Augmented Narrative: Assisting the Reader with Sound

Susana Sanchez Perez, Naohito Okude and Kai Kunze

1 Introduction

Designing an assistive augmentation for narrative texts requires a solution-based creative process that begins with a deep understanding of the topic at hand. Therefore, it is essential to study literature and how narrative texts are perceived to answer questions as to why readers need assistance and how to implement the assistive augmentation.

The study and analysis of literature is a scholarship that corresponds to literary theory, with its different schools that discuss and write about writing. One of these schools is the reader-response criticism that views literature as a performing art where readers create their own text-related performance. This division of literary theory began in the 1960s, and considers the meaning of a narrative text is completed through interpretation [3]. Amongst the different views of reader-response theorists, Individualists focus only on the reader's experience, while Uniformists assume that text and reader have a shared responsibility to convey meaning. This makes reading both, subjective and objective. One must look into reading processes to create meaning, and experience to understand the narrative text. Here, there are two levels of understanding: the information explicitly presented in the narrative text, and the integration of the different pieces of information from that narrative text.

The reader-response criticism is based on Kintsch's construction-integration model [37]. This general theory contends that comprehension arises from the interaction between the narrative texts to comprehend, and the reader's general knowledge and lived experiences. In fact, studies based on the construction-integration model stress the importance of background knowledge and the reader's ability to generate

S. Sanchez Perez (✉) · N. Okude (✉) · K. Kunze (✉)
KMD, Keio University, Tokyo, Japan
e-mail: su_saps@hotmail.com

N. Okude
e-mail: naohito.okude@gmail.com

K. Kunze
e-mail: kai.kunze@gmail.com

inferences to fill in gaps of missing information, supporting the idea of a multidimensional scale of text complexity [12, 23].

The construction-integration model has many similarities with the possible world (PW) theory, adapted to narratives texts by Lewis [42]. Lewis postulates the idea of "modal realism" to make a distinction between real worlds and the actual world. Here, all PW are real, as they exist even if it is only in the imagination, but there is only one actual world (AW). This AW serves as a model to mentally construct other storyworlds that can differ from the AW. Readers imagine fictional worlds as close as possible to the AW, changing only what the narrative text mandates, in an interpretative rule called "the principle of minimal departure". For example, if the narrative text describes Pegasus as a winged stallion, the reader will image a creature that resembles in every respect a real world horse, real world meaning the AW, except for the fact that this horse has wings [61].

Contrary to the reader-response criticism are the text-oriented schools. Formalism claims readers can understand narrative texts while remaining ambivalent about their own culture. However, if said readers belong to a culture where horses do not exist, because of their geographical location, would they be able to understand Pegasus? In this case, the lack of previous knowledge and experience could affect the perception of the narrative text, frustrating readers that do not have enough information to mentally construct the mythological creature. Another example is the Victorian novel, which represents a culture and social norms of old fashion didacticism [39]. Set in a historical present, period literature turns out to be problematic to today's readers, since this literary genre confers them with a simple lesson from the past that has become meaningless with time [67]. The perception of the Victorian novels is then affected by the culture of contemporary readers, who would need assistance to fill in the gaps of information about context and social norms of Victorian societies in order to extract meaning.

That being the case, "Augmented Narrative" is proposed as an assistive augmentation. In line with the reader-response criticism this augmented narrative also uses the construction-integration model as framework. The augmentation allows the narrative text to provide an embodied experience to the reader, supplementing information for the lack of previous lived experiences, and thus assisting the multidimensional scale of the narrative text complexity.

2   Embodiment

Augmented Narrative looks at embodiment to assist perception and retrieve meaning from the narrative text, using the qualities of sound to add non-verbal information to the narration. The sonic augmentation allows an embodied experience, where the reader perceives the narrative through mind and body: the written word and sound.

The phenomenologist Alfred Schtz calls literary meaning "monothetic", since it relies on idealizable and objectifiable semantic content that makes the written word time transcendent, sign-oriented, and conceptual [63]. But to truly conceive

reading as a performing art, we have to look for a "polythetic" meaning. To illustrate, theatrical performances are a form of communication, and more appropriately of storytelling, which do not rely on semantic content. In theatre, meaning is time-immanent, fully perceived in an embodied experience of the moment [17]. In this way, plays transmit a 'polythetic' meaning, assisting audiences to experience within the embodied consciousness, and not as an object of consciousness.

Narratological studies agree that orality and sound, which are in a close connection with theatre and the natural world, have nothing in common with literature. This field of literature considers the experience of the world is detached from the omniscience of the written word, as literature offers readers the artificial effect of experiencing and viewing outside the natural world [16]. By doing so, narratology has failed as a comprehensive discipline as regardless of literature's disconnection with its oral past, it still shares some characteristics that allow for an embodied experience.

Embodiment, in a simple definition, is the biological and physical presence of the body [43]. However, Maurice Merleau-Ponty, phenomenological philosopher, defines embodiment by separating the objective body, the physiological entity, from the phenomenal body, the sense of ones motor capacities with which to experience and perceive the world [49]. Merleau-Ponty's definition of embodiment has served to connect phenomenology to cognitive sciences and neuroscience, necessary fields of study to understand how readers retrieve meaning from written language. The connection between mind and body also concerns literary theory, as the brain is able to simulate actions through language processing. Read words automatically elicit neural activations similar to those that occur during the perception of events, because verbs evoke mental representation of objects to which the action refers [45].

In parallel with the construction-integration model, neuroscientists have found evidence that the brain has two dimensions: mirroring and self-projection. Mirroring allows to physically resonate what others are experiencing, whereas self-projection implies imagining what should be felt, and then attributing the imagined experiences to others. Imaginative capacity, which involves previous knowledge for self-projection, is required to attenuate the cognitive distance gap, as mirroring is only felt in low intensity [36]. Mirroring allows an intuitive and immediate comprehension of actions, and self-projection uses an inferential process of self experience to reason intentions into emotions [9]. In Augmented Narrative sound is used to assist imagery to further attenuate the cognitive gap between mirroring and self-projection, augmenting the experience of the phenomenal body to physically resonate the storyworld.

The embodiment postulation is a remediation of print to convey a polythetic meaning through a co-evolutionary process in a double dimension of the interface: human-media and media-media [64]. The proposed embodied reading experience follows Andy Clark's "extended mind" theory, where cognition is a cycle that runs from brain through body, world, and back. In fact, Clark sees the boundary of "skin and skull" as cognitively meaningless. The human-media interface is a *biofeedback*, a text-body interaction where the narrative text knows when to assist the reader by measuring attention and memory work. On the other hand, the remediation of the

media-media interface follows a coevolution where different media has contaminated one another to create the *sonic assistance*[48]. Theatrical sound design is combined with the latest available technology to reinterpret literature as an embodied experience. Here, sound operates in a cognitive ecology of multidimensional content, where the reader uses mind and body to read, listen, feel, remember, and imagine.

3 Cognitive Ecology of Literature

The changes made to transform literature into a medium that assists its reader affects not only the object, in this case the book, but also the social structure around it and the cognitive architecture of written communications. Therefore, it is important to consider how the object's assistive augmentation will affect the processes of how literature is created and consumed, as no technology can serve as a reading space in the absence of writers and readers. The framework of cognitive ecology is proposed because it supports extending the medium's capability to be an assistive storyteller within the simplest communication model of sender-receiver. Furthermore, the framework is used to translate literary theory into cognitive and computational terms bringing literature into the field of HCI.

Edwin Hutchins defines cognitive ecology as the study of context, in particular the mutual dependence between elements in an ecosystem [29]. Here, the mind arises within a physical system distributed over space and time. Hutchins suggests sensory and motor processes are not peripheral, making the relations of brain-and-body interactions with the environment an important unit of analysis. This ecological approach uses the distributed cognition theory to describe, in computational terms, human-work-systems where knowledge lies not only within the individual, but the individuals social and physical environment [27]. The goal of the theory is to describe how distributed units are coordinated by analyzing the interactions between individuals, the media used, and the environment within which the activity takes place.

The focus of a cognitive ecology of literature is in the sender-receiver model, by augmenting the medium's capability to better transmit the message in absence of the writer, who is not in the same physical space as the reader. When the reader has not enough background knowledge to model the narrative world, the result is a lack of understanding that hinders the ability of the writer to transmit her message; if there is not enough shared understanding between the two, communication cannot take place. This brings antagonism to the reader, drifting attention away from the narrative text, making it difficult to continue reading.

For the writer, her readers are unknown, leaving it up to these readers to retrieve meaning through decoding and interpretation. The medium of print allows the narrative to talk to the reader, but not to listen. The reader can interpret, but will not be reassured or assisted. However, if the medium changes into an empathetic narrator capable to understand the reading experience, then literature becomes an intrinsic co-constitution in a dialog between three agents: author, reader, and book. In Hutchins' terms, the unit of writer affects the reader through the narrative text, whereas the

unit of reader affects the writer on how her work is interpreted. Here, there is a shared responsibility to achieve meaning, even in print, but with different weights. The difference is that in print both units, writer and reader, have a delimiting line that suppresses, not only the natural dialog found in oral communications, but also relegates auditory and any other sensuous complexity. For that reason, it is important to look into Plato's advices to see the boundaries between units more as joints, where connectivity is no longer relatively long, but strong and reciprocal. The three agents: writer, reader, and book, can then interact within a cognitive ecology of literature where perception and attention are technologically mediated, in a system distributed across social structures, objects, and cognitive architectures. The assistive remediation comes when literature is augmented to become aware of the human cognitive architecture: sensory memory, working memory and long-term memory through a biofeedback.

The cognitive ecology principle has been used to create a framework for Shakespearean Studies [69]. It served to analyze theatre across a system of neural and psychological mechanisms, bodily and gestural norms, physical environments, cognitive artefacts, and technologies of sound and light; elements that affect and modulate each other. In this theatrical cognitive ecology sound design is used to mediate the audience's attention and perception during the play. Here, sonic information is detected as perceivable opportunities for action in the environment, where sound psychoacoustics allow audiences to distinguish between a variety of sounds, and how to interact with them [55].

In this cognitive ecology sound is used to lessen the communication problem between writer and reader increasing the available information about the storyworld. The ecological approach is based on perception of information rather than sensation [18]. Even though ecological psychology seems to contradict information processing (recognition is a multi-stage process between perceptual qualities of sound source, abstract representation in memory, meanings, and associations), it explains why sound can carry meaning even if that same sound has been created artificially, as it is concerned with the invariant properties of the sound [19, 66]. For instance, one is able to recognize someone's voice on the phone, even if that person has a terrible cold.

4 Biofeedback

Based on the relation between mind and body, computers are capable to understand emotional states. Consequently, computers can take actions by recognizing likes and dislikes of users, evolving from mere tools to personal companions. Sensors give logic and reason to mechanical devices enabling them to empathize with their users, looking for positive and negative reactions to a task. These sensors have been used to measure the user's emotional stand in digital entertainment to allow interaction in narratives. For instance, Interactive Storytelling captures user's emotions to create an affect-based interactive narrative. This affective gaming creates new experiences

by adapting the game to the player's emotions in two modalities: managing the story-line to achieve the game's emotional goal or adapting the narrative only to generate positive emotions [74].

Rather than adapting the narrative text to balance emotions, an augmented narrative focuses on keeping the reader engaged without the need of multiple storylines or changing the author's work. Here, sensors are used to measure the reader's cognitive responses, reflected in the body, to what is being read. As the proposed augmentation is concerned with assisting cognitive processes of perception and attention rather than emotional responses, the biofeedback is based on mental workload. Mental workload increases with memory work, used in reading comprehension, where previous knowledge is required, and correlates to engagement tasks of sustained attention [4]. Thus, the biofeedback looks for levels of engagement to know when to assist. Based on the reader's mental workload, the augmented narrative can detect if the reader has been or not engaged in the storyworld.

A higher mental workload signals the reader is being engaged in the story. On the other hand, a low mental workload signals a lack of engagement, when the reader is finding difficulties to create her own mental performance of the narrative text, and is in need of assistance. Consequently, the proposed biofeedback operates as a bridge for understanding between author, medium, and reader, re-designing the medium to be an assistive narrator that acts and delivers the story according to the readers needs. In this way, authors can be assured that their work will be meaningful to their readers.

Perception and emotion are intricately related, and the line between the two is blurred at the best of times. For instance, even though engagement is related to attention and memory work, is a positive feeling. While there are different physiological metrics to create a biofeedback, some might be more adequate to find emotional states and others to find cognitive processes. To design a biofeedback for an augmented narrative, the physiological metrics of mental workload in the heart rate variability (HRV) and nose temperature are examined. The first of these metrics served to design a circumflex model of engagement, while the second was used to develop smart-glasses to detect reading engagement.

4.1   Metrics of Engagement

The study of the empathetic relations readers build with fictional characters has focused in verbal feedback. For example, measuring engagement with expressions such as: "*I felt I was there right with Phineas*" about the book "A Separate Peace" by John Knowles [58]. The verbal feedback is possible since books, with their authorial omniscience place non-natural characters into a natural frame, allowing authors to engage readers with the artificial effect of experiencing and viewing outside the natural world. In turn, readers attribute a mental stance to these non-natural characters, in the same way they do in everyday life, building mental reconstructions of read emotions and actions [2, 21].

Empathy is the mind-reading ability that allocates mental states to fictional characters and is essential for engagement [26]. The segment of the story affects the degree of engagement, which is more prominent during the climax of the story, where the reader is expected to be paying more attention [1]. This deeper sense of involvement is part of Csikszentmihalyi's flow theory [11]. Here, positive experiences come when engaged in task demands where the person is in deep sense of control, and the activity feels rewarding. Flow can also be considered to involve straining tension and mental workload. However, the stress found in flow is a positive experience referred to as eustress. Based in this theories an in order to understand the demanding character of flow activities, and find engagement in reading tasks, some studies in the metrics of the heart are reviewed.

4.1.1 Heart Rate Variability

Involuntary responses controlled by parasympathetic components slow the heart rate (HR) and sympathetic components raise it. In the absence of arousal, attraction and aversion can be detected in the variations between heartbeats in the heart rate variability (HRV), called valence [35]. Engagement can be associated with a decrease in HR, contrary to the increment found in emotional responses. Thus, information of engagement can be determined by a positive valence, detected in the ratio of low frequency energy to high frequency energy that represents the extent of sympathetic and parasympathetic influence in the HRV. On the other hand the highest frequency indicates boredom, in negative valence, reflecting the lowest mental workload.

Mental workload is affected by engagement, as information-processing increments mental workload. Keller et al. looked into the impact of flow in conditions of boredom, fit, and overload while completing knowledge tasks with questions from the TV show "Who wants to be a millionaire?". Results revealed the highest HRV was found during boredom, reflecting the lowest mental load, while a decreased HRV reflected involvement in the fit condition, showing that higher levels of flow can be associated with a low HRV. In the absence of arousal, the valence of different conditions of involvement such as engagement, boredom, and overflow, can be detected in the variations between heartbeats. In the HRV, the variations between heartbeats, time interval is referred as the QRS complex in which the ratio of low frequency energy to high frequency energy represents the extent of sympathetic and parasympathetic influence on the heart.

To find the metrics of engagement, Keller's frequency data was compared to McCraty's et al. typology of six HRV patterns that denote different modes of psychophysiological interaction [47]. The graphic from McCraty et al. describes wave forms of emotional states, divided into normal every day life emotional experiences and hyper-states of emotional experiences. These were compared to Keller et al. HRV data using the axis of arousal and balance. In the findings 'fit' corresponds to serenity, while 'boredom' corresponds to apathy. Even though both scholars look for different processes, cognitive and emotional, the wave patterns are similar, linking emotional and cognitive metrics of the heart.

Fig. 1 Circumflex model of engagement based on the metrics of HRV
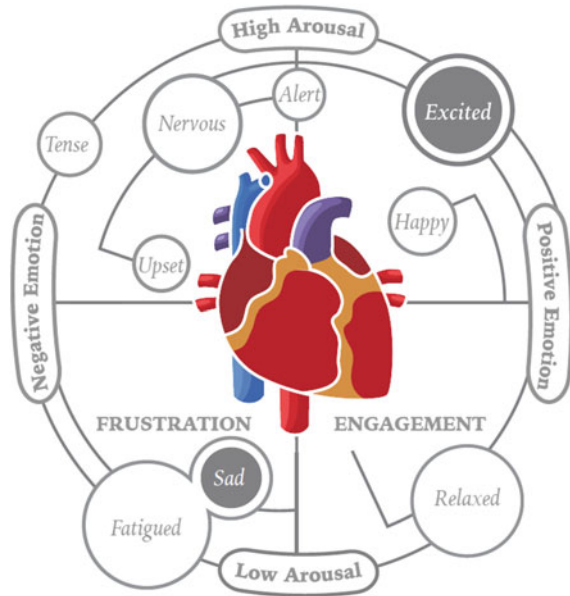
Based on this findings, a circumflex model of engagement is suggested (See Fig. 1). In this model, starting clockwise, quadrant four and quadrant one represent arousal, an emotional stimulations by the narrative text. In quadrant two the reader is in a state of relaxation or eustress, meaning engagement. Finally, in quadrant three, the system has an input of frustration or stress, which in this case will be defined as non-engagement. Quadrant two and three are low arousal, meaning there is an absence of demanding tasks in emotional processing, where the mental workload takes over, and is regarded as a cognitive process involving attention.

4.1.2 Nose Temperature

The metrics of facial temperature have been found to reliably discriminate between positive and negative emotions, and cognitive tasks. Here, the sympathetic system is responsible of lowering the temperature of the nose in mental workload tasks. On the other hand, frustration, which is related to boredom, increases the blood volume into supraorbital vessels along with the skin temperature of the nose [25, 31].

Israel Waynbaum, in his vascular theory of emotional expression, attributes the experience of emotions to follow facial expressions rather than preceding them; relating it to the James-Lange theory [7, 73]. The vascular theory is based on the fact that the supply of blood to the brain and face comes from the same source, the carotid artery. Therefore, the reactions to circulatory perturbations in the facial artery produce disequilibrium in the cerebral blood flow. The facial muscles contract and push against the skull bone structure, acting as a tourniquet on arteries and veins.

This serves to regulate the blood flow, affecting the cerebral blood flow, reducing or complementing it. Blood flow alterations cause temperature changes that modify the neurochemistry of the brain. The thermoregulatory action influences peptides and neurotransmitters, temperature dependent, found to produce emotional changes. Cooling is associated to pleasant and warming to unpleasant feelings [8, 13, 30]. For example, facial skin temperature of nose, forehead, and cheeks, decreases when laughing and increases when being angry. The decrease in nose temperature can be the most dramatic, dropping as much as 2.0 °C in 2 min [52]. Moreover, the temperature changes that affect the region of the nose, have been considered reliable to detect cognitive tasks. For example, nose temperature has been associated to mental workload, where in a driving test study simulator drives led to a higher subjective workload score and a greater nose temperature drop than real driving [53].

Mental workload increases with memory work, used in reading comprehension, and correlates to engagement tasks of sustained attention [4]. To create a biofeedback related to engagement in reading tasks, preliminary work showed mental workload can be used as an indication of engagement versus non-engagement, in a negative correlation between nose temperature and subjective immersion [39]. In reading tasks the temperature of the nose decreases when the reader is engaged, while increasing when not being engaged. In our preliminary study a correlation in engagement between the metrics of the heart and those of the temperature of the nose was found. When engaged, the skin temperature of the nose decreased (in red), and the temperature when non-engaged increased (in blue) (See Fig. 2). The metrics of heart showed that the HRV had higher and lower values in the non-engaging exercise than in the engaging exercise in accordance to the "Graph of psychophysiological interaction distinguished by the typology" from the Institute of HeartMath.

In synthesis, the metrics of nose temperature could be related to an engaged experienced of sustained attention related to the flow theory in reading tasks, as mental workload increases with memory work. It has been suggested that working memory is an indication of efficiency to sustain attention in multiple task-relevant representations, even when there is distracting irrelevant information [15]. For instance, Kintsch and Ericsson argue that working memory serves to hold a few concepts used
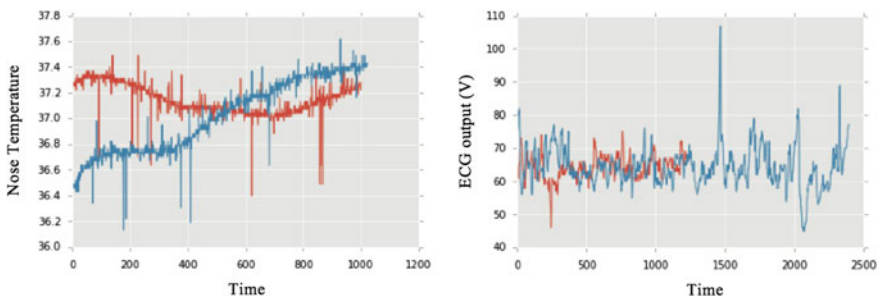

Fig. 2 Correlation of engagement (*red*) versus non-engagement (*blue*) between nose temperature (*left*) and HRV (*left*)
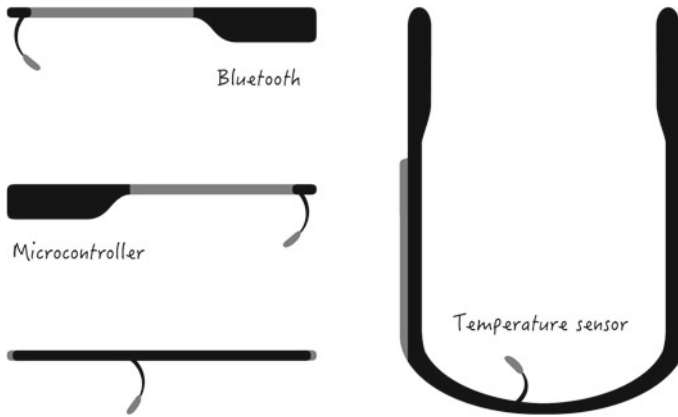
Fig. 3 Sketch of suggested smart-glasses to detect engagement in reading tasks

as cues to link what is stored in long-term memory [38]. Readers compose their own episodic structure for comprehension using the general knowledge to decode the written words and personal experience to give meaning [14]. Thus, for the biofeedback required to create an augmented narrative, the design of smart glasses that follow the reader's engagement with the metrics of memory workload reflected on the changes in nose temperature is proposed (See Fig. 3).

5   Sonic Assistance

The use of sonic assistance in literature is the result of studying the evolution of storytelling, from orality to literacy, with an emphasis on theatre [71]. When analyzing the medium's past, sound was found to remain as a constant component of storytelling. Thus, the proposed sonic augmentation looks into reintroducing components of orality to literature in a dual-coding across senses, where sound effects assist readers by addressing a channel other than the visual. Sound is used to fill in information gaps, with non-verbal information, allowing for an embodied experience. These sound effects act as positive disruptions awakening the mind to increase engagement.

Eric Havelock suggests the middle point between orality and literacy is in Attic theatre, which in itself arose from the need of Athenians in the sixth century to rediscover their own identity as a single city-state [24]. Attic theatre was a supplement for Homer, whose epics furnished the Greek identity, moral, politics and history. But Homer's tales were Ionic and Pan- Hellenic, and not part of the native dialect with which to address new Athenians. More importantly, Attic theatre gave birth to authors that were more producers than writers. They composed their vision, in the tension of oral and written communication by dictating to a literate assistant, and

hearing it back from this assistant to verbally edit, retaining only the ear to compose. Havelock suggests the secret for the brilliance of Attic theatre is due to the tension caused by the transition from orality to literacy, which has not been repeated in history.

This tension is still visible in what Walter Ong calls residual oral cultures. For example, the acclaimed playwright William Shakespeare was part of these residual oral cultures. As such, it is unlikely that Shakespeare was involved in publishing his own plays, since his writing was meant to be spoken, and not read. In fact, John Marston, a contemporary playwright claimed: "*scenes invented to be spoken should not be enforced to the public for reading*" [57]. This is because Shakespeare's plays entail an interaction that is directly addressed to a particular audience, in a particular moment. However, generations of editors have added layers of silent emendations, holding meaning through grammatical punctuation, with commas and periods that set off clauses for the eye. These editorial revisions make Shakespeare more readable, since print is sophisticated and precise, but take away the rhetorical and auditory punctuation, disconnecting our embodied experience of theatrical narration [44].

With the evolution of storytelling, from orality to literacy, the perception of the narrative became purely visual, where the reader depicts the immediate external stimuli to the organism as neural representations, and relates it to internal portrayals encoded in memory. However, perception is not limited to only one sense. Events in everyday life are registered by more than one modality, integrating the different information from various sensory systems in a unified perception. For example, in Elizabethan plays not all sights of the action could be portrayed on stage, but even when these actions could not be seen, they all could be heard, making sound an important part of the narration. These off stage sounds were so important that packs of hounds and soldiers with their full arsenal were hired to produce authentic noises on cue, allowing for a multimodal perception.

Specific cues, such as sound effects or music fall into different dimensions, all of which need to be informational in nature to support off-the-moment events [40]. Even when the source is not in the field of vision, sound effects are powerful tools for stimulation. Audiences are able to recognize sounds by linking them to past sonic experiences, allowing them to mentally identify the source of each sound, and the significance it has to the narration. Even when sensory information is insufficient for the listener, the perceptual system still analyzes the situation taking into consideration previous knowledge acquired from the surrounding sonic-world [46].

In ecological psychology the physical nature of the sounding object, the means by which it has been set into vibration and the function it serves to the listener, is described to be perceived directly without any intermediate processing [50]. For example, studies show that listeners formulate the same cognitive organization based on the mechanics of the sound source: machine and electric device sounds, liquid sounds and aerodynamic sounds, even when these are sound-generated human-made illusions [28, 41]. Overall, sound has the ability to carry on non-verbal information. Sonic cues are able to structure perception by associating themselves to images and meanings, contributing for a significant experience, placing sound at the heart of interpretation [70].

5.1 Dual Coding Across Senses

Allan Paivio studied the unique human skill to deal simultaneously with verbal and non-verbal information [54]. In Pavio's dual-coding theory, the general assumption is that there are two classes of phenomena handled by separate cognitive subsystems: one for representation and processing of information, and the other for dealing with language. Language is a peculiar system that deals directly with speech and writing, while at the same time serves with a symbolic function with respect to non-verbal objects, events, and behaviours. For example, literature makes use of non-verbal information assistance adding illustrations to improve comprehension. In fact, students can perform better in reading comprehension test when there is a mix of text and images [32].

On the other hand radio drama, also called theatre of the mind, uses a dual-coding that addresses not the eyes, but the ears. Without any visual component, radio drama holds its meaning in the auditory dimension, depending on dialog, music, and sound effects to deliver enough information for the imagination. Here, sound effects round out the dialog, filling the absence of visual cues to convey meaning, and allow listeners to become part of the intricate moments of the drama. Sound is so powerful for the imagination that the auditory experience consents the absence of reason and logic. Listeners can lose structured thoughts to invite ideas that cannot be explained [72]. For example, in 1938 the radio play "War of the Worlds" caused hysteria amongst listeners who believed the Martians were invading. Imagery, also referred as listening with the mind, provided the radio audience in 1938 an aural visualization of extraterrestrial chaos, making them forget they were only tuning in to hear their regular program.

Regardless of the advantages of sonic non-verbal cues in radio drama, audiobooks have only retained the printed word. Even so, components of oral storytelling are still present. For instance, listening to an audiobook can make narrative texts that seem tedious to read, reveal the fullness of the literary work; especially if the work is narrated by a good actor [65]. Moreover, listening to what is being read can assist comprehension. One example is the modernist writer James Joyce, who experimented with aural reading to achieve comprehension. In Joyce's "Finnegans Wake" there is an acknowledgement of the importance of the sensory system, where understanding emerges from sound. The language developed by Joyce forces his readers to become aware that "Finnegans Wake" has to be pronounced, preferably out loud, to be able to understand the story [60]. To retrieve meaning readers needs to get into a rhythm that only can be achieved through orality.

In aural reading there is a redundancy of information, presented in a combination of auditory and visual channels, where readers process the same information twice. Language presented in these two modalities produces and enhanced memory recall [56]. Applied in multimedia learning, this verbal redundancy can facilitate the narration. When words are presented in both channels, learners are able to extract meaning from both modalities with no cognitive overload, since working memory and auditory working memory are independently processed [51].

In synthesis a dual-coding assists comprehension. In literature this is done through the visual channel with text and images, while in radio drama this is done through the auditory channel with speech and sound effects. Finally a dual-coding seems to be effective in a redundancy of verbal information across both channels: visual and auditory. However, there is an unexplored application using the potential of sound to carry non-verbal information mixed with the verbal information that the narrative text conveys.

In an augmented narrative verbal information, retrieved through the visual channel, is augmented with non-verbal information using sound effects in a dual-coding across senses. The non-verbal sonic information is used to technologically mediate the economy of attention, assisting the reading task. The framework depends on the mind-reading mechanisms, which literature simulates and capitalises on to mentally construct the storyworld [59]. It depends on the ability of the reader to find behavior patterns based on narrow slices of self experience [20]. Thus, an augmented narrative assists by giving more sensory information to the reader.

5.2 Sonic Disruption

Shakespeare and his plays hold the clue on how sound can assists readers: disruption. Shakespeare made a profound impact in the English language, refashioning oral expression through pen and paper. He has been credited for over 1700 original words, by changing nouns into verbs, verbs into adjectives, and by connecting words never before used together. However, Shakespeare's new words have a deeper impact into human conscience than just transforming oral expression. For example, in his play "Coriolanus", the main character returns to Rome for revenge. Menenius, well regarded by Coriolanus, is sent to persuade him to halt his crusade for vengeance. After sending Menenius back with no truce, Coriolanus recognizes: "*This last old man, whom with a crack'd heart I have sent to Rome, loved me above the measure of a father; nay,* **godded** *me, indeed*". In this dialog, Shakespeare stimulates the reader by turning the noun God into a verb, where the new word 'godded' causes a grammatical disruption that triggers a P600 wave in the brain [68].

The P600 wave is a positive voltage variation peak in an electroencephalogram recording; a response found in the posterior part of the scalp that starts about 500 ms after hearing or reading an ungrammatical word. The wave reaches its maximum amplitude in around 600 ms, giving the reason for the name P600 [22]. More importantly, P600 is related to violations in the probability of occurrence of a stimulus, placing it in the P300 family [10]. This makes Shakespeare's new verbs unexpected stimulation rather than just grammatical disruptions.

The design of an augmented narrative involves translating Shakespeare's grammatical disruptions into sonic disruptions that act as assistive augmentations, favouring an embodied experience of literature with a dual-coding of information across

senses, but also awakening the mind calling for the reader's attention. Here, sound is also used as spontaneous stimuli to technologically mediate attention when engagement is low, using the previously described smart glasses. Finally, it is important to mention disruption is not unique to Shakespeare's work. Alberto Moreiras, academic and cultural theorist, refers to literary disruptions as a "reversal of expectations", and argues that if properly managed, can prove more enabling and empowering to the story that was interrupted [34].

6   Augmented Narrative

Augmented Narrative is literature that assists readers to remember, feel, and imagine, triggering sound effects when engagement is low (See Fig. 4). The visual channel is enhanced by adding appropriate auditory information in a dual-coding across senses where information is perceived in an embodied experience. Here, interpretation of read actions is influenced by information available in the auditory channel. The augmented narrative operates in a cognitive ecology, following the body's reaction to the narrative text, technologically mediating attention and perception with multidimensional content.
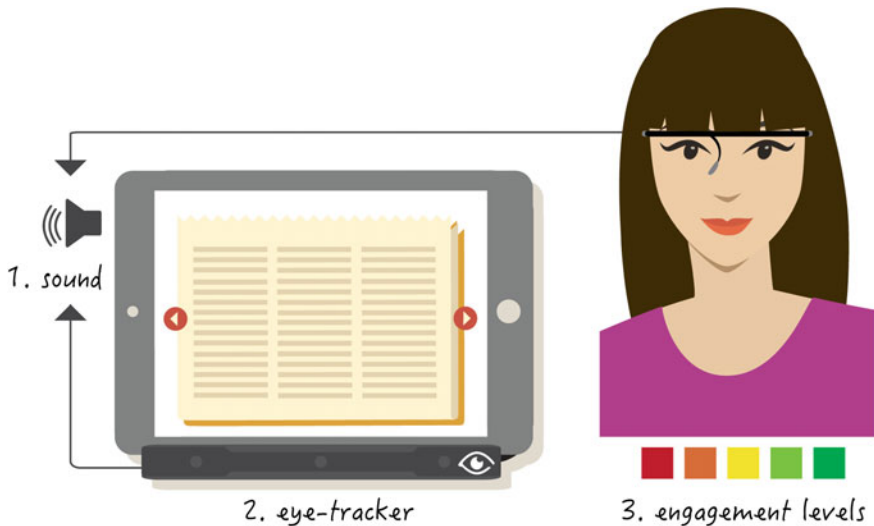

Fig. 4   The concept of augmented narrative has 3 components: (*1*) Sound effects embedded in the narrative text, (*2*) eye tracker for input of reading location, (*3*) smart-glasses to measure engagement. Sonic assistance comes when two conditions are met: the eyes are over the span region, and engagement is low

6.1 Eye-Tracker

Reading skills vary from reader to reader, but eye tracking can serve as input of location, to know the area of the text the reader's gaze is, regardless of reading speed [6]. With the use of eye-tracking technology, the sonic augmentation is designed to work as an eye-driven-stimuli using an off-the-shelf eye-tracker technology to process the eye gaze, and link the sound to the perception of the narrative text. In this way individuals can read at their own pace, and when their gaze is in the 'span region' the sound is triggered. The eye-tracker tobii EyeX is used to design the eye-driven command during the reading task, in a transparent interaction where the reader is not distracted from the task to activate the sonic content (See Fig. 5). The "Gaze Track Plugin" developed by Augusto Esteves was used to connect the eye tracker to the Processing sketch. This plugin allows the tobii EyeX eye-tracker to trigger the embedded sound, marked in a span region. The eye gaze then instructs the sketch to pull the sound from a library of prerecorded *.wav* files. The eye tracker serves only as an input of location, leaving the command of assistance to the smart glasses. Thus, for the sound to be triggered two conditions need to be met: the gaze has to be over the span region and the nose temperature has to increase.

6.2 Engagement Levels

The smart-glasses create a biofeedback, giving the narrative text a cognitive component, allowing for a reciprocal communication where the reader perceives the narrative text, and the narrative text perceives the reader's level of engagement. The eye-driven sonic stimuli is only triggered if engagement is low, a sign that the reader is in need of assistance. Being triggered by involuntary reactions of the body, and not by the conscious mind, these sonic stimulations are set to disrupt prediction in

Fig. 5 Reading tablet with the tobii EyeX eye-tracker and engagement glasses

reading, calling for the readers attention. If the level of engagement is high, then there is no need for assistance, and the augmented narrative can leave the reader to continue the reading experience on her own, avoiding unnecessary disruption.

6.2.1 Measuring Engagement

In a first study we looked for the correlation between the changes in the temperature of the nose and reading engagement. We used our prototype of multimodal literature: (1) smart glasses with skin temperature to measure engagement; (2) Four original short stories from the series "Encounters" (*The Watchmaker, John, The Last Space, That Which Lives in the Attic*) in two modalities: sound and no-sound for control; (3) The tobii EyeX eye tracker mounted on the Dell Venue 11 Pro tablet. Eight volunteers (4 woman, 23–34 years) agreed to participate in testing the augmented narrative. The sample was based on the target readers for digital publishing demographics. All volunteers were given an explanation of the experimental setup and assigned two of the four short stories in a latin square design, one with sound and one with no-sound. Before they could start reading, each volunteer was calibrated for the nose temperature sensor and eye tracker. Volunteers took 2–6 min to read each short story depending on which short stories they were assigned. After each reading they were asked to fill in an immersion questionnaire [33] to measure subjective engagement for both short stories.

A Spearman's rank-order correlation was run to determine the relationship between reader's immersion levels and nose temperature. Combining the values of the two modalities, namely sound and no- sound, a negative correlation was found between immersion and temperature, which was statistically significant ($r_s(16) = -0.517$, $p = 0.04$) (See Fig. 6). Regarding the difference between modalities, we noticed subjects in the sound modality seemed to drop their nose temperature in more occasions than the no-sound modality 4:1, but did not find a statistical significance.

6.3 Sound

Through the analytical reflection that spontaneity is good, an augmented narrative carefully plans to be spontaneous, embedding sound that addresses particular passages that vanish with the passage itself. The narrative text is carefully edited with sound effects to stimulate the reader's mental simulation in an embodied perception of the story. The process begins by identifying sonic cues based on their capacity to effectively transmit non-verbal information relevant to the text, in an approach that fits the artistic process of writing fiction. These cues have been divided into five categories for assistance.

Fig. 6 Visualization of the negative correlation between nose temperature and engagement for the 8 participants in the 2 modalities (16 sets). Temperature decreases with higher engagement
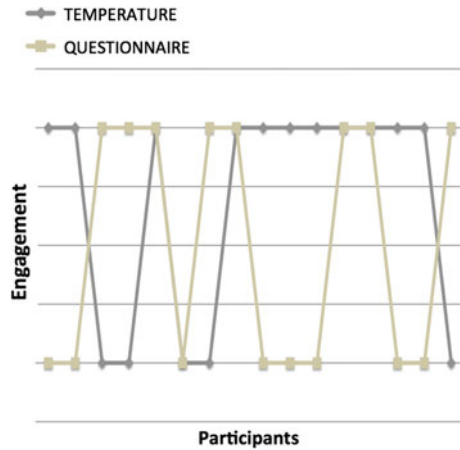
1. Redundancy. There are two sources of information, verbal and non-verbal that relate and complement each other creating an intermodal redundancy. For example when reading: *the phone rang*, the sound can complement the text by giving information about that phone. Is it a mobile, rotary dial or a touch-tone dialling phone?

2. Valid co-occurrences. To bring simultaneous sensory information of an event. For example, when the narrative talks of an explosion there is a synchronization of sound [5].

3. Inference. Writers give cues to their readers to infer situations and emotions, such as a blush or a sweaty hand to tell readers a certain character is in love. These cues can be given with sound, with a palpitating heart, where the reader can hear the intense heart beating of the character.

4. Differentiate. Sonic cues can be use to differentiate between parallel worlds or parallel times using a soundscape.

5. Ambience. The description of the story's setting is critical to understand the fictional world, as it helps to establish time and place. Sounds of an old and squeaky house could be the ambience of a horror story.

6.3.1   Measuring Sonic Assistance

To measure sonic assistance we looked for the best condition to increase reading comprehension and/or imagery between sonic non-verbal cues and sonic verbal redundancy. In this study an extract of "The Legend of Sleepy Hollow" was presented in three modalities: augmented with non-verbal sound effects, augmented with the audiobook taken from LibriVox for verbal redundancy, and without any sound for control. The short story was presented on the Dell Venue 11 Pro tablet computer mounted with the tobii EyeX eye tracker to fifteen volunteers (7 woman, 23–44 years) that agreed to participate in testing the augmented narrative.
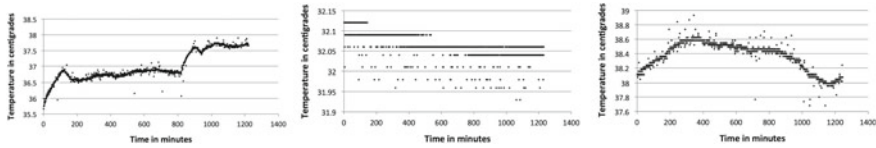
Fig. 7   Nose temperature samples for audiobook (*left*), control (*middle*), and sound effects (*right*)

All volunteers were given an explanation of the experimental setup and assigned one of the three modalities randomly. Before they could start reading, each volunteer was calibrated for the nose temperature sensor and eye tracker. Volunteers took 20 to 30 min to read the extract of the short story. After each reading we asked them to fill in an immersion questionnaire [33] to measure subjective engagement for both short stories and to retell [62] the story in order to be assessed in comprehension and imagery. The retell interview looked for the level of detail provided by each volunteer, comparing it to a previously prepared outline of five categories: characters, event details, climax, setting, and personal connections.

Results form a Kruskal-Wallis H test indicate that there was a statistically significant difference in nose temperature levels between the different modalities, $\chi^2(2) = 6.020$, $p = 0.049$, with a mean rank temperature level of 5.00 for the control group, 7.20 for non-verbal audio cues and 11.80 for participants exposed to audiobooks (See Fig. 7). We found no significant effects for immersion ($\chi^2(2) = 1.067$, $p = 0.587$) or comprehension ($\chi^2(2) = 2.226$, $p = 0.328$). However, were able to visually identify the sound effect modality was better for imagery. Sound appeared to change the temperature of the nose, positively and negatively when compared to the control modality, as this modality had only slight changes in temperature and average immersion. We observed sound effects seemed to enhance imagery, while verbal redundancy helped with comprehension. Even though comprehension was not significant, we noticed participants in the redundancy modality were more confident in the retell interview.

6.3.2   Conclusion

The concept of an augmented narrative follows a historical progression in a co-evolutionary process between literature and technology using familiar instruments and spaces to exploit the brain's natural strengths. Following Gutenberg's approach, where the success of print rests in preserving the quality of scripts, as well as solving production strains, we looked for an interaction that preserves the strengths of literature, but trying to open the reader's cognitive system to new possibilities when consuming it.

Preliminary studies suggest that the nose temperature could be a good indication for engagement levels. The term 'engagement' is proposed as it is part of the flow theory, allowing us to link immersion, more suitable for gaming, with attentional

levels related to memory work found in the nose temperature. However, one limitation is attention overload. If the reader struggles retrieving the verbal information from the text, it could lead to stress, where the smart glasses would starts to work against the goal. In this case the disruption of sonic augmentation proved to only distract and frustrate the reader more. Moreover, if there is no textual information to relate the sound to, the cue becomes purposeless. Nevertheless, engagement could be found using the smart glasses, as in the study frustration seemed to lead to a much higher increase in nose temperature than engagement, by around four degrees.

The sonic assistive augmentation took most participants by surprise, since they did not know when the next sound would come. Regardless the disruption, we noticed participants were able to modify their biological minds and integrate the sonic information to the information of the narrative text. We suggest sound, as music and literature, is consumed across time. To understand the narrative, one needs to link all available information. We saw some indication of this in the debriefing session, even when participants could not immediately relate the sound to the text, as they progressed in the story they could remember and associate those sounds to what was read before. The sonic assistance seemed to encourage participants to imagine the setting of the story or even to correct their perception of the narrative. Some of them trusted more the recognizable sounds than the words, giving more weight to the assistive augmentation.

References

1. Angelotti M, Behnke RR, Carlile LW (1975) Heart rate: a measure of reading involvement. Res Teach Engl 192–199
2. Ashby J, Rayner K (2012) Reading in alphabetic writing systems: evidence from cognitive neuroscience. Neurosci Educ Good Bad Ugly 61
3. Benton M (2005) Reader-response criticism. Understanding children's literature: key essays from the second edition of the international companion encyclopedia of children's literature, 86
4. Berka C, Levendowski DJ, Lumicao MN, Yau A, Davis G, Zivkovic VT, Olmstead RE, Tremoulet PD, Craven PL (2007) Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat Space Environ Med 78(Supplement 1):B231–B244
5. Bertelson P, de Gelder B (2004) The psychology of multimodal perception. Crossmodal space and crossmodal attention, pp 141–177
6. Biedert R, Buscher G, Schwarz S, Möller M, Dengel A, Lottermann T (2010) The text 2.0 framework. In: Workshop on eye gaze in intelligent human machine interaction. Citeseer, pp 114–117
7. Cannon WB (1927) The james-lange theory of emotions: A critical examination and an alternative theory. Am J Psychol 106–124
8. Cohen B, Izard C, Facial Simons R (1986) physiological indexes of emotions in mother-infant interactions. In Psychophysiology, vol. 23, Soc Psychophysiol Res, (1010) Vermont Ave Nw Suite 1100. Washington, DC 20005:429–429
9. Corradini A, Antonietti A (2013) Mirror neurons and their function in cognitively understood empathy. Conscious Cogn 22(3):1152–1161

10. Coulson S, King JW, Kutas M (1998) Expect the unexpected: event-related brain response to morphosyntactic violations. Lang Cogn Process 13(1):21–58
11. Csikszentmihalyi M, Csikzentmihaly M (1991) Flow: the psychology of optimal experience, vol 41. Harper Perennial, New York
12. Eason SH, Goldberg LF, Young KM, Geist MC, Cutting LE (2012) Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. J Educ Psychol 104(3):515
13. Ekman P, Levenson RW, Friesen WV (1983) Autonomic nervous system activity distinguishes among emotions. Sci 221(4616):1208–1210
14. Engle RW, Tuholski SW, Laughlin JE, Conway AR (1999) Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. J Exp Psychol Gen 128(3):309
15. Ericsson KA, Kintsch W (1995) Long-term working memory. Psychol Rev 102(2):211
16. Fludernik M (2002) Towards a 'natural' narratology. Routledge
17. Folkerth W (2014) The sound of Shakespeare. Routledge
18. Gibson JJ. The senses considered as perceptual systems
19. Gibson JJ (2014) The ecological approach to visual perception: classic edition. Psychology Press
20. Gladwell M (2007) Blink: the power of thinking without thinking. Back Bay Books
21. Goldman AI (2005) 2 imitation, mind reading, and simulation. Perspectives on imitation: imitation, human development, and culture, 79
22. Gouvea AC. How to examine the p600 using language theory: what are the syntactic processes reflected in this component?
23. Graesser AC, McNamara DS (2011) Computational analyses of multilevel discourse comprehension. Top Cogn Sci 3(2):371–398
24. Havelock EA (1980) The oral composition of greek drama. Quaderni Urbinati di Cultura Classica, 61–113
25. Healy SD (1984) Boredom, self, and culture. Fairleigh Dickinson University Press
26. Herman D (2000) Narratology as a cognitive science. Image Narrative 1:1
27. Hollan J, Hutchins E, Kirsh D (2000) Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans Comput Hum Interact (TOCHI) 7(2): 174–196
28. Houix O, Lemaitre G, Misdariis N, Susini P, Urdapilleta I (2012) A lexical analysis of environmental sound categories. J Exp Psychol Appl 18(1):52
29. Hutchins E (2010) Cognitive ecology. Topics. Cogn Sci 2(4):705–715
30. Ioannou S, Ebisch S, Aureli T, Bafunno D, Ioannides HA, Cardone D, Manini B, Romani GL, Gallese V, Merla A. The autonomic signature of guilt in children: a thermal infrared imaging study
31. Ioannou S, Gallese V, Merla A (2014) Thermal infrared imaging in psychophysiology: potentialities and limits. Psychophysiology 51(10):951–963
32. Jalilehvand M (2012) The effects of text length and picture on reading comprehension of iranian efl students. Asian Soc Sci 8(3):329
33. Jennett C, Cox AL, 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(9):641–661
34. Kadir D (2006) Comparative literature in a world become tlon. Comp Crit Stud 3(1):125–138
35. Keller J, Bless H, Blomann F, Kleinböhl D (2011) Physiological aspects of flow experiences: skills-demand-compatibility effects on heart rate variability and salivary cortisol. J Exp Soc Psychol 47(4):849–852
36. Kiesling LL (2012) Mirror neuron research and adam smith's concept of sympathy: three points of correspondence. Rev Austrian Econ 25(4):299–313
37. Kintsch W (1988) The role of knowledge in discourse comprehension: a construction-integration model. Psychol Rev 95(2):163
38. Kintsch W, Patel VL, Ericsson KA (1999) The role of long-term working memory in text comprehension. Psychologia 42(4):186–198

39. Kunze K, Sanchez S, Dingler T, Augereau O, Kise K, Inami M, Tsutomu T (2015) The augmented narrative: toward estimating reader engagement. In: Proceedings of the 6th augmented human international conference. ACM, pp 163–164
40. Lebrecht J, Kaye D (1999) The art and technique of design, sound and music for the theatre
41. Lemaitre G, Heller LM (2013) Evidence for a basic level in a taxonomy of everyday action sounds. Exp Brain Res 226(2):253–264
42. Lewis D (1978) Truth in fiction. Am Philos Q 15(1):37–46
43. MacDonald RA, Hargreaves DJ, Miell D, Davidson JW, North AC (2002) Musical identities, vol 13. Oxford University Press, Oxford
44. MacLuhan M (2005) The effect of the printed book on language in the 16th century. Gingko Press
45. Mar RA, Oatley K (2008) The function of fiction is the abstraction and simulation of social experience. Perspect Psychol Sci 3(3):173–192
46. McAdams SE, Bigand EE (1993) Thinking in sound: the cognitive psychology of human audition. In: Based on the fourth workshop in the tutorial workshop series organized by the hearing group of the French acoustical society. Clarendon Press/Oxford University Press
47. McCraty R, Atkinson M, Tomasino D, Bradley RT (2009) The coherent heart: heart-brain interactions, psychophysiological coherence, and the emergence of system-wide order. Integr Rev 5(2):10–115
48. Mcluhan HM (2010) Understanding me: lectures and interviews. McClelland & Stewart
49. Merleau-Ponty M, Smith C (1996) Phenomenology of perception. Motilal Banarsidass Publishe
50. Michaels CF, Carello C (1981) Direct perception. Prentice-Hall Englewood Cliffs, NJ
51. Moreno R, Mayer RE (2002) Verbal redundancy in multimedia learning: when reading helps listening. J Educ Psychol 94(1):156
52. Nakanishi R, Imai-Matsumura K (2008) Facial skin temperature decreases in infants with joyful expression. Infant Behav Dev 31(1):137–144
53. Or CK, Duffy VG (2007) Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. Occup Ergon 7(2):83
54. Paivio A (1978) A dual coding approach to perception and cognition. Modes of perceiving and processing information, pp 39–51
55. Pauletto S (2014) Film and theatre-based approaches for sonic interaction design. Digit Creativity 25(1):15–26
56. Penney CG (1989) Modality effects and the structure of short-term verbal memory. Mem Cogn 17(4):398–422
57. Pollard AW (2010) Shakespeare's fight with the pirates and the problems of the transmission of his text. Cambridge University Press
58. Purves AC, Rippere V. Elements of writing about a literary work—a study of response to literature
59. Rabinowitz PJ (1997) Before reading: narrative conventions and the politics of interpretation. The theory and interpretation of narrative series, ERIC
60. Rose PA (2012) The classical trivium: the place of thomas nashe in the learning of his time (by marshall mcluhan). Can J Commun 37:4
61. Ryan M-L (1991) Possible worlds, artificial intelligence, and narrative theory. Indiana University Press
62. Sadoski M (1983) An exploratory study of the relationships between reported imagery and the comprehension and recall of a story. Read Res Q 110–123
63. Schütz A (1951) Making music together: a study in social relationship. Soc Res 76–97
64. Scolari CA (2012) Media ecology: exploring the metaphor to expand the theory. Commun Theory 22(2):204–225
65. Shokoff J (2001) What is an audiobook? J Popular Cult 34(4):171
66. Sowa JF. Conceptual structures: information processing in mind and machine
67. Spacks PM (1995) Boredom: the literary history of a state of mind. University of Chicago Press

68. Thierry G, Martin CD, Gonzalez-Diaz V, Rezaie R, Roberts N, Davis PM (2008) Event-related potential characterisation of the shakespearean functional shift in narrative sentence structure. Neuroimage 40(2):923–931
69. Tribble E, Sutton J et al (2011) Cognitive ecology as a framework for shakespearean studies. Shakespeare Stud 39:94–103
70. Verstraete P. The frequency of imagination: auditory distress and aurality in contemporary music theatre
71. Walter O (1982) Orality and literacy. TJPress, London, The technologizing of the word
72. Weaver P (2012) Radio drama: a "visual sound" analysis of John, George and drew baby. PhD thesis, University of Central Florida Orlando, Florida
73. Zajonc RB (1985) Emotion and facial efference: a theory reclaimed. Science 228(4695):15–21
74. Zhao H (2012) Emotion-driven interactive storytelling. PhD thesis, Bournemouth University