger or thirst. It is thus also possible to model the reciprocal case in the same way, such that when a stimulus from the induced reality is sufficiently strong to draw attention (back) toward that domain, it acts as a 'distractor' from the real world. However, it seems likely that such purely exogenous shifts of attention might serve mainly to interrupt immersion and would be unlikely to result in immersion on their own. In order to support immersion in an induced reality, a set of stimuli must be sufficiently compelling to hold the focus of attention to the exclusion of stimuli from outside that reality, and we would argue that for the most part this would require at least some degree of endogenous attention to (at least) some internal representations associated with that reality. Indeed, we can conceive of cases in which one's attention is drawn back again and again to an immersive experience, for example when one cannot stop thinking about a compelling novel or engrossing film. One might argue that such cases represent something other that true immersion, such as perhaps presence or transportation. However, if we do think of them as representative of immersion to some degree, which seem typically to involve a high degree of attention to internal representations, especially narrative structures and conceptual schemas, these cases further support our argument that immersion depends heavily on attention to internal representations.

3. Conceptual taxonomic model

In this section, we establish a conceptual taxonomic model to study immersion using the aforementioned concepts. To do so, we introduce a simple mathematical framework which might support future experimental designs by quantitative means. We start with an attentional focus vector $\mathbf{a} \in \mathbb{R}^2$ (we use bold letters for vectors) in a two dimensional plane with the actual reality and the induced reality as different dimensions, see Fig. 1(a). Orthogonality assures that the mapping to the induced vs. the actual reality are mutually exclusive. It is worthwhile to emphasize that this 'attentional focus vector' is a simplified mathematical construct and we make no attempt at this point to relate it to (bio–)psychological substrates. It is simply used to provide a continuous (smooth) transition between the two–competing dimensions of resource allocation. Immersion requires now the projection of the attentional focus vector on the induced reality axis, i.e., the component of $\mathbf{a}$ which is represented on the ir–axis. We denote this component by $\mathbf{a_{ir}}$. Along this line, we denote the projection of the attentional focus vector to the actual reality, i.e., the ar–axis, by $\mathbf{a_{ar}}$, see Fig. 1(a). So we have that $\mathbf{a} = \mathbf{a_{ar}} + \mathbf{a_{ir}}$. Note that this can be written as $\mathbf{a} = a_{\mathrm{ar}}\mathbf{e}_1 + a_{\mathrm{ir}}\mathbf{e}_2$, where a_{ar} and a_{ir} is the magnitude (or length) of $\mathbf{a}_{\mathrm{ar}}$ and $\mathbf{a}_{\mathrm{ir}}$, respectively. The vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are the standard basis vectors of a two dimensional Cartesian coordinate system defining the competing directions of the actual and the induced reality. In a bistable perception, i.e., either you perceive the induced or the actual reality (see blue/green transition at $\varphi = \pi/4$ in Fig. 1 (left)), the model can also map a distraction framework (e.g., see McRae et al., 2010). For instance, for the assessment of 'immersion' in the induced reality, the projection a_{ar} represents the amount of distraction from the induced reality.

The vectorial model in Fig. 1 (left) represents how the attentional focus varies between the competing dimensions. The dominant stream in the individual reality is the momentary "winner" of a stream selection based on the exogenous and endogenous weights to several input streams. A probabilistic stream selection for this is described in Trenado et al. (2009) but an alternative is suggested below. In module 1 below, we will first have a closer look at the dynaics of the attentional focus between these "winners" or dominant streams in the respective reality and their associated weights.

3.1. Module 1: exogenous and endogenous sensory dnamics

The direction of the attentional focus vector $\mathbf{a}$ is now given by the exogenous and endogenous weights of these dominant streams in the individual realities. In dynamic regimes, the vectors as well as the

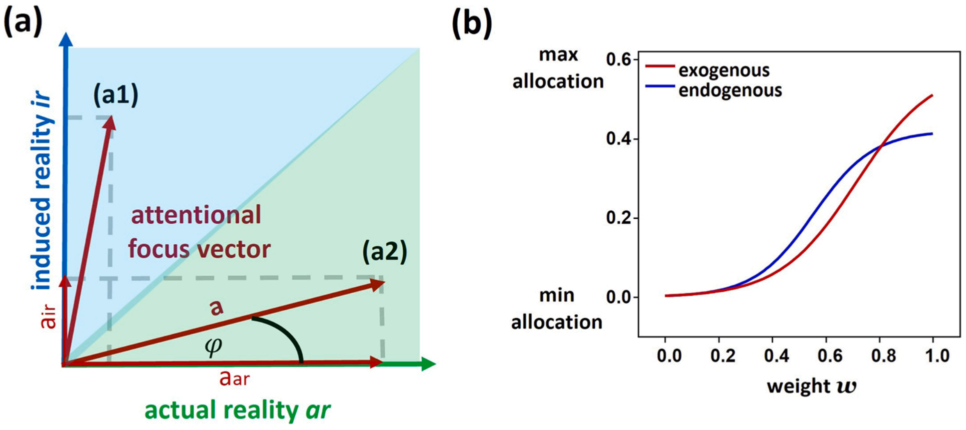


Fig. 1. (a) The two–competitor model along the orthogonal axes of the actual and the induced reality. In state (a1) the projection of the attentional focus vector is mainly on the induced reality which 'wins' the competition; a necessary condition for a high 'immersion'. In state (a2) the attention is mainly on the actual reality making it the 'winner' of the competition. The decomposition of the attentional focus vector in its component a_{ar} and a_{ir} is also shown. (b) The transfer function $\Omega_{\alpha,\beta,\gamma}(x)$ for $x \in [0, 1.2]$ and $\alpha = 0.58$, $\beta = 7$, $\gamma = 5$ (exogenous) and $\alpha = 0.42$, $\beta = 9$, $\gamma = 5$ (endogenous).

weights depend on time $t \in \mathbb{R}$. We denote the exogenous and endogenous weights in $\mathbb{R}$ by $w^{\mathrm{exo}}_{\mathrm{ar}}(t)$ and $w^{\mathrm{endo}}_{\mathrm{ar}}(t)$ for the actual reality and by $w^{\mathrm{exo}}_{\mathrm{ir}}(t)$ and $w^{\mathrm{endo}}_{\mathrm{ir}}(t)$ for the induced reality. We define a transfer function $\Omega : \mathbb{R}^+ \mapsto [0, \alpha]$ to map the (possibly not normalized) weights (see below) to a fixed interval $[0, \alpha]$, where $\alpha \in \mathbb{R}_{>0}$ is the maximum resource allocation. Here we use the sigmoid function $\Omega_{\alpha,\beta,\gamma}(\cdot) = \frac{\alpha}{1+e^{-\beta\cdot-\gamma}}$ but other possible choice are, for instance, $\Omega_{\alpha,\beta}(\cdot) = \alpha\tanh(\beta\cdot)$ or $\Omega_{\alpha}(\cdot) = \min\{\cdot, \alpha\}$ $(\alpha, \beta, \gamma \in \mathbb{R}_{>0})$. Here β and γ determine the transfer sensitivity of the mapping, see Fig. 1(b).

There is a long-standing debate in the literature over how the exogenous and endogenous mechanisms of attentional allocation interact, especially across modalities, e.g., see (Müller and Rabbitt, 1989; Spence and Driver, 1994; Berger et al., 2005; Sani et al., 2020; Jigo et al., 2021; Ren et al., 2021). Some models suggest that there should be a stronger interaction (i.e., larger weights in our model) under high demand due to a competition for resources, see Berger et al. (2005). While we make explicit choices here for the purpose of developing a working computational model, we acknowledge that further research is likely to suggest the need for more detailed refinements in this regard. For example, due the evolutionary significance for survival of the auditory modality for automatic orienting in primates (e.g., Carretié, 2014; Strauss et al., 2020; Olszanowski et al., 2023, see also "Advantages & Costs of Front–Facing Eyes" in Allman, 2000), we give priority to the exogenous channel to 'override' endogenous weights in extreme conditions, e.g., due to abrupt, very intense sounds. Note that the parameters used in Fig. 1(b) allow for this 'exogenous override'. However, this weighting could be modified depending on the specific method for inducing immersion, i.e., sound-damping headphones, goggles, etc. For the present model, we use different transfer functions $\Omega_{\alpha,\beta,\gamma}(\,\cdot\,)$ for the exogenous and endogenous channels, see Fig. 1(b) for examples. When using the notation α', β', γ' for the parameters of the endogenous transfer function, the individual attentional projections are given by

$$a_{\mathrm{ar/ir}}(t) = \Omega_{\alpha,\beta,\gamma}\left(w^{\mathrm{exo}}_{\mathrm{ar/ir}}(t)\right) + \Omega_{\alpha',\beta',\gamma'}\left(w^{\mathrm{endo}}_{\mathrm{ar/ir}}(t)\right).$$

Using these projections, the magnitude (or length) of the attentional focus $\mathbf{a}(t)$ is given by $m_a(t) = \|\mathbf{a}(t)\|_2 = \sqrt{a_{\mathrm{ar}}(t)^2 + a_{\mathrm{ir}}(t)^2}$ with the corresponding angle

$$\varphi_a(t) = \tan^{-1}\frac{a_{\mathrm{ir}}(t)}{a_{\mathrm{ar}}(t)}.$$

Then, in this two–competitor model, φ_a defines the projection of attention to the induced reality, i.e., fully focused for $\varphi_a = \pi/2$ and not focused at all for $\varphi_a = 0$.

3.2. Module 2: external vs. internal attention and affect

In order to assign the weights $w^{\mathrm{exo/endo}}_{\mathrm{ar,ir}}$ to sensory (or internally generated) streams or objects, we employ a conceptual link inspired by the taxonomy of internal versus external attention of Chun et al. (2011) as a more (psychologically) representative alternative to the probabilistic stream selection in (Trenado et al., 2009). The demands for a generally valid model of these attention modes in immersion are of course immense as, especially regarding internal attention, they crucially depend on many factors including the induced reality type (computer games, educational settings, virtual meetings, entertainment etc.), the nature and coherence of the narrative, the participant's personal interest, and several other individual factors. As a first step, we suggest an orthogonalization of these attention modes which we have used in Strauss and Francis (2017) to analyze effortful listening. Here the resulting attentional vector is used to define an (attentional) effort response in the sense of Sarter et al. (2006). This conceptual model is then further coupled with affect in a closed–loop system that includes related modules such as motivation and reward by Schneider et al. (2019) to represent cognitive fatigue effects in coordination with the effort response. The interaction of resource allocation and affect was also carefully analyzed in Francis and Love (2020) and Herrmann and Johnsrude (2020). Models of this type are not necessarily restricted to effortful tasks in the induced reality and can be generalized to provide an endogenous response reflecting affect related factors. In fact, when using a proper normalization, this affect-dependent endogenous response may directly define, at least in a first approximation, the weights $w^{\mathrm{endo}}_{\mathrm{ar,ir}}$ in Module 1 as described earlier.

For instance, when losing the motivation to solve an effortful task in the induced reality (e.g., the difficulty of an 'immersive' computer game is too high, resulting a bad effort/reward ratio), the endogenous weight will decrease over time. But the very same decrease in motivation happens when watching a boring movie over a longer time, even though this task not necessarily effortful. For the sake of simplicity, for the moment we represent models of this type by a multivariate map $\Psi_{\mathbf{p}} : \mathbb{R}^m \mapsto \mathbb{R}$, representing the endogenous response to m sensory or internally generated streams or objects, denoted by $\mathbf{s}(t) = (s_1(t), s_2(t), \ldots, s_m(t))^T$, and specified by a set of parameters $\mathbf{p} \in \mathbb{R}^n$. Thus, the weights $w^{\mathrm{endo}}_{\mathrm{ar,ir}}$ are simply given by

$$w^{\mathrm{endo}}_{\mathrm{ar,ir}}(t) = \Psi_{\mathbf{p}}(\mathbf{s}(t)), \quad t \in \mathscr{T}.$$

Here $\mathscr{T}$ denotes the interval of the simulation time. Also a multisensory integration (see Sec. 4) may be included in this function (or module) as a missing multisensory fusion or a lack of concurrency of the sensory inputs (see Calvert et al., 2004). Such a feature might reflect the damage to immersion that might be caused by a low–fidelity system that generates a (multisensory) induced reality in which the different sensory streams are poorly coordinated in time (think of a poorly dubbed movie in which the movement of the lips does not line up well with the sounds of the dubbed speech). Such mis-synchrony may well reduce the feeling of presence (Nilsson et al., 2016) or other quality important for the immersive experience, see the discussions in Marucci et al. (2021), and would be reflected in a loss of endogenous attention directed toward internal mental representations (schemas, narrative structures, etc.) associated with the coherence of the reality, that is as smaller endogenous weight $w_{\mathrm{ir}}^{\mathrm{endo}}(\cdot)$, see Sec. 4.3 for further discussions.

To model the exogenous weight, we define the parameterized, multivariate function $\Phi_{\mathbf{p}}: \mathbb{R}^m \mapsto \mathbb{R}$. It reflects the attention directed toward sensory stimuli based on their perceptual, and affective salience, e.g., brightness, loudness, valence and arousal. Higher order features, e.g., computationally extracted voice features (Eyben et al., 2010; Trinh et al., 2022) or facial features (Flotho et al., 2022) in case of overt attention can also be integrated in the model $\Phi_{\mathbf{p}}$ (see Sec. 4). Thus we have that

$$w_{\mathrm{ar,ir}}^{\mathrm{exo}}(t) = \Phi_{\mathbf{p}}(\mathbf{s}(t)), \quad t \in \mathscr{T}.$$

The entire two-competitor model with these maps and the sensory input is shown as schematic diagram in Fig. 2.

Exemplary implementation of an effortful, fatiguing task in the induced reality: As an example, we present an implementation of an effortful task in the induced reality such as playing a computer game with an increasing level of difficulty. For this, we focus on the map $\Psi_{\mathbf{p}}(t)$ which is implemented using the fatigue model in Schneider et al. (2019) in which the sensory input streams $\mathbf{s}$ are represented directly in terms of their internal and external demands[2]. The effort response to these demands (Strauss and Francis, 2017; Schneider et al., 2019) defines the endogenous weight $w_{\mathrm{ar}}^{\mathrm{endo}}$ whereas the other weights are modeled by simple oscillatory functions in time with a transient peak (the slamming door, see below), see Fig. 3. However, these exogenous weights can be easily adjusted to the task at hand. In Fig. 3 (1. row) different (actual) motivation profiles are shown. On the left, playing the game is rewarding (good reward to effort ratio Schneider et al., 2019) and we have a rather constant motivation over the 20min time of playing. On the right–hand side, the skills of the player cannot catch–up with the difficulty, resulting in a bad reward to effort ratio, fatigue, and thus a decease in actual motivation over time up to the point of giving up completely in the end. The resulting endogenous weight mapped by $\Omega_{\alpha,\beta,\gamma}(\,\cdot\,)$ is shown below (blue line) along with the other mapped weights using simple oscillatory functions as model. Note that around 10min there is a short intense exogenous activation in the actual reality reflecting a sudden, very loud sound such as a slamming door (black line). The endogenous weights directed toward the actual reality are moving up over the time in this model because we assume that there is necessarily increasing pressure to do something else in the actual reality, if only due to the biological necessity of survival. The increasing difficulty in the game is, in contrast, associated with a more rapid exogenous stimulation in the induced reality which is modeled by the (up–)chirp function. The resulting magnitude $m_a(t)$ and angle $\varphi_a(t)$ are shown below. It is noticeable that φ_a reflects the competition between the individual realities for the allocation of attention.

The quantitative model in summary: In our two–competitor model, sensory streams emanating from the induced reality compete with those from the actual reality, i.e., the physical world around us and inside us, for the allocation of our attention. The induced and the actual reality are represented as orthogonal axes in a two dimensional plane in which an attentional focus vector is mapped. The model consists of two modules. Module 1 defines the projection of the attentional focus vector to the actual and induced reality by accumulating the exogeneous and endogenous weights to stimuli in the respective reality. These weights are proportional to the attentional allocation in the respective reality. The resulting two projections, i.e., to the actual and induced reality, define the attentional focus vector and the corresponding angle between this vector and the respective reality. The closer this angle is to 90° (which is $\pi/2$ in radians), the maximum angle in the defined setting, the more attention is allocated to the induced reality. So, everything depends on the weights. The weights are defined by module 2 using the taxonomy of internal and external attention. In fact, we defined a (generic) map which maps the sensory streams in the respective reality to the exogeneous and endogenous weights using the internal and external mode of attention. As presented, this map can be any model which provides this, though an explicit, generally valid formulation is beyond the scope of this paper. However, by employing a model of cognitive fatigue which links internal and external attention to reward and motivation (i.e., affect-related quantities) from (Schneider et al., 2019), we provided an example for such a map.

4. Discussion

Although preliminary in terms of detail, the present model provides a conceptual framework linking the neuroscience of attention to fundamental and applied immersion research. As such, it should provoke new, more specific research questions related to the role of attention in immersion and its interactions with affect. In addition, we hope that it also stimulates researchers in cognitive and affective neuroscience to apply their knowledge to immersion research. To that end, in this section we present some preliminary observations from the model that suggest some potential avenues for future research.

4.1. Optimizing immersion in induced realities: internal attention, motivation, and affect

In this subsection, we outline some ways in which the present model might contribute to studying the optimization of immersion; an important problem particularly in the development of more effective immersive technologies.

For example, the presented two–competitor model maps the role of attention in immersion using two distinct but related taxonomies of attention, and these must be considered together when optimizing immersion. To optimize immersion, the angle $\varphi_a(t) \in [0, \pi/2]$ has to be maximized in the two–competitor model. Conceptually, the model was designed such that it can in principle cover the full range of examples of immersion discussed in the literature, from being 'immersed' in a daydream or in reading a book to the more commonly considered case of multimodal immersion in the metaverse projected using high fidelity audiovisual technology. Reflecting the diversity of these examples in terms of the diversity of attentional systems, this model is capable of achieving angles close to $\pi/2$ (i.e., 90°) (representing a very high degree of immersion) in very different ways. In the purely narrative driven case of reading a book, an interesting narrative absorbs your internal resources and creates an induced reality, e.g., by unfolding a story in your mind. In the technology driven case, a highly detailed array of multisensory information stimulates a complex interplay of internal attention to the narrative and external attention to multiple sensory channels, driving a highly dynamic exogenous and endogenous allocation of attention. Whereas in the former case, internal attention and the associated interplay with motivation and affect drives immersion, in the latter case, exogeneous input and external attention play a crucial role as

---

[2] The exemplary model implemented in Matlab2022b/SIMULINK, MathWorks, Inc, Natick, MA, USA can be download from https://github.com/farahic/The-Role-of-Attention-in-Immersion-The-Two-Competitor-Model. Note that the absolute values in the distress transfer function were adjusted for the current setting as compared to Schneider et al. (2019)).

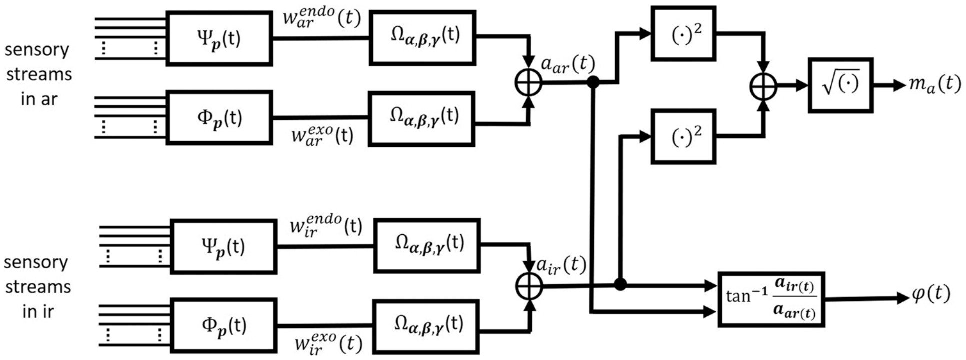


Fig. 2. A schematic diagram of the dynamic two–competitor model. The input is given by the sensory streams as well as defined by the parameters $\mathbf{p}$ of the functions $\Phi_\mathbf{p}$ and $\Psi_\mathbf{p}$. The output is the angle $\phi(t)$ representing how much attention is projected to the induced reality. The magnitude of the attentional focus vector is given $m_a(t)$.

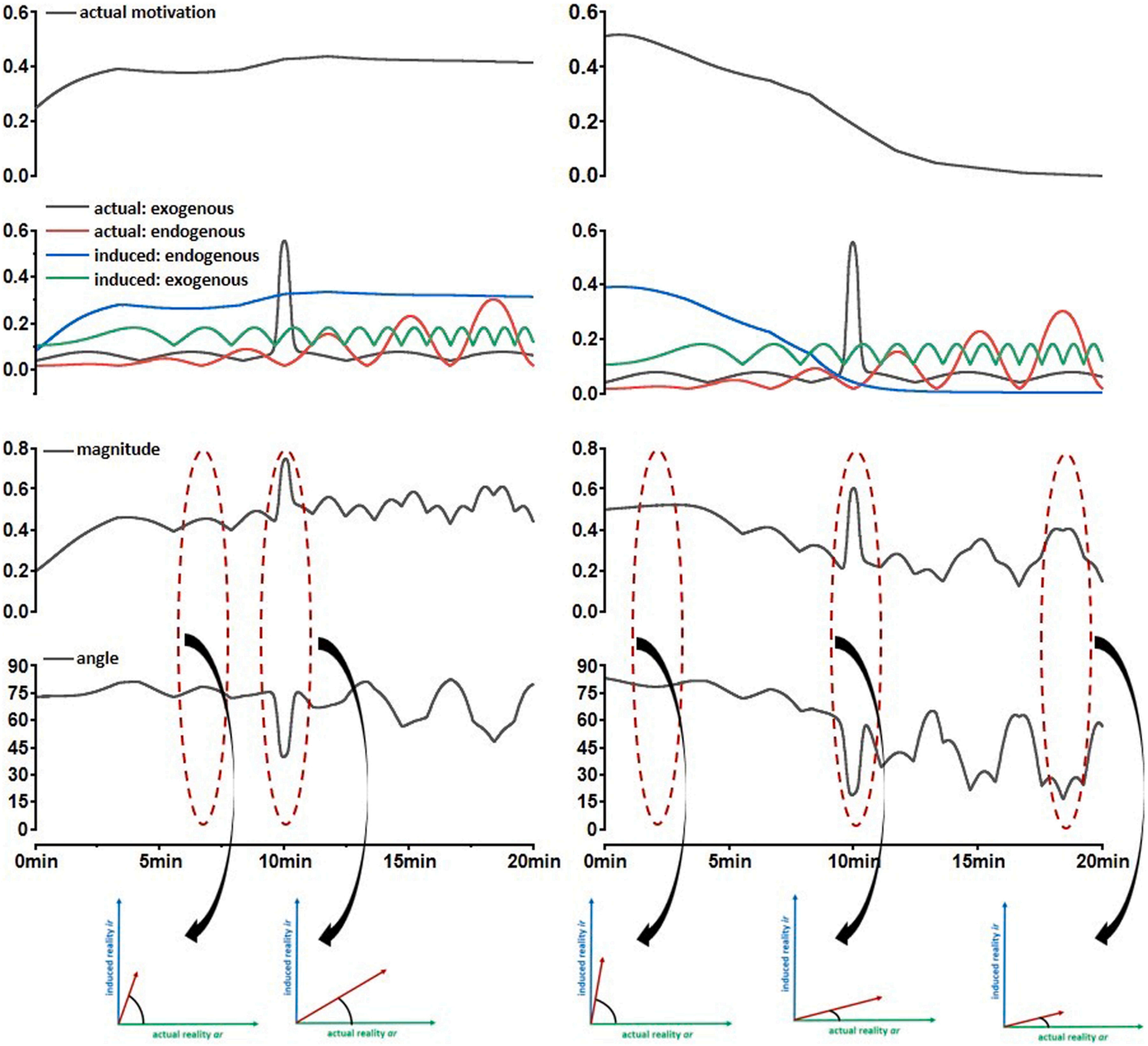


Fig. 3. The first row (top) shows the actual motivation profile for a rewarding activity (left) and a very effortful fatiguing activity (right). Second row: The weights mapped by $\Omega_{\alpha,\beta,\gamma}(\cdot)$; here $\Omega_{\alpha,\beta,\gamma}(w_{\mathrm{ar}}^{\mathrm{endo}})$ stems from the actual motivation above whereas the others are modeled by simple oscillatory functions. Third row: The magnitude of the attentional focus vector $m_a(t)$. Fourth row: The angle $\varphi_a(t)$ of the attentional focus vector. Beside the angles $\varphi_a(t)$ in degrees, all units are normalized such that 1.0 is the maximum allocation.

well. In summary, both the amount of attention (length of the arrow) that is being directed, and also the direction (angle) it is distributed toward, contribute to the degree of immersion, but equivalent values of these measures can in principle be achieved through different mechanisms.

In many ways, it seems that the technology driven approach would have more flexibility, or perhaps greater capacity, to create immersion, but this may only be superficially the case. Most obviously, modern "virtual reality" immersive systems can attract attention exogenously through visual, auditory, and even tactile stimulation simultaneously (though this may have disadvantages as well, see below). More generally, effects such as abrupt onsets of emotionally loaded stimuli may represent an excellent exogenous means to pull and keep someone's attention to an induced reality – not an unknown concept in the Hollywood film industry (e.g., the infamous "jump scares" of formulaic horror movies). However, it also seems likely that the effectiveness of such emotional stimuli at maintaining immersion, no matter how good they are initially attracting attention, will still depend in the long run on their consistency with the narrative and with observers' expectations, i.e., with mental representations that depend on the endogenous commitment of internal attention, and such narrative consistency can also be achieved through less technological means. Nevertheless, these cases highlight the close relationship between attention and intention (Snyder et al., 2000) or motivation, and thus also emotion (Lang et al., 1997) and associated psychophysiology.

4.2. *Objective assessment of immersion*

In this section, we review and discuss a few issues that arise when considering how we might use objective physiological measures to assess immersion in terms of the attentional and affective properties identified in the presented model. As a myriad of methods can be used for this purpose (e.g., electro– and magnetoencephalography, functional magnetic resonance imaging and near–infrared imaging), we focus mostly on widely available and ecologically valid electrophysiological assessment methods, in particular, electroencephalographic methods in the following. Such methods also allow an inverse fit of the presented forward model, e.g., to deduce endogenous and exogenous model parameters. Thus these measurements can also be complemented by instruments from psychophysics to track the exogenous components in the respective modalities.

Induced realities, mainly in the form of (visual) virtual reality concepts, are of increasing interest to researchers intending to enhance the ecological validity of their experimental designs in cognitive neuroscience and affective research, e.g., see (Beck et al., 2007; Wong et al., 2014; Reggente et al., 2018; Nicol et al., 2019; Marucci et al., 2021; Hofmann et al., 2021; Yu et al., 2022). However, for such methods to be broadly applicable, it is important for researchers to be able to verify that participants are indeed responding to stimuli in a (relatively) uniformly immersed condition. The assessment of immersion has been addressed by a variety of neuroscientific and neurotechnological measurement methods, e.g., to differentiate between concentration and immersion (Lim et al., 2019), to quantify the degree of presence (Clemente et al., 2014), or to characterize binaural sound immersion (Nicol et al., 2019). A very recent review by (Souza and Naves, 2021) has analyzed different electroencephalographic methods to assess attention, work load (see Matthews et al., 2015; Yu et al., 2015), and mental fatigue for possible applications in attention detection in virtual environments, focusing on technologically induced realities. In particular, these authors along with others (Vortmann et al., 2019) have highlighted the importance of being able to distinguish between the internal and external modes of attention in virtual and augmented reality environments when using neurotechnological measurements, a task further highlighted by the presented model.

Magnitude and Angle of Attentional Focus: The suggested two–competitor model, with its emphasis on distinct taxonomies of attention, highlights the importance of attention allocation between the competing realities but also across the internal/external and exogenous/endogenous dichotomies, distinctions which provide greater rigor for the design of experiments assessing the depth of immersion. In particular, the two–competitor model stresses that beside assessing the magnitude of the described attentional focus vector $m_a(t)$ to internal or external processes (related to manifold electrophysiological correlates of attention, e.g., see Näätanen, 1992; Hopfinger et al., 2000; Parasuraman and Wilson, 2008; Hillyard and Picton, 2011; Mangun, 2013; Magosso et al., 2021), it is also important to get the direction of attention represented by the angle $\varphi_a(t)$. A straightforward way to implement this is to assess the attentional factors not in total, i.e., the magnitude $m_a(t)$, but rather to specific events in the respective reality, i.e., its components $a_{\mathrm{ar}}(t)$ and $a_{\mathrm{ir}}(t)$. In other words, if we know about the occurrence of events in a particular reality, we can assess the attention paid to them as well as any affective response(s) associated with them. This requires the co–registration of several measurements, including evoked, event–related, and induced electroencephalographic measures which can be associated to 'events' (identifiable fluctuations in ongoing measurements) in the respective reality, e.g., see Pfurtscheller and Lopes da Silva (1999); Mangun (2013), a methodology that can also be referred to as entrainment assessments (see Lakatos et al., 2008; Horton et al., 2013). Electroencephalographic measures of external (selective) attention paid to short auditory, visual, or somatosensory events can be applied for such purposes, e.g., see (Hillyard et al., 1973; Hansen and Hillyard, 1980; Hillyard et al., 1998; David et al., 2006; Lehser et al., 2018; Hillyard and Picton, 2011; Mangun, 2013). Within the respective reality, more specific types of attention can be assessed by neuroimaging techniques such as feature based, spatial or temporal attention, see (Luck and Hillyard, 1994; Giesbrecht et al., 2003; Olivers, 2008; Hillyard and Picton, 2011; Mangun, 2013) to further analyze the attentional dynamics within the modality if necessary. Even incongruities between internally generated patterns/predictions and the attended sensory streams can be assessed with such electroencephalographic methods and oddball paradigms, see, e.g., (Kutas and Hillyard, 1980; Polich, 2007); thus such methods are closely linked to internal attention and the narrative processing of the induced reality. Research of this sort is particularly well-represented in the auditory modality, where electroencephalographic temporal response function or stimulus reconstruction methods are frequently used to assess to which speech stream external attention is directed to (Mesgarani and Chang, 2012; Power et al., 2012; Wong et al., 2018; Schäfer et al., 2018). Recent work has also shown that a non–invasive surface electromyogram of the vestigial auricular muscles can be used to decode the direction of exogeneous and endogenous attention to short sounds and competing speech, see Strauss et al. (2020).

While directing selective attention to exogenous events in the induced reality is certainly a significant condition for immersion in the typical sensory–driven "virtual" reality, as we have suggested, internal attention must also play a necessary role. Research on assessment of internal attention has tended to focus on α–power (Ray and Cole, 1985), but other measures may be useful, including pupillometry (Unsworth and Robison, 2018), and future research should also not rule out combining attention assessments with physiological measures of orienting and affect (see Valenza and Scilingo, 2014; Marucci et al., 2021) or other autonomic nervous system responses, e.g., see (Valenza and Scilingo, 2014; Francis and Oliver, 2018), that might help link attention to the emotional qualities associated with the internal representations (narrative structures, schemas, etc.) relevant to the induced reality.

Detecting and Annotating Events: In order to quantify the angle $\varphi_a(t)$, a variety of measures from selective attention research combined with affective psychophysiology can be used, but it is critical to be able to link events of interest in a particular reality and modality (whether these events are strictly sensory, or rather significant from a more narrative or otherwise internal perspective) to measurable physiological events. This may seem trivial when considering more traditional experimental contexts in which the two "realities" across which attention may be divided are relatively simple (i.e., a mostly static visual scene containing two competing objects). However, this task becomes markedly more difficult when considering more advanced scenarios such as when the induced reality is emulating properties of an actual (physical) reality (e.g., in a 3D or 4D movie clip with immersive sound), or cases of augmented reality in which the induced stimuli are presented in direct competition with the full richness of the real world in which the participant is completely engaged. In such cases, the identification or annotation 'by hand' of physiologically relevant events in each reality for assessing degree of immersion can be impractical or even impossible, but recent computational efforts in machine learning, computer vision, and affective computing might help to identify/anntoate such events in the sensory streams automatically, see (Eyben et al., 2010; Hausfeld et al., 2018; Trinh et al., 2022; Höfling et al., 2020; Flotho et al., 2022).

Closed Loop Systems: While we have focused here on the use of physiological measures to assess the degree of immersion or perhaps the effectiveness of particular immersive methods, it is also possible to conceive of systems that use such measures as part of the immediate construction and ongoing maintenance of an induced reality. For example, virtual reality body suits (e.g., the Teslasuit, VR Electronics Ltd, London, UK) already incorporate physiological measures from the wearer that could be used to modulate stimulus delivery in a contextually–specific manner. In applications in psychotherapy, adaptive VR can already be used to modulate the degree of stress imposed by a virtual scene (e.g. the height or narrowness of a virtual walkway above a pit in exposure therapy for fear of heights) Baker and Fairclough (2022). Similarly, in applications in which modulating the degree of immersion might be important for the rehabilitation outcome, e.g., see Georgiev et al. (2021), the proposed model could support close–loop systems to keep the immersion at the intended level, e.g., by sending an appropriate stimuli when a drift of attention away from the induced reality is detected (significant reduction of the angle $\varphi_a(t)$), but perhaps also reducing properties conducive to immersion if the level of anxiety is identified as being too high.

4.3. Multimodal integration

If we are to apply this model to understanding immersion in the context of existing, let alone future, virtual reality technologies, we must consider the case when multiple sense are stimulated to evoke the induced reality. In this article, we have remained vague about the specific nature of the stimuli in question, and have simply mapped abstract sensory information streams in the model which might or might not be induced by any one of several senses. However, multimodal integration likely plays a significant role in immersion, especially in technology driven induced reality settings.

A few decades ago, neuroscientists studied the world one sense at a time. Visual neuroscientists would study the world from a visual perspective and auditory neuroscientists would study the world from an auditory perspective. More recently, cognitive neuroscientists have realized that no sense operates in a vacuum. Instead, all senses interact in a process known as multisensory integration (Spence and Santangelo, 2009; Thesen et al., 2004; Calvert et al., 2004). Indeed, Stein and Meredith (1993) argued that there is no animal in which there is known to be a complete segregation of sensory processing. Several senses can summate to produce superadditive signals in the brain where $1 + 1 = 3$. Alternatively, if the senses add up in the wrong way, they can produce subadditive integration and the senses can detract from each other. This depends on factors such as the relative timing, spatial coincidence, and semantic congruency (e.g., pairing a duck quack with a picture of a dog) of the different unimodal sensory stimuli in the multisensory perception. Stein and Meredith (1993) said that "Integrated sensory inputs provide far richer experiences than would be predicted from their simple coexistence or the linear sum of their individual products." In a similar line to our discussion in Sec. 4.1, imagine Alfred Hitchcock's 'Psycho' without Bernard Herrmann's music or several of the latest hit movies without Hans Zimmer's compositions.

Based on the above principles and other principles such as sensory dominance, it may eventually be possible to determine the ideal integration of multisensory information for the strongest perceptual experience by considering stimulus aspects such as sensory modality, spatial location, relative timing, and other congruencies or incongruencies. In general, especially for the purposes of attentional capture, vision dominates the other senses, and it has been said that 70% of attentional capture is based on vision, 20% on audition and 10% the rest, see Zimmerman (1989). However, even when multiple sensory signals are equally accessible, which sense dominates in a particular context or task depends on many factors, including perhaps their different ecological roles. For example, as discussed, e.g., by Francis (2022), the human auditory system seems to have evolved to be particularly suited to the function of alerting, being sensitive to the occurrence of events in the environment over large distances and essentially in all directions (Murphy et al., 2017; Olszanowski et al., 2023), serving to quickly evaluate their relevance for subsequent action Murphy et al. (2016), and guiding action (including the direction of vision) appropriately, see Arnott and Alain (2011). In addition, audition can be more reliable for judging relative timing, and under certain circumstances can even overrule visual information, for example, creating two visual blinks out of one in the cross-modal double flash illusion (see Shams et al., 2000). Thus, competition between stimuli across different realities and different modalities can be quite complex, and further research will be necessary to determine how, or to what degree, information presented in different modalities can facilitate or interfere with the development and maintenance of immersion.

A variety of factors influence multisensory integration. For example, Spence and Santangelo (2009) showed that the distance of the multisensory stimulus from the observer is important with the brain expecting an audiovisual stimulus that is far away (e.g., lightning and thunder) to have the visual aspect arrive first, while this expected gap becomes smaller when the observer is closer to the stimuli. Similarly, Marucci et al. (2021) showed that task load matters for multisensory perception. When task load is high having combined visual, auditory, and tactile stimuli led to improved performance and an increased sense of immersion in the virtual environment compared to visual stimulation alone. For these reasons much effort has been devoted to integrating more senses into the VR experience. VR systems have been built that integrate vision, touch, audition, and smell (e.g., Sensync, Panagiotakopoulos et al., 2022) to produce optimally immersive environments. However, care must be taken to appropriately coordinate the information presented in each modality, in order to minimize conflicting cues and improve fidelity across modalities.

Thus, fidelity between multisensory stimuli is important and seems to enhance immersion. Thus, factors such as timing, spatial coincidence and semantic congruency play a large role for the immersive experience. Perceptually, audiovisual speech is typically far more understandable than audio or visual alone, but even relatively small asynchrony between visual (e.g., lip movements) and audio (speech sounds) signals can not only eliminate, but actually reverse this benefit (Van Wassenhove et al., 2007; Conrey and Pisoni, 2006). Semantically, a quacking dog might make us instinctively react and snaps us out of the immersion. Thus, it seems like our brain has a particularly strong reaction to incongruencies to our expectations. This mechanism could potentially carry over to other types of rules as well such as when our expectations for multisensory stimuli are violated. Integrating these concepts in the two–competitor model could be an interesting line of future research.

4.4. Limitations

The presented model is necessarily very preliminary, and much work remains to be done for both, to further elucidate the model and to assess the validity of certain predictions made based on it.

Specific Immersion Settings: The current two–competitor model is very general as our intention was to provide a model of attention in immersion that was broad enough to cover, at least preliminarily, circumstances as simple as reading a book to a far more technologically advanced settings using augmented and virtual reality technology. Because of this, certain scenarios have not been analyzed in depth. For instance, our discussion concentrates on the state of immersion in the induced reality *per se*, and we disregard the "wow-effect" likely to be associated with using a new advanced technology and digital presence for the first time. Moreover, we have ignored unique properties of specific technologies, e.g., certain virtual reality headsets may be designed to physically block exogenous stimulation from the actual reality. As currently written, the model is intended to highlight the importance of competition between the two realities (induced and actual) for exogenous and endogenous attention, and we have also emphasized the essential role of internal attention to competing streams or objects as a potential source of distraction. Future research could absolutely investigate the development of the transfer functions $\Phi_{\mathbf{p}}(\mathbf{s}(t))$ and $\Psi_{\mathbf{p}}(\mathbf{s}(t))$ for specific settings, e.g., reading, watching movies, or playing computer games, in fully virtual, augmented, or mixed realities as desired. We hope that the suggested two–competitor model paves the way in such directions and helps to specify the problems that arise by allowing them to be reduced to general functionalities mapped by explicit mathematical functions.

Interactivity and Motor Action: Technological approaches to induced realities allow for interactivity and motor action, a topic that has been studied in virtual reality based learning (Petersen et al., 2022) using the Cognitive Affective Model of Immersive Learning (CAMIL), see Makransky and Petersen (2021). Moreover, the interplay of attention and motor action as such is an active field of current research, see Song (2019) and some recent theories of attention have begun to suggest that a major role of attention is to subserve planning for action (Hommel et al., 2019; Allport, 2016). Indeed, if we consider that "the purpose of the human brain is to use sensory representations to determine future actions" Wolpert et al. (2003), it seems likely, that one way to increase the engagement of attention and hence immersion with an induced reality would be to increase (physical) interactivity with it as well Park et al. (2017).

Limits of a Two–Competitor Approach: Finally, in this article we have attempted to reconcile terminology and phrasing used in research on virtual, augmented, and mixed reality, which we have subsumed as induced reality, with that used in traditional attention research. This reconciliation has limits. For instance, we have tended to use a very broad conceptualization of immersion, to the extent that it would fit out definition to say that, when reading your tax report instead of a book, you might be 'immersed' in the sense that we have used here, i.e., paying attention to such a high degree that other stimuli are excluded. We might even expect that this state would exhibit many of the affective physiological states associated with such immersion, and yet the phrase 'induced reality' or 'virtual reality' (which we have tended to treat as the thing one is *immersed in*) might, for most of us unfortunately, not fit here. That is, it would seem incongruent to talk about being 'immersed in the induced reality' of one's tax report. Thus the suggested two–competitor model just maps the case in which an induced reality is competing with an actual reality. Ultimately, we believe that this problem returns us to the deeper questions of how or to what degree 'immersion' must be distinguished from simple selective attention, on the one hand, and more complex and cognitively engaged phenomena such as 'presence,' 'transportation,' or 'flow,' on the other (see Agrawal et al., 2020). We have suggested that the role of internal attention is key to investigating this question, but further research is clearly necessary to provide a clearer picture.

5. Conclusions

We have analyzed the role of attention in immersion based on the hypothesis that attention is a necessary but not sufficient condition for immersion. By utilizing different modes of attention, we have developed the two–competitor model.

In this model, an induced and the actual reality are represented as orthogonal dimensions competing for the projection of attention. The two–competitor model allows for a quantitative implementation with an easy graphical interpretation, and helps to highlight the important link between different modes of attention in studying immersion.

Even though the two–competitor model is a preliminary conceptual approach to study immersion, it is intended to provoke new, more specific research questions related to the role of attention in immersion and its interaction with affect and neurophysiological signals. In addition, we hope that it also stimulates researchers in cognitive and affective neuroscience to apply their knowledge to immersion research.

CRediT authorship contribution statement

Alexander L. Francis: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Daniel J. Strauss:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Farah I. Corona-Strauss:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Supervision, Visualization, Writing – review & editing, Validation. **Jonas Vibell:** Formal analysis, Investigation, Writing – original draft.

Declaration of Competing Interest

The authors have declared no competing interest.

Data Availability

No data was used for the research described in the article.

Acknowledgments

The first author is partially supported by the German Federal Ministry of Education and Research (BMBF), Grant 13FH050KX1 and the European Union (European Regional Development Fund, ERDF) / state of Saarland, Germany, via the Project "Multi-Immerse". The authors would like to thank Kyra GauthierDickey and David Thinnes for early discussions and technical support. The first author would like to thank also Steven A. Hillyard for his helpful feedback to some early thoughts on the topic.

References

E. Adams and A. Rollings.Fundamentals of Game Design.Pearson College Div, Glenview, 2006.

Agrawal, S., Simon, A., Bech, S., Bærentsen, K., Forchhammer, S., 2020. Defining immersion: literature review and implications for research on immersive audiovisual experiences. Audio Engineering Society International Convention 68, 404–417.

Albrecht, J.E., O'Brien, E.J., 1993. Updating a mental model: maintaining both local and global coherence. J. Exp. Psychol. Learn. Mem. Cognit 19, 1061–1170.

Allman, J.M., 2000. Evolving Brains. Scientific American Library, New York.

Allport, A., 2016. Perspectives on Perception and Action, Chapter Selection for Action: Some Behavioral and Neurophysiological Considerations of Attention and Action. Routledge, pp. 395–415.

Arnott, S.R., Alain, C., 2011. The auditory dorsal pathway: orienting vision. Sci. Rep. 35, 2162–2173.

Baker, C., Fairclough, S.H., 2022. Adaptive virtual reality. Academic Press, pp. 159–176.

L. Beck, M. Wolter, N. Mungard, T. Kuhlenand W. Sturm.HCI and Usability for Medicine and Health Care, volume 4799, chapter Combining Virtual Reality and Functional Magnetic Resonance Imaging (fMRI): Problems and Solutions, 335-348.Springer Berlin Heidelberg, 2007.

Berger, A., Henik, A., Rafal, R., 2005. Competition between endogenous and exogenous orienting of visual attention. J. Exp. Psychol. Gen. 134, 207–221.

Berger, M.J., Calhoun, D.A., Helzel, C., LeVeque, R.J., 2009. Logically rectangular finite volume methods with adaptive refinement on the sphere. Philos. Trans. R. Soc. Lond. A: Math. Phys. Eng. Sci. 367, 4483–4496.

Bilandzic, H., Brusselle, R.W., 2017. Beyond metaphors and traditions: exploring the conceptual boundaries of narrative engagement. volume 27. In: Linguistic Approaches to Literature: Narrative Absorption. John Benjamins Publishing Company, pp. 11–27. volume 27.

Biocca, F., Delaney, B., 1995. Immersive virtual reality technology. Communication in the Age of Virtual Reality. Lawrence Erlbaum Associates Inc.

M.R. Cagnina and M. Poian.How to compete in the metaverse: The business models in second life. Working Paper No. 01-2007, U of Udine Economics, 2008.

Calvert, G.A., Spence, C., Stein, B.E., 2004. The handbook of multisensory processes. editors.. MIT Press.

Carretié, L., 2014. Exogenous (automatic) attention to emotional stimuli: a review. Cogn. Affect Behav. Neurosci. 14, 1228–1258.

Chasid, A., 2017. Imaginative content, design-assumptions and immersion. Review of Philosophy and Psychology 8, 259–272.

Cheng, Y., Wang, Y., Zhao, W., 2022. Shared virtual reality experiences during the covid-19 pandemic: exploring the gratifications and effects of engagement with immersive videos. Int. J. Environ. Res. Public Health 19, 5056.

Chun, M.M., Golomb, J.D., Turk-Browne, N.B., 2011. A taxonomy of external and internal attention. Ann. Rev. Psychol. 62, 73–101.

Clemente, M., Rodriguez, A., Rey, B., Alcañiz, M., 2014. Assessment of the influence of navigation control and screen size on the sense of presence in virtual reality using EEG. Expert Syst. Appl. 41, 1584–1592.

Conrey, B., Pisoni, D.B., 2006. Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. J. Acoust Soc. Am. 119, 4065–4073.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006. Dynamic causal modeling of evoked responses in EEG and fMRI. NeuroImage 30, 1255–1272.

Deci, E.L., Ryan, R.M., 2000. The "what" and "why" of goal pursuits: human needs and the self–determination of behavior. Psychol. Inquiry 11, 227–268.

Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. Annual Rev. Neurosci. 18 (1), 193–222.

Ermi, L., Mäyrä, F., 2005. Fundamental components of the gameplay experience: analysing immersion. Worlds in play: International Perspectives on Digital Games Research 37 (2), 37–53.

F. Eyben, M. Wöllmer, and B. Schuller.Opensmile: the Munich versatile and fast open-source audio feature extractor.In: Proceedings of the 18th ACM international conference on Multimedia, 1459-1462, 2010.

P. Flotho, H. Cosmas, G. Steidl, and D.J. Strauss Lagrangian motion magnification with landmark-prior and sparse pca for facial microexpressions and micromovements.In: Annu Int Conf IEEE Eng Med Biol Soc, volume 2022: 2215 2218, 2022.

Francis, A.L., 2022. Adding noise is a confounded nuisance. J. Acoust. Soc. Am. 152, e1375.

Francis, A.L., Love, J., 2020. Listening effort: are we measuring cognition or affect, or both? WIREs. Cogn. Sci. 11, e1514.

Francis, A.L., Oliver, J., 2018. Psychophysiological measurement of affective responses during speech perception. Hearing Res. 369, 103–119.

Georgiev, D.D., Georgieva, I., Gong, Z., Nanjappan, V., Georgiev, G.V., 2021. Virtual reality for neurorehabilitation and cognitive enhancement. Brain Sci. 11, e221.

Giesbrecht, B., Woldorff, M.G., Song, A.W., Mangun, G.R., 2003. Neural mechanisms of top-down control during spatial and feature attention. Neuroimage 19, 496–512.

Green, M.C., Brock, T.C., 2000. The role of transportation in the persuasiveness of public narratives. J. Personality Social Psychol. 79 (5), 701–721.

Green, M.C., Brock, G.F., Kaufman, T.C., 2004. Understanding media enjoyment: the role of transportation into narrative worlds. Comm. Theor 14, 311–327.

Hansen, J.C., Hillyard, S.A., 1980. Endogenous brain potentials associated with selective auditory attention. Electroencephalogr. Clin. Neurophysiol. 49, 277–290.

Hausfeld, L., Riecke, L., Valente, G., Formisano, E., 2018. Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. NeuroImage 617–626.

Herrmann, B., Johnsrude, I., 2020. A model of listening engagement (MoLE). Hearing Res. 397, e108016.

Hillyard, S.A., Picton, T.W., 2011. Electrophysiology of Cognition. John Wiley & Sons, Ltd, pp. 519–584.

Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. Science 182, 177–180.

Hillyard, S.A., Vogel, E.K., Luck, S.J., 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philos. Trans. R. Soc. Lond. BBiol. Sci. 353, 1257–1270.

Höfling, T., Tim, A., Gerdes, A.B.M., Föhl, U., Alpers, G.W., 2020. Read my face: automatic facial coding versus psychophysiological indicators of emotional valence and arousal. Front. Psychol. 11 https://doi.org/10.3389/fpsyg.2020.01388.

Hofmann, S., Mariola, F.K.A., Nikulin, V., Villringer, A., Gaebler, M., 2021. Decoding subjective emotional arousal from eeg during an immersive virtual reality experience. eLife 10, e64812.

Hommel, B., Chapman, C.S., Cisek, P., Neyedli, H.F., Song, J.H., Welsh, T.N., 2019. No one knows what attention is. Attent. Percept. Psychophys. 81, 2288–2303.

Hopfinger, J.B., Buonocore, M.H., Mangun, G.R., 2000. The neural mechanisms of top-down attentional control. Nature Neurosci. 3, 284–291.

Hopkins, E., 2022. Virtual commerce in a decentralized blockchain-based metaverse: Immersive technologies, computer vision algorithms, and retail business analytics. Ling. Philos. Invest. 21, 203–218.

Horton, C., D`Zmura, M., Srinivasan, R., 2013. Suppression of competing speech through entrainment of cortical oscillations. J. Neurophysiol. 109, 3082–3093.

Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A, 2008. Measuring and defining the experience of immersion in games. Int. J. Hum.-Comput. Stud 66, 641–661.

Jigo, M., Heeger, D.J., Carrasco, M., 2021. An image-computable model of how endogenous and exogenous attention differentially alter visual perception. Proc. Natl. Acad. Sci. 17, e2106436118.

Kahneman, D., 1973. Attention and Effort. Prentice–Hall, Englewood Cliffs, NJ.

Kurzban, R., 2016. The sense of effort. Curr. Opin. Psychol. 7, 67–70.

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.

Lakatos, P., Karmos, G., Metha, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320, 110–113.

Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1997. Attention and orienting: Sensory and motivational processes, chapter Motivated attention: Affect, activation and action. Erlbaum, Hillsdale, NJ, pp. 97–135.

Lavie, N., 2005. Distracted and confused?: selective attention under load. Trends in Cogn. Sci. 9, 75–82.

Lehser, C., Wagner, E., Strauss, D.J., 2018. Somatosensory evoked responses elicited by haptic sensations in mid-air. IEEE Trans. Neural. Syst. Rehabil. Eng. 26, 2070–2077.

Lim, S., Yeo, M., Yoon, G., 2019. Comparison between concentration and immersion based on EEG analysis. Sensors 19, 1669.

Luck, S.J., Hillyard, S.A., 1994. Spatial filtering during visual search: evidence from human electrophysiology. J. Exp. Psychol.: Hum. Percept. Perform. 20, 1000–1014.

Mack, A., Rock, I., 1998. Inattentional Blindness. MIT Press, Cambridge.

Magosso, E., Ricci, G., Ursino, M., 2021. Alpha and theta mechanisms operating in internal-external attention competition. J. Integr. Neurosci. 20, 1–19.

Makransky, G., Petersen, G.P., 2021. The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. Educational. Psychol. Rev. 33, 937–958.

Mangun, G.R., 2013. Cognitive electrophysiology of attention: Signals of the mind. Academic Press, San Diego, CA.

Marucci, M., DiFlumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E., Babiloni, F., Betti, V., Arico, P., 2021. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. Sci. Rep. 11, e4831.

Matthews, G., Reinerman-Jones, L.E., Barber, D.J., Abich 4th, J., 2015. The psychometrics of mental workload: multiple measures are sensitive but divergent. Hum Factors 57, 125–143.

McMahan, A., 2003. Immersion, engagement, and presence: a method for analyzing 3-d video games. The Video Game Theory Reader 67–86.

McRae, K., Hughes, B., Chopra, S., Gabrieli, J.D.E., Gross, J.J., Ochsner, K.N., 2010. The neural bases of distraction and reappraisal. J. Cogn. Neurosci. 22, 248–262.

Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233–236.

Michailidis, B.-B.E., He, L.X., 2018. Flow and immersion in video games: the aftermath of a conceptual challenge. Front. Psychol. 9, e1682.

Müller, H.J., Rabbitt, P.M., 1989. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. J. Exp. Psychol. Hum. Percept. Perform 15, 315–330.

Murphy, S., Spence, C., Dalton, P., 2016. Auditory attentional selection is biased by reward cues. Sci. Rep. 6, e36989.

Murphy, S., Spence, C., Dalton, P., 2017. Auditory perceptual load: a review. Hearing Research 352, 40–48.

Murray, J.H., 2017. Hamlet on the Holodeck: The Future of Narrative in Cyberspace. MIT Press.

Näätanen, R., 1992. Attention and Brain Function. Routledge, London.

R. Nicol, O. Dufor, L. Gros, P. Rueff,and N.Farrugia.EEG measurement of binaural sound immersion.In: EAA Spatial Audio Signal Processing Symposium, 73-78, 2019.

Nilsson, N.C., Nordahl, R., Serafin, S., 2016. Immersion revisited: a review of existing definitions of immersion and their relation to different theories of presence. Human Technology 12 (2), 108–134.

Olivers, C.N.L, Meeter, M., 2008. A boost and bounce theory of temporal attention. Psychol. Rev. 115, 836.

Olszanowski, M., Frankowska, N., Tolopilo, A., 2023. "rear bias" in spatial auditory perception: Attentional and affective vigilance to sounds occurring outside the visual field. Psychophysiology 60, e14377.

Panagiotakopoulos, D., Marentakis, G., Metzitakos, R., Deliyannis, I., Dedes, F., 2022. Digital scent technology: toward the internet of senses and the metaverse. IEEE: IT Professional 24, 52–59.

Parasuraman, R., Wilson, G.F., 2008. Putting the brain to work: neuroergonomics past, present, and future. Hum. Factors 50 (3), 468–474.

Park, D.S., Lee, D.G., Lee, K., Lee, G., 2017. Effects of virtual reality training using xbox kinect on motor function in stroke survivors: a preliminary study. J. Stroke Cerebrovasc. Dis. 26, 2313–2319.

Petersen, G.P., Petkakis, G., Makransky, G., 2022. A study of how immersion and interactivity drive VR learning. Comput. Educ. 179, e104429.

Pfurtscheller, G., LopesdaSilva, F.H., 1999. Handbook of Electroencephalography and Clinical Neurophysiology: Event–related Synchronization. Elsevier, Amsterdam.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148.

Power, A.J., Foxe, J.J., Forde, E.J., Reilly, R.B., Lalor, E.C., 2012. At what time is the cocktail party? a late locus of selective attention to natural speech. Eur. J. Neurosci. 35, 1497–1503.

Ray, W.J., Cole, H.W., 1985. Eeg alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. Science 228, 750–752.

Reggente, N., Essoe, J.K., Aghajan, Z.M., Tavakoli, A.V., McGuire, J.F., Suthana, N.A., Rissman, J., 2018. Enhancing the ecological validity of fMRI memory research using virtual reality. Front Neurosci. 12, 408.

Ren, Y., Zhang, Y., Hou, Y., Li, J., Bi, J., Yang, W., 2021. Exogenous bimodal cues attenuate age-related audiovisual integration. Iperception 12, 20416695211020768.

Richter, M., 2013. A closer look into multi–layer structure of motivational intensity theory. Soc. Pers. Psychol. Compass 7, 1–12.

Ryan, M.L., 2003. Narrative as virtual reality: immersion and interactivity in literature and electronic media. The Johns Hopkins University Press, Baltimore.

Ryan, R.M., Deci, E.L., 2000. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp. Educ. Psychol. 25, 54–67.

Sanderson, D.J., McHugh, S.B., Good, M.A., Sprengel, R., Seeburg, P.H., Rawlins, J.N., Bannerman, D.M., 2010. Spatial working memory deficits in GluA1 AMPA receptor subunit knockout mice reflect impaired short-term habituation: evidence for wagner's dual-process memory model. Neuropsychologia 48, 2303–2315.

Sani, I., Stemmann, H., Caron, B., Bullock, D., Stemmler, T., Fahle, M., Pestilli, F., Freiwald, W.A., 2020. The human endogenous attentional control network includes a ventro-temporal cortical node. Nat Commun 12, 360.

Sarter, M., Gehring, W.J., Kozak, R., 2006. More attention must be paid: the neurobiology of attentional effort. Brain Res. Rev. 51, 145–160.

Schäfer, P.J., Corona-Strauss, F.I., Hannemann, R., Hillyard, S.A., Strauss, D.J., 2018. Testing the limits of the stimulus reconstruction approach: Auditory attention decoding in a four-speaker free field environment. Trends Hear 22, 1–12.

Schellenberg, S., 2013. Belief and desire in imagination and immersion. J. Philos. 110, 497–517.

E.N. Schneider, C. Bernarding, A.L. Francis, B.W.Y. Hornsby, and D.J. Strauss.A quantitative model of listening related fatigue.In: Neural Engineering (NER), 9th International IEEE/EMBS Conference on, 619-622, 2019.

Shams, L., Kamitani, Y., Shimojo, S., 2000. What you see is what you hear. Nature 408, 788.

Shinn-Cunningham, B.G., 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12 (5), 182–186.

Simons, D.J., 2000. Attentional capture and inattentional blindness. Trends in Cogn. Sci. 4, 147–155.

Slater, M., Usoh, M., Steed, A., 1994. Depth of presence in virtual environments. Presence: Teleoperators Virtual Environ. 3, 130–144.

Snyder, L.H., Batista, A.P., Andersen, R.A., 2000. Intention-related activity in the posterior parietal cortex: a review. Vision Res. 40, 1433–1441.

Song, J.H., 2019. The role of attention in motor control and learning. Curr. Opin. Psychol. 261–265.

Souza, R.H.C., Naves, E.L.M., 2021. Attention detection in virtual environments using EEG signals: a scoping review. Front Physiol 12, 727840.

Spence, C., Santangelo, V., 2009. Capturing spatial attention with multisensory cues: a review. Hear Res. 258, 134–142.

Spence, C.J., Driver, J., 1994. Covert spatial orienting in audition: exogenous and endogenous mechanisms. J. Exp. Psychol. Hum. Percept Perform 20, 555–574.

Stein, B.A., Meredith, M.A., 1993. The merging of the senses. MIT Press, Cambridge.

Strauss, D.J., Francis, A.J., 2017. Toward taxonomic model of attention in effortful listening. Cogn. Affect Behav. Neurosci. 17, 809–825.

Strauss, D.J., Corona-Strauss, F.I., Schroeer, A., Flotho, P., Hannemann, R., Hackley, S.A., 2020. Vestigial auriculomotor activity indicates the direction of auditory attention in humans. eLife 9, e54536.

Swann, C., Keegan, R.J., Piggott, D., Crust, L., 2012. A systematic review of the experience, occurrence, and controllability of flow states in elite sport. Psychol.Sport Exerc. 13, 807–819.

Tamborini, R., Skalski, P., 2006. The role of presence in the experience of electronic games. Playing Video Games: Motives, Responses Consequences 225–240.

Thesen, T., Vibell, J., Calvert, G.A., Osterbauer, R., 2004. Neuroimaging of multisensory processing in vision, audition, touch and olfaction. Cogn. Processing 5, 84–93.

Thon, J.-N., 2008. Immersion revisited: On the value of a contested concept. Extending Experiences. Structure, Analysis and Design of Computer Game Player Experience. Lapland University Press.

Trenado, C., Haab, L., Strauss, D.J., 2009. Corticothalamic feedback dynamics for neural correlates of auditory selective attention. IEEE Trans. Neural. Syst. Rehabil. Eng. 17, 46–52.

Trinh Van, L., Dao Thi Le, T., Le Xuan, T., Castelli, E., 2022. Emotional speech recognition using deep neural networks. Sensors 22, 1414.

Unsworth, N., Robison, M.K., 2018. Tracking arousal state and mind wandering with pupillometry. Cogn Affect Behav. Neurosci. 18, 638–664.

Valenza, G., Scilingo, E.P., 2014. Autonomic Nervous System Dynamics for Mood and Emotional-State Recognition. Springer, Heidelberg.

Van Laer, T., De Ruyter, K., Visconti, L.M., Wetzels, M., 2013. The extended transportation-imagery model: a meta-analysis of the antecedents and consequences of consumers' narrative transportation. J. Consumer Res. 40 (5), 797–817.

Van Wassenhove, V., Grant, K.W., Poeppel, D., 2007. Temporal window of integration in auditory-visual speech perception. Neuropsychologia 45, 598–607.

Visch, V.T., Tan, E.S., Molenaar, D., 2010. The emotional and cognitive effect of immersion in film viewing. Cogn. Emotion 24, 1439–1445.

Vortmann, L.M., Kroll, F., Putze, F., 2019. EEG-based classification of internally- and externally-directed attention in an augmented reality paradigm. Front. Hum. Neurosci. 13, 348.

Wolpert, D.M., Doya, K., Kawato, M., 2003. A unifying computational framework for motor control and social interaction. Philos. Trans. R. Soc. Lond. Series B: Biol. Sci. 358, 593–602.

Wong, C.W., Olafsson, V., Plank, M., Snider, J., Halgren, E., Poizner, H., Liu, T.T., 2014. Resting-state fMRI activity predicts unsupervised learning and memory in an immersive virtual reality environment. PLoS One 9, e109622.

Wong, D.E., Fuglsang, S.A., Hjortkjær, J., Ceolini, E., Slaney, M., de Cheveigné, A., 2018. A comparison of regularization methods in forward and backward models for auditory attention decoding. Front. Neurosci. 12 https://doi.org/10.3389/fnins.2018.00531.

Yu, K., Prasad, I., Mir, H., Thakor, N., Al-Nashash, H., 2015. Cognitive workload modulation through degraded visual stimuli: a single-trial eeg study. J. Neural. Eng. 12, e046020.

Yu, M., Shasha, X., Hua, M., Wang, H., Chen, X., Tian, F., Li, Y., 2022. EEG-based emotion recognition in an immersive virtual reality environment: From local activity to brain network features. Biomed. Signal Process Control 72, 1746–8094.

Zimmerman, M., 1989. Human Physiology (2nd. Complete Ed.). Chapter the Nervous System in the Context Of Information Theory. Springer.