Technische Universität Berlin
Fakultät I - Geistes- und Bildungswissenschaften
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation

Masterarbeit für die Prüfung zum Master of Science im
Studiengang Audiokommunikation und -technologie
an der Technischen Universität Berlin,
Fakultät I – Geistes- und
Bildungswissenschaften

Effects of auditory stimuli on visual attention and subsequent memory performance

David Du Bruyn
[redacted]

1. Gutachter
Prof. Dr. Stefan Weinzierl
FG Audiokommunikation
Technische Universität Berlin

2. Gutachter
PD Dr. Jochen Steffens
FG Audiokommunikation
Technische Universität Berlin

“Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.”— Marie Curie

Abstract

Human attention is subject to different influences from its surrounding environment. Attention is focused on objects of subjective relevance. Due to external influences, the focus of attention can change. Affect and emotion play a vital role in human attention, as changes in affect can change the subjective appraisal and relevance of surroundings and objects. Affect can influence the perception and appraisal of situations and objects, therefore, affective influences can change the focus of attention (Gasper & Clore, 2002; Wadlinger & Isaacowitz, 2006). Recent research has found that positive affect can broaden visual attention, with a focus on general concepts. In contrast, negative affect narrows down attention with focus towards details and local concepts. (Gasper & Clore, 2002) Further, high arousal is indicated to cause a narrowing of visual attention, in the direction of prior affective or stimulus induced affective state.

Similar to attentional processes is memory. Recent research has indicated that positive affect enhances memory performance, especially in short-term (Bradley, Greenwald, Petry, & Lang, 1992). Higher levels of arousal have indicated better memory performance in long-term memory. Memory under positive affect is indicated to follow a globally and more schematic representation, whilst negative affect influences memory for details.

Aim of this research was to find out how affective congruence of auditory and visual stimuli influenced visual attention towards affectively congruent visual stimuli and subsequent memory. Considering that memory performance is higher for stimuli actively attended to, in comparison to stimuli unattended, memory performance has been used as an indication for visual attention on stimulus. An online audio-visual memory experiment with 75 participants used affectively pre-rated film music excerpts (Eerola & Vuoskoski, 2011) and photographs (Kurdi, Lozano, & Banaji, 2017), of different affective ratings, to test whether a concomitant congruent auditory stimulus facilitates visual attention towards an affectively congruent visual stimulus. Collages containing 16 photographs with four photographs each per quadrant in the affective space (Russell, 1980), were presented with one music excerpt. Each participant carried out 12 trials. After each trial the participant was asked to indicate which pictures were memorized. In total 12 image collages have been paired with 12 different music excerpts. Four music excerpts were chosen from each quadrant in the affective space. Total amount of images used was 192.

Results indicate that the overall congruence of auditory and visual stimuli facilitates memory, indicating heightened visual attention towards affectively congruent visual stimuli. Valence congruence indicated a higher influence on memory performance than arousal congruence. However, high arousal visual stimuli have been better memorized, regardless of affect of auditory stimulus. Additionally, this effect was stronger in participants scoring high on the BFI openness score. Concluding, it is expected, that in the context of music and images, a mechanism of visual

attention guiding on the basis of affective feature congruence is underlying. A main influence is expected to come from semantic networks, being more readily activated in congruent audio-visual pairings, leading to heightened visual attention and therefore better memory performance in respect to congruent visual stimuli. Despite this highly contextualized experimental setting, the author expects this mechanism to hold true to at least some extend in real-life visual attention guidance, especially when music is listened to.

Zusammenfassung

Die menschliche Aufmerksamkeit wird durch diverse Faktoren der Umwelt beeinflusst. Aufmerksamkeit fokussiert sich auf Objekte von subjektiver Relevanz. Äußeren Einflüssen geschuldet kann sich der Fokus der Aufmerksamkeit verändern. Affekte und Emotionen spielen dabei eine ausschlaggebende Rolle. Das bedeutet: Veränderungen des Affekts können auch Änderungen in der subjektiven Beurteilung und Bedeutung von Umgebung und Objekten auslösen. Der Affekt kann die Wahrnehmung und die Einschätzung von Situationen und Objekten beeinflussen, somit können gefühlsbedingte Einflüsse auch den Fokus der Aufmerksamkeit verändern.

Aktuelle Forschung zeigt, dass positiver Affekt die visuelle Aufmerksamkeit verbreitert und auf weitere Stimuli mit neutraler oder positiver Valenz lenkt. Die Aufmerksamkeit unter positivem Affekt ist global und schematisch orientiert. Negativer Affekt lenkt die visuelle Aufmerksamkeit auf Details. Hohes Arousal verengt die visuelle Aufmerksamkeit und lenkt sie auf Details in Richtung der zuvor herrschenden Emotion. (Gasper & Clore, 2002; Wadlinger & Isaacowitz, 2006) Die Gedächtnisleistung folgt einer ähnlichen Struktur wie die Aufmerksamkeit. Positiver Affekt unterstützt die Kurzzeitgedächtnisleistung (Bradley et al., 1992). Positiver Affekt fördert ein schematisches und globales Gedächtnis, hingegen fördert negativer Affekt ein detailorientiertes Gedächtnis.

Ziel der Arbeit war es zu untersuchen, inwieweit Affektkongruenz zwischen visuellen Stimuli und gleichzeitigem auditivem Stimulus, die visuelle Aufmerksamkeit auf kongruente visuelle Stimuli lenkt und die nachfolgende Gedächtnisleistung für diese steigert. Ein audio-visuelles Onlineexperiment mit 75 Teilnehmern hat untersucht, auf welche Bilder einer Collage die visuelle Aufmerksamkeit eines Teilnehmers gelenkt wird, während der gleichzeitigen Wiedergabe eines

Musikstückes. Jeder Teilnehmer absolvierte 12 Versuchsdurchläufe. Jede Collage bestand aus 16 Bildern (Kurdi et al., 2017), welche bereits mit Affektbewertungen versehen worden sind. In jeder Collage befanden sich jeweils vier Bilder, die einem Quadrant des circumplex Modells (Russell, 1980) zugeordnet werden konnten. Die Musikstücke (Eerola & Vuoskoski, 2011) waren ebenfalls nach Affektausprägung bewertet. Je vier Musikstücke konnten einem der Quadranten des circumplex Modells zugeordnet werden. Nach jeder Stimulus-Präsentation wurden gemerkte Bilder abgefragt. Da Gedächtnisleistung für Stimuli höher ist, welche sich im Fokus der Aufmerksamkeit befinden (Desimone, 1996), wird die Gedächtnisleistung als Indikator für Stimuli verwendet, auf die die visuelle Aufmerksamkeit gerichtet war.

Die Ergebnisse der Untersuchung zeigten, dass die generelle Kongruenz einen positiven Einfluss auf die Gedächtnisleistung hat. Die Valenzkongruenz hat im Vergleich zur Arousalkongruenz mehr Einfluss auf die Gedächtnisleistung. Daraus wird abgeleitet, dass die visuelle Aufmerksamkeit in kongruenten audiovisuellen Paarungen auf kongruente visuelle Inhalte gelenkt wird. Zusätzlich zeigte sich, dass Bilder mit hoher Arousal Bewertung, ungeachtet der Affektausprägung der begleitenden Musik, besser gemerkt wurden, folglich also auch im Fokus der visuellen Aufmerksamkeit standen. Im Kontext von Musik- und Bildpräsentation zeigte sich ein Mechanismus, der die visuelle Aufmerksamkeit auf affektkongruente Inhalte lenkt. Zusätzlich zeigte sich dieser Effekt verstärkt bei Personen, die hohe Werte bei dem BFI Faktor ‚Offenheit' erzielten. Es wird vermutet, dass dieser Mechanismus auch zum Teil in anderen lebensnahen Situationen auftritt, insbesondere wenn Musik gehört wird. Zusätzlich wird vermutet, dass in kongruenten audiovisuellen Paarungen semantische Netzwerke leichter aktiviert werden, was wiederum die Aufmerksamkeit in der visuellen Modalität auf kongruente Stimuli richtet.

Eidesstattliche Erklärung

„Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.“

Berlin, den

Unterschrift

Inhalt

1. Introduction

The natural environment exists in a multimodal fashion. The perception of the environment by an organism, especially humans, is similarly organized as a multimodal construct. Humans use multiple senses, such as hearing and seeing, to orient themselves. All sensory modalities can function independently. However, Welch and Warren (1986) stated that despite their independence, all senses might not fully operate without affecting – or being affected by – other senses. Working together, different senses help to construct a complete picture of one's surroundings. Influences in the perception of one modality through to changes in another are called intersensory interactions.

To cope with the inflow of sensory information, attention occurs at the end of a series of complex processes. Because processing capacity is limited, attention helps humans to distinguish relevant from irrelevant stimuli. Depending on the context of perception, the process of paying attention can be influenced by many factors. The subjective relevance of stimuli varies, depending on cognitive states, which in turn can be influenced by the inflow of new sensory information. When the relevance of stimuli is subject to change, attention can also be subject to change.

Input from sensory modalities influences cognition and subsequently behaviour. To adapt to changes in the environment physiologically or psychologically, people can elicit emotions (Lane, Chua, & Dolan, 1999). In addition, affective responses elicited by input from one sensory modality can alter the perception and appraisal of input from another sensory modality. The stimulus relevance can be altered as well. Because attention thus depends on stimulus relevance, this research investigates the influence of affective-feature correspondence between auditory and visual stimuli in visual attention guiding.

Recent research has indicated that positive affect, as a construct to represent emotion, can broaden visual attention. Positive affect also generally directs visual attention towards global features. The broadening effect can occur regarding visual aspects having neutral and positive valence. By contrast, negative affect has been associated with visual attention being focused on details and local features. Positive affect influences memory towards a schematic representation, whereas negative affect appears to enhance memory for detail. (Gasper & Clore, 2002; Mather & Sutherland, 2009; Wadlinger & Isaacowitz, 2006)

The influence of affective responses due to the processing of secondary stimuli has been the focus of recent studies. These studies have investigated how the perception of the primary stimulus is altered by the presentation of a secondary stimulus. Other research has mainly focused on emotion induction with use of a secondary stimulus, prior to the presentation of the stimulus of interest or primary stimulus (cf. Gasper & Clore, 2002; Wadlinger & Isaacowitz, 2006). The primary stimulus is the stimulus of which the processing is investigated. There is a need to investigate more on the effects of a concomitant secondary stimulus on processing of primary stimulus. .

In light of recent research, this study explored the effects of concurrent auditory stimuli on visual attention and short-term memory. The research focused on how attention and subsequent memory in the visual domain are guided by auditory characteristics, specifically affective qualities. Because stimuli attended to are better remembered than unattended stimuli (Desimone, 1996), short-term memory performance was used to indicate visual stimuli attended to. Of special interest was how affective correspondences between auditory and visual stimuli guide one's visual attention towards congruent or incongruent, stimuli in the visual domain. The stimuli are said to be affectively congruent, when valence and arousal levels in both stimuli are similar or same. Congruence therefore defines a state of agreement. Correspondence of stimuli in both domains was examined on the basis of affect, in two dimensions of valence and arousal. Valence can be defined as the quality of pleasantness or unpleasantness of an experience (Scherer, 2004) or the intrinsic level of attractiveness or averseness (Frijda, 1986). Arousal is the level of physiological activation, as stated by Juslin and Sloboda (2001) or the level of an emotional response (Husain, Thompson, & Schellenberg, 2002).

It is expected that the affective-feature congruence between stimulus dimensions has a positive effect on visual attention and memory performance. That is, images are better remembered when the concomitant auditory stimulus has the same affective rating in at least one dimension (valence or arousal) as that of the image stimulus. An affective rating, is when a stimulus is rated along a rating scale, for a level of arousal or a level of valence. These ratings are usually obtained from large groups and averaged. Recent research has also indicated that images rated as highly arousing (images evoking high activation or alertness of viewer) are better remembered than low-arousal images, regardless of the affective features of the concomitant auditory stimulus.

The described expectation for cross-modal effects on visual attention and memory were investigated in this study by an online audiovisual memory experiment. On the basis that attention to a stimulus facilitates subsequent memory (Desimone, 1996), influences acting on attentional focus were derived from memory performance. In contrast to the reviewed research efforts, this experiment focused on the simultaneous presentation of auditory and visual stimuli. The results gave an indication of the direct cross-modal influences during stimulus presentation. The stimuli chosen for this investigation were taken from previous studies (Eerola & Vuoskoski, 2011; Kurdi et al., 2017) and were provided with ratings of arousal level and valence level (affective rating). The ratings of the stimuli were obtained in these previous studies. The advantages of pre-rated stimuli were, first, comparability between the stimuli, and second, comparison with past research results.

Considering the limits of this investigation and the specific context of the experiment, various disciplines can benefit from the findings of this research. The understanding of a contextualized aspect of human cross-modal perception and cognition can aid the development of new technologies aimed at human–environment interaction. It also serves as a basis for further research on intersensory influences in cognition and attention.

As this field of research is rather complex and is relevant to multiple disciplines, a brief overview of scientific advances is given. Indications for a model of attentional guidance in multimodal stimulus display are highlighted, in consideration of affective-feature correspondence. The method and results of the online multimedia memory experiment are provided. Preliminary concluding thoughts on the interplay of concomitant audiovisual stimulus processing, in light of affective-feature correspondence, are also given.

2. Theory and current state of research

The basis of this research is human perception and the influences between intersensory interplay and cognitive processes. A brief outline of the main concepts used in this research is given. The aim of this research is to gain insight into the influence of auditory stimuli on the focus of visual attention and subsequent memory. To understand this, it is important to discuss recent research on similar approaches and the general functionality of the sensory modalities. The following sections provide a brief overview of important advances in research.

2.1 Perception

Irtel and Goldstein (2011) provided a basic understanding of the fundamental processes of perception. Perception is a conscious sensory experience. Sensory organs receive a stimulus input, which is then transformed into electrical signals that represent the stimulus. These signals are transformed by the brain to produce the sensations of hearing, seeing and so on. This process is called transduction.

Perception is the mere uptake of sensory input. By contrast, according to Irtel and Goldstein (2011), recognition is a further step in processing that input. Recognition means the ability to categorize an object according to its purpose and meaning. Perception and recognition are thus two discrete processes. Perception can happen independently, without necessarily leading to recognition. An example occurs in brain-damaged patients who are able to perceive features of objects, but are not able to recognize the object itself. (Irtel & Goldstein, 2011)

Another step in the process of perception is the resultant action. Resultant actions are movements of the body or body parts. These movements can be voluntarily or involuntarily. (Irtel & Goldstein, 2011)

A further process of perception that occurs internally is neural processing. Neural processing refers to an alteration in the electrical behaviour of neurons. Further recognition and resultant action derive from internal processes, and are the most important results of the perceptual process. (Irtel & Goldstein, 2011)

The process of perception and making sense of the perceived input are forms of information intake. There are two parts to information input: bottom-up and top-down. Bottom-up information describes the stream of external information input, such as basic sensory perception. The other aspect is the top-down input of information, namely existing knowledge. (Irtel & Goldstein, 2011)

2.2 Dimensional model

Information presented by stimuli can be processed at a higher level too, which means that a stimulus can elicit an affective response. Emotions and moods can alter cognitive states and processes – for example, thoughts. This in turn influences the subsequent perception and processing of stimuli. When exploring influences on perception, it is thus vital to portray emotional aspects. To classify affective responses and stimuli, researchers have developed diverse theories and models.

Several models have been proposed (cf. Schimmack & Grob, 2000, Eerola & Vuoskoski, 2011) to rate affective responses. These models are mainly based on either a dimensional or a categorical approach. The categorical approach provides discrete emotions, whilst a dimensional model provides two or more dimensions, such as valence and arousal, from which emotions can be constructed. Therefore both approaches can be compared. A two-dimensional model by James A. Russell is widely used. Russell (1980) stated that affect can be described by a set of dimensions, such as pleasure and excitement. Russell stated that such dimensions are strongly related to each other and these relationships can be ordered along a circle in two-dimensional bipolar space. Russell (1980) suggested ordering affective concepts in this circular order as follows: pleasure (0°), excitement (45°), arousal (90°), distress (135°), displeasure (180°), depression (225°), sleepiness (270°), relaxation (315°). Thus, varying along the bipolar dimensions of valence (x-axis) and arousal (y-axis), experiences of affective responses can be represented in two-dimensional space as shown in Figure 1.

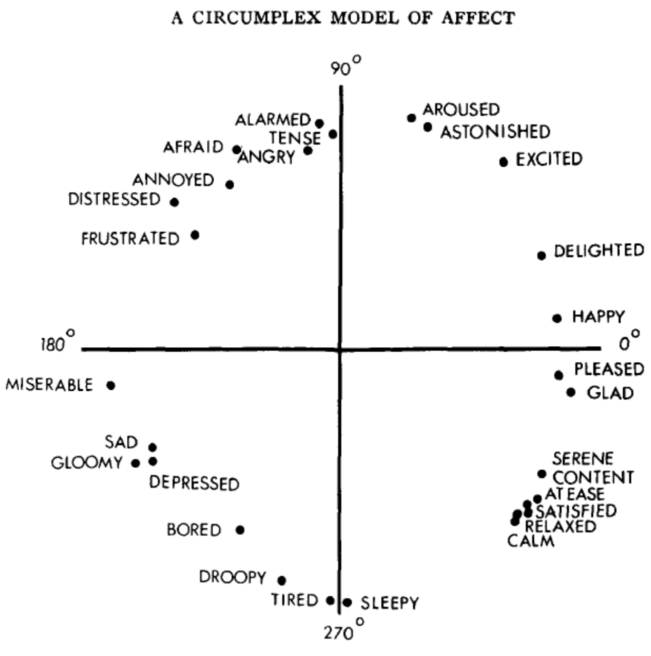

Figure 1 Psycho Circular scaling of mean (n = 36) rated arousal-valence coordinates of 28 affect words. From " A Circumplex Model of Affect " by James A. Russell, J. Personality and Social Psychology, 1980, 39, 1161-1178. Copyright 1980 by the American Psychological Association.

To classify emotions, independent neural systems for different emotions have been proposed. Instead of this view, the two-dimensional model assumes that all affective states arise from only two independent dimensions. One is the level of pleasure or displeasure, the so-called valence. The other dimension represents the arousal, defining the level of physiological activation. The theory suggests that when these are combined, a strong correspondence exists between the two elementary dimensions of valence and arousal. In addition, all emotions can be broken down into varying degrees of arousal and valence. (Eerola & Vuoskoski, 2011)

Despite the existence of different models to classify affective states, the two-dimensional model of valence and arousal has become common in recent research. Especially in the affective sciences and psychological research on music, the two-dimensional model is widespread. (Scherer, 2004) This is advantageous, as it allows for comparability between research efforts and stimuli. The two-dimensional theory is used in this study. The stimuli utilized in this research had been rated along these two dimensions (OASIS image stimuli). The film music excerpts have been rated in discrete emotions, which can be broken down into levels of valence and arousal (Film music excerpts Eerola & Vuoskoski, 2011). (For detailed description of the stimuli used, refer to chapter 3.2.)

Arousal, according to Sloboda and Juslin (2001), refers to the level of physiological activation in a person or to the level of emotional response (Husain et al., 2002, p. 153). Welch and Warren (1986) stated that arousal has a varying status, with bipolar characteristics ranging "from drowsiness to alertness and frantic excitement".

Valence, according to Scherer (2004), can be defined as the quality of pleasantness or unpleasantness of an experience. Valence of objects, events, or situations can be defined as well as the intrinsic level of attractiveness – or the opposite, averseness. (Frijda, 1986)

Eerola and Vuoskoski (2011) stated that two-dimensional models have been criticized for their inaccuracy, when differentiating between adjacent emotions in the affective space. However, discrete models have also been criticized for poor resolution in ambiguous examples, for which two-dimensional models have an advantage. (Eerola & Vuoskoski, 2011) Despite these criticisms, Eerola and Vuoskoski (2011) reported that in recent research on music and emotions, both discrete and dimensional models are frequently used.

2.3 Attention

Another important concept in human perception, which underpins this research, is attention. Attention is an important mechanism for any organism as it manages the allocation of the limited processing capacity of the brain. This is done by ordering sensory input according to the subjective relevance. (Broadbent, 1957)

Attention can change on the basis of different factors, such as the relevance of stimuli or the motivational state of the person. Attention can be selective; that is, it may focus on specific information while blending out the rest. The brain can effectively select relevant sensory inputs and enhance their neural representations. Thus, irrelevant stimuli can be ignored. Furthermore, selective attention occurs when the observer deliberately attends to certain stimuli or only certain features or aspects of stimuli, while ignoring irrelevant stimuli. (Amezcua, Guevara, & Ramos-Loyo, 2005) (Welch & Warren, 1986)

Visual attention is the process of directing attention in the visual domain. Visual attention, as is the case for attention in other domains, can voluntarily be directed towards stimuli or may be involuntarily attracted. Visual attention can be based on different domains of perceptual representation. Attention can be object-based, spatial or feature-based. Feature-based attention, for example, enhances parts of the image in the field of vision that are related to a certain feature. (Maunsell & Treue, 2006)

Recent psychological and neuroscientific research has made gains in understanding how human attentional processes work, and how these are susceptible to influence. Attention and memory are both focus areas in this research, as some mechanisms for both processes are similar and partially intertwined. Memory is often better for stimuli that have been actively attended to (Desimone, 1996) than for unattended stimuli. Hence, an investigation of attention and memory should be made in tandem. Firstly, advances of research in the field of attention and especially visual attention will be portrayed.

Recent investigations provide evidence that multiple factors can act upon attention. Depending on the stimulus, the degree of influence of secondary stimuli on perceptive processes varies. Generally, attention limits the number of stimuli processed in complex displays and selectively enhances processing of the stimuli that are selected. (Lane et al., 1999)

The relevance of attentional processes in connection with auditory stimuli, especially music, is that these stimuli can modulate motivational and emotional states. In turn this affects attentional processes. (Amezcua et al., 2005; Steffens, Steele, & Guastavino, 2016)

Gasper and Clore (2002) investigated on the effects of positive and negative moods in visual processing and the level-of-focus in visual attention. Central to their investigation was the affect-as-information theory and the level-of-focus theory, which have been the topic of research in recent years. Gasper and Clore adopted a functional view of affect. The affect-as-information theory presumes affect to be the conscious and accessible result of ongoing unconscious appraisal processes. Affect seems to influence the way a person judges their surroundings. Additionally, as long as affect is felt as being task relevant, affect can influence the subjective interpretation of accessible information and knowledge. Gasper and Clore argued for a level-of-focus theory, in which persons in a negative mood focus their visual attention on details, whereas persons in a positive mood focus their visual attention more on global aspects and general concepts. The results from Gasper and Clore (2002) suggest that positive mood enhances a general global perspective and the use of general concepts in many areas of cognition and perception. This leads to a more global, category-level visual attention. In contrast, negative mood and negative affective cues are thought to direct general perception towards local features and details, thus leading visual attention towards local details. Gasper and Clore argued that globally orientated attention is most common, as the average resting affective state is neutral with a tendency towards the positive.

As an example, Gasper and Clore (2002) stated that, with attention to category-level global cues, a person would consider the meaning of an artwork (such as a painting) in its larger context. In contrast, detail-orientated local attention could be focused on brush strokes and textures in a painting. One experiment by Gasper and Clore (2002) used visual stimuli in the form of geometric shapes. These shapes were individually made up of smaller geometric shapes (see Figure 2). For instance, a triangle made of smaller squares was shown to participants, next to a square made of small triangles. The participant was asked to indicate which of these two examples seemed to best represent a target stimulus (either a triangle or a square) (see figure 2). By considering the participant's choice, it was possible to see whether the participant focused on the individual small geometric shapes or on the resulting global shape. Thus, attention to global versus local features in visual attention could be determined. The main factor in determining attentional focus was that affective feelings indicated what qualities a participant attended to. However, Gasper and Clore did not use a secondary stimulus for mood induction. For the purpose of assigning participants to mood groups, participants had to write either about a positive life experience or a negative life experience, prior to the main experiment.

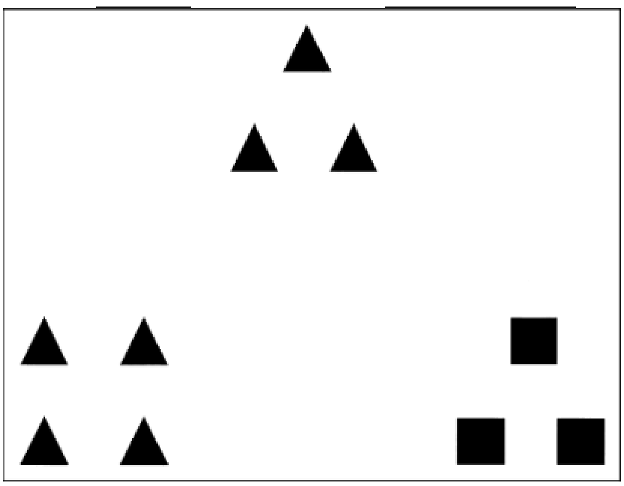

Figure 2 Sample item from the global-local focus test used by Gasper & Clore (2002). Original figure from Kimchi and Palmer (1982), p. 526. From (Gasper & Clore, 2002).

Gasper and Clore (2002) mentioned that other theories suggest that the influence of mood on visual attention arises from an altered processing capacity due to mood. Positive moods are theorized to facilitate a larger network of associations, whereas negative moods save resources for intensive processing. Thus, fewer global concepts are focused on with a negative mood.

As a basis for understanding how affective states can alter the level or focus of visual attention, the research by Gasper and Clore could highlight fundamental mechanisms. It is also of interest what visual contents a person’s attention is guided to, especially how secondary stimuli – such as auditory stimuli – influence this visual attention guiding. Of special interest is what connection might exist between the affect-eliciting stimulus and the visual stimulus attended to on the basis of affective congruence.

Influences on attention are usually indicated by shifts of attention, narrowing of attention or broadening of attention. A study on influences of positive mood on visual attention by Wadlinger and Isaacowitz (2006) indicated that positive mood influenced visual attention by broadening the attention paid to neutral and positive-valence pictures. This was indicated by frequent saccades – meaning quick eye movements between two or more points of fixation – when images having a neutral or low and medium positive valence were seen. The broadening of attention by participants in a positive mood has also been observed through participants looking longer at peripheral images having a neutral or positive valence. This study, however, presented only stimuli in the visual modality. A positive mood was induced by handing out candy to participants.

Arousal plays an important role in attentional processes. Typically, activation and arousal narrow the sensitivity of an organism so that only the most relevant cues are processed. (Easterbrook, 1959) Not only does positive mood evoke a selective broadening of visual attention, but emotional arousal also seems to selectively impact visual attention. Lane et al. (1999) researched the effects of valence, arousal and attention on neural activation in image processing. The results showed that emotional arousal modified the allocation of attentional resources. Arousal seemed to selectively broaden attention and heighten sensitivity to environmental cues, which is similar to motivational states being induced by the initial provoking stimulus. This selectively broadened attention was not only guided by the induced motivational states but also by current emotional states. Lane et al. (1999) used image stimuli to elicit affect.

Bishop, Karageorghis, and Kinrade (2009) suggested that much-liked and arousing music may be a cause of exploratory behaviour and tendency to action. The receptiveness to visual surroundings is also heightened. Another factor strongly contributing to attention is expectation. Not only are attention and perceptual sensitivity heightened towards the expected stimulus in the expected modality, but also for other modalities. Even when stimulus is unlikely to occur in certain modalities, sensitivity due to expectation remains heightened across modalities. (Driver & Spence, 1998)

The prioritized processing of selected stimuli is based on motivational relevance, thus further interpretations of visual stimuli are constrained by expectations. These expectations arise from the prior likelihood of certain events. (Summerfield & Egner, 2009) As previously stated, visual attention can be selectively broadened either by a current emotional state or by an induced motivational state. The question arises as to how these shifts generally happen. According to Maunsell and Treue (2006), visual attention can be directed to specific visual dimensions or features. These features include colour, orientation and the direction of motion. In an attempt to model how feature cues guide attention, Hamker (2004) demonstrated how attention can be guided and "locked" by features. Findings indicated that in an experimental setting, participants split their visual attention between two locations if two targets were behaviourally relevant and both contained the same feature, to which a participant was paying attention.

Further support for the idea that stimuli in one modality can attract attention in another modality was provided by Driver and Spence (1998). Their investigations showed that with the exception, that visual input did not lead to auditory attention under certain circumstances, a salient but irrelevant stimulus in one modality could lead to covert spatial attention in other modalities.

An example of attention guidance in an everyday context is that visual attention is attracted by sudden sounds in one’s surroundings. There are two forms of orientating oneself to attend to this sudden sound. Firstly, the process called overt orienting of attention. In this process one's physical orientation is changed towards the object that is attracting attention. Secondly, covert orienting, would be an internal shift in attention without physically reorienting oneself. (Shams & Kim, 2010) Recent research showed that not only is one’s visual attention drawn in the direction of a sudden sound, but the perception of subsequent visual stimuli that appear close to the attended location improves. (McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2003)

Concerning attention, it can be summarized that multiple factors act as influences. The secondary auditory stimulus does not itself guide the attentional and memory processes in the visual domain. Rather, the motivational and affective states elicited by auditory stimuli are responsible for such influences. Mood and affect are driving factors in determining the level of focus of visual attention. As discussed, positive mood leads to a broadened, more global and schematic attention, whilst negative affect is associated with narrow and detail orientated attention.

2.4 Memory

Memory enables a person to retain information it has taken up and to recall it later. Memory consists of different stages. After the uptake of information, the first step is the encoding of information. The second step is retaining the information, and the third step is to effectively recall stored information when needed. (Schermer, 1998)

According to Schermer, encoding occurs when a person devotes its attention and perception to a stimulus. Although this is important to memory, the process of retaining information that was taken in by the person forms the essential basis of memory. (Schermer, 1998)

Human memory systems have long been the focus of research. Many theories have been postulated in the last century. Several earlier theories proposed that memory was a unified system to store all information. Around the 1960s, newer research started to challenge this view and new models were developed. These new models abandoned the idea of a unified memory system and proposed multi-store models. In the multi-store model, information that is taken up can pass through all the stages, and this passing through the stages of memory is responsible for the encoding of information. (Schermer, 1998)

A prominent multi-store model is that of Atkinson and Shiffrin (1968). This structural model posits that memory consists of different stores with varying retention times, capacity and structure of encoding. This structure comprises sensory memory, short-term memory and long-term memory. Sensory memory takes up sensory stimuli, which can partially be transferred to short-term memory. The last part of this model is long-term memory, which stores information for long periods. (Atkinson & Shiffrin, 1968; Schermer, 1998)

A slightly different model was given by Craik and Lockhart (1972), who posited different levels of processing. Key in this concept is that memory performance is not due to different memory stores but rather to different levels of processing. This makes memory processing dependent on differences in stimulus processing. Analysis of stimuli happens at various levels, depending on the intensity and characteristics of the stimuli. In a flat-hierarchical processing, only simple physical and sensory features are attended to and processed, such as brightness, colour and contour. (Schermer, 1998)

Another store of memory is the working memory. Working memory is a subsystem that retains information while tasks such as learning, comprehension or reasoning are completed. (Baddeley, 2010) Two important stores of memory are episodic memory and semantic memory. The episodic memory stores information in a spatial and temporal relationship to other events within the episodic memory. In contrast, semantic memory refers to organized knowledge about relations and concepts. It is strongly influenced by cognition, as it does not directly store input received from the senses. Remembering the concept of a chair would prompt the representation of any chair, or associations to some object of similar typical use. Generalizing is a main use of semantic memories. (Tulving, 1972)

An important term when working with memory is the term "chunk". According to Simon (1974) it can roughly be described as information pieces. These chunks, however, are not restrained to individual information but can also refer to groupings of information material. A chunk can comprise a set of, for example, numbers or phonemes. Simon (1974) stated that a chunk of any sort of material is something that the short-term memory can hold five items of. The average chunk capacity was found experimentally to be between five and seven chunks.

Most newer models of human memory use the concept of a multi-store memory, implying different parts with various functional characteristics. The types of memory stores differ in their capacity for processable information. Other than capacity, different memory types also differ in their ways of encoding and the time they can store information for – or rather recall the retained information. (Schermer, 1998)

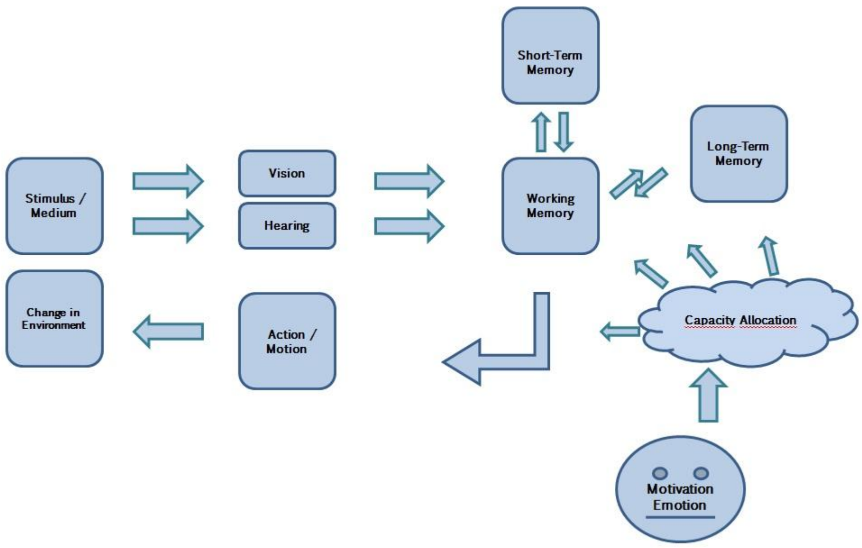

Figure 3 Memory structure and resource allocation (created by author based on Mangold (2016)).

In the most basic concept, arousal by an object seems to act as the most important factor for memory. Bradley et al. (1992) suggested that the level of arousal elicited by a source is responsible for good memory representation, regardless of the pleasantness of the source.

Mather and Sutherland (2009) stated that enhanced memory for emotional objects and intrinsic features of these are due to the arousal evoked by these objects. These findings were partially corroborated by Kensinger, Garoff-Eaton, and Schacter (2007), who showed that visually arousing objects elicited better memory for gist and for detail. Figure 3 schematically illustrates the influence of affective states on resource allocation and resultant memory performance.

Beyond arousal, more complex factors also contribute to memory. Kitayama and Niedenthal (1994) stated that semantic priming is a facilitative concept for memory and other cognitive tasks. When a semantic representation of affect is activated due to a stimulus, it affects cognitive processes and thoughts. This can possibly affect attention in a semantically congruent fashion.

In relation to memory, semantic priming elicits mood-congruence effects that are a facilitating factor in memory encoding and decoding. Affect can trigger links to parts of the semantic memory, hence activating thoughts and memories congruent to affect. These mood congruence effects, according to Kitayama and Niedenthal (1994), enable better memory encoding and memory retrieval for material that is congruent with a person's mood. (Kitayama & Niedenthal, 1994)

A study by Lehmann and Murray (2005) found that participants' ability to accurately detect image repetition during a recognition and memory task increased when pictures were initially presented with a semantically congruent sound. Participants were presented with images that were either new to the participant or that have been priorly presented with a semantically congruent sound. Participants were to detect whether an image has been priorly presented. The presentation of a congruent sound evidently facilitated the memory of the image objects. Interestingly, however, such ability was not affected by prior presentation with incongruent sounds, but impaired when initially presented with a 1000-Hz tone.

Recent research has identified arousal and semantic factors as main influences on memory performance. In addition, several researchers have shown that pleasantness impacts memory performance as well. Concerning the question of how valence or pleasantness and arousal affect memory, Bradley et al. (1992) conducted experiments to investigate memory performance. They investigated the memory performance for both short-term and long-term, with the long-term period being one year. Participants rated images according to valence and arousal and were asked to indicate which images they memorized, shortly after the rating session. After one year participants were asked to indicate images they memorized from the initial session. This investigation provided evidence that memory performance is affected by both valence and arousal of initial stimuli. However, the arousal dimension rendered stable results. Highly arousing picture stimuli were remembered better than low-arousal stimuli for both long- and short-term recall. Pleasant material had an advantage over unpleasant material in both low- and high-arousal combinations for short-term recall. This effect was not evident for long-term recall, where highly arousing stimuli, whether pleasant or unpleasant, were remembered best. Certain categories of pictures were also remembered better than others. For example, pictures of attractive members of the opposite sex were better remembered than weapons. The researchers indicated that this effect was more related to the gender of a participant than to the arousal rating. Hence, certain categories were better remembered by each gender.

Bradley et al. (1992) mentioned the findings of Maitlin and Stang and dubbed them “the Pollyanna effect”. This effect indicates better memory for pleasant materials. The experiments by Maitlin and Stang used only verbal stimuli.

Mather and Sutherland (2009) also stated that affect plays a vital role in memory. They argued that positive emotions lead to better schematic memory whereas negative emotions tend to lead to a more accurate memory for detail. These findings are similar to the previously mentioned findings by Gasper and Clore (2002) regarding levels of focus. The latter work indicated that positive affect seems to direct visual attention to global features and concepts, whereas negative affect directs visual attention to details.

In conclusion, arousal by stimuli seems to have a strong effect on memory recall. Bradley et al. (1992) suggested that highly arousing stimuli would facilitate an elaborate representation of elements associated with the elicited arousal. Therefore, poorer representations and thus poorer memory performance are associated with peripheral information that is unrelated to arousing aspects. This was corroborated by earlier findings by Walker (1958), who postulated the action decrement theory, suggesting that highly arousing events cause a stronger consolidation process (Bradley et al., 1992). This leads to increased long-term memory. However, memory retrieval is inhibited during this consolidation process, which in turn seems to cause poorer short-term memory results. Impairment of short-term memory for arousing material was not supported by the research of Bradley et al. (1992).

In a study that investigated the impact of visual information on music perception and memory, Boltz, Ebendorf, and Field (2009) suggested that the role of music in film is supportive. They stated that their findings indicated that music in film served to highlight certain information, which led to a more integrated memory that facilitated later recall of the visual information in the film. These findings underline that multi-sensory integration is an important factor for judgement of visual aspects and memory performance.

Proverbio et al. (2015) reported that emotionally touching music and silence improved the recall of faces in a recognition test. They hypothesized that emotionally touching music modified people’s visual perception in memory processes, by binding the facial properties with emotional and auditory information. Emotionally touching music, and to some extent joyful music, were shown to lead viewers to perceive pictures of faces as more arousing than in a no-music condition. In contrast to touching music or silence, joyful music or background sounds interfered with memory encoding. Furthermore, Proverbio et al. (2015) stated that multimodal display led to better binding in memory. Such would be the strengthening of memory engrams (the trace of a memory and hypothetical structural brain alteration that follows learning) through music, to facilitate multimodal audiovisual memory encoding.

Boltz (2001) found similar indicators. Three ambiguous film scenes were shown to participants, accompanied by music of either positive or negative valence. A no-music condition served as the control group. In contrast to the control group, accompanying negative or positive music influenced participants' interpretations. Memory of film objects was observed in a music-mood-congruent fashion.

Additionally, music seems to prove its facilitating role in perception. According to Yamasaki, Yamada, and Laukka (2015), the perception of emotions of film characters by viewers was consistent with particular emotions of concomitant music. Interestingly, several acoustic characteristics have been associated with certain emotions, making a cross-modal connection (Boltz et al., 2009). These cross-modal connections, that have been identified, possibly ease the detection and induction of emotion in music

Additionally, environmental perception can be regulated by musical features. Iwamiya (1997) showed that not only emotional characteristics affected the impression of landscapes; musical tempo, for instance, also altered the way a person judged the environment. This was observed in a music-feature-congruent way. This effect especially applies in incongruent settings. When an environment is otherwise calm, activating music can alter that impression, making the surrounding perceived as more active. Already active surroundings such as busy intersections, however, were not perceived as significantly more active when accompanied by active music.

Generally, secondary auditory stimuli do affect the perception of simultaneously viewed visual stimuli. The perception of ambiguous content can be altered in the direction of a perceived or induced emotion. Not only does arousal limit attentional resources to the most important cues for processing, but memory is affected similarly.

Research has shown that at a basic level, attended stimuli are better remembered than unattended stimuli. Arousal seems to be the basis of memory performance, regardless of affect. Aspects such as semantic priming and mood-congruence effects further influence which stimuli are remembered. Concerning mood congruence, two modes of interference seem to be at play. First, enhanced memory for objects that are mood-congruent can be seen in some conditions. Second, memory for and interpretation of visual scenes are noted to follow a mood-congruent direction.

The effects of valence have not been conclusively explored; it seems that context strongly affects the functionality of affective valence with regard to memory. Positive rather than negative affect tends to boost short-term memory and immediate recall. However, these effects are lost in the longer term, where high-arousing objects are best remembered, regardless of valence.

Similar to the effects of visual attention are the effects of mood on detail and structure of memory. Positive affect is thought to be linked to schematic representation in memory, whereas negative affect enhances memory for detail.

Generally, bimodal audiovisual presentations of stimuli, and highly arousing stimuli, have been found to best facilitate memory encoding, partially due to enhanced binding. Additionally, semantically congruent object pairings are well remembered.

2.5 Cross-modal correspondence

An important aspect of intersensory interactions is cross-modal correspondence. A literature review of recent research indicates that many different and contextualized reasons may underlie the connection in sensory processing. Although beyond the scope of this research, a brief introduction to cross-modal correspondences is given here. The aim is to highlight possible underlying mechanisms in aspects discussed in this research.

Regarding intersensory interactions, science has offered several theories and suggestions in recent years. Mondloch and Maurer (2004) found cross-modal correspondences between the visual and auditory modalities. In experiments where participants were asked to match objects with fitting sounds, adults matched high-pitched tones with small and bright lights; they also matched high-pitch tones with lighter coloured squares. The findings of Mondloch and Maurer (2004) were corroborated by experimental findings of Evans and Treisman (2010). They reported experiment participants to make positive relations between the visual feature of vertical position versus the auditory feature of pitch. Participants assigned high pitch tones to objects of high vertical position.

Interestingly, despite lacking cultural and musical experience, 6-month-old babies showed similar cross-modal correspondences. Young infants matched up or down sweeping tones with arrows pointing either up or down. (Evans & Treisman, 2010) Early research efforts by Carl Stumpf in 1883 found similar correspondences between verbal qualities and the sensation of pitch. Stumpf found that all world languages use the respective labels of "high" and "low" for the pitch of tones. (Evans & Treisman, 2010)

Spatial and temporal similarity or coincidence is usually a prominent cue for the integration of sensory inputs. However, the brain also seems to rely on feature correspondence between different sensory inputs. (Evans & Treisman, 2010) Similar effects have been observed for the impression of a landscape as influenced by concomitant music. The various effects that music had on the impression of a landscape, according to Iwamiya (1997), were based on perceived similarities between the auditory and visual aspects. Such facilitated cross-modal integration through feature correspondence is expected to be a vital mechanism that underlies congruence effects

Furthermore, cross-modal similarities or congruence plays a facilitating role. In experiments on natural object identification, Schneider, Engel, and Debener (2008) examined cross-modal priming. The results showed shorter reaction times for object identification when visual objects were paired with a congruent sound. Not only does multimodal information processing give a broad impression of an organism’s surrounding; it also facilitates perceptual processes. An accurate and efficient representation of information is the result of multimodal processing of information.

However, Evans and Treisman (2010) experimented on cross-modal mappings between visual and auditory stimuli, and reported that congruence effects occurred even when irrelevant to a given task. This raises assumptions that cross-modal congruency effects do not only occur as conscious processes but seem to operate at many levels of awareness. Findings by Shams and Kim (2010) corroborate that multimodal input can affect visual perception in many ways. Multisensory stimulation does not only influence cognitive processing but also the basic properties of visual perception. These multisensory enhancements even occur when they are not task-relevant and not directly related. For the brain, however, to place events or objects into causal relation it is important that these are in spatial as well as temporal proximity. This was indicated by the findings of Lewald and Guski (2003).

Evans and Treisman (2010) suggested a possible reason for some cross-modal correspondences. For example, auditory loudness and visual brightness might correspond due to similar neural codes. Brightness and loudness could be assumed as cross-modal analogies; for example, an increase in each represents an increase in stimulation. Similar underlying mechanisms could occur at a basic level.

Concerning cross-modal correspondences, it can be summarized that all the senses interact to give a complete impression of the world. The interplay of the different sensory modalities not only paints a fuller picture of the world but also facilitates the processing of different modalities, such as object discrimination and orienting. Recent research has indicated that certain impressions and associations which humans have with certain aspects and features might be inherent in the human brain. Further, recent research has shown connections and correspondences between the different modalities. The question arises as to what extent these cross-modal correspondences influence the perception of a person's surroundings. As a basis for future research on attention and memory, this study could be extended to further explore the relationship between affective-feature congruence and cross-modal correspondences. Even though beyond the scope of this research, cross-modal correspondences should not be neglected when portraying effects of congruence.

2.6 Conclusion regarding research to date

Comparing processes of visual attention and memory, some similarities can be observed. Recent research has indicated that in combination with positive affect, the focus of a person's visual attention is broader and category-level oriented. Negative affect is associated with attention to the local level and to detail. Memory for visual stimuli has been found to follow a similar pattern as attention, where positive affect enhances the memory of stimuli at a global level. In contrast, negative affect has been found to enhance memory for detail.

Arousal seems to have the main effect on memory performance. Several researchers suggest that arousing stimuli and states of arousal contribute to enhanced encoding of information. Arousing primary or secondary stimuli can enhance memory performance for objects through multimodal binding. This relationship has been especially noted for long-term memory.

Highly arousing objects have been reported to enhance memory performance, regardless of valence. Concerning visual attention, arousal narrows down the focus of attention in a mood-congruent fashion, by directing visual attention towards stimuli corresponding to initial emotional or motivational states. Concluding the current state of research, a model of attentional focus can be assumed in the context of affective influences.

It has been shown in different contexts that secondary stimuli in the form of music can influence general perception in a mood-congruent fashion, especially with emotionally ambiguous content.

After singular auditory stimuli, the impression of a landscape may be perceived in a mood-congruent and primary-stimulus-congruent fashion. In the context of affective influences, a relationship between stimulus, attentional directivity and memory has been observed.

The secondary stimulus elicits affect, which in turn leads to unconscious or partially conscious appraisal and judgement processes. The outcome of these processes determines the affective state of the person. This point is critical for further attentional processes. Here the affective state or the perceived affect a person focuses on can determine which mechanism is initiated. Considering the level-of-focus theory, affect determines whether the attention of a person is broadened to a more general and schematic representation or is focused on local features and details. Two important connections can be observed at this point. The narrowed-down and detailed attention seems to be based on the efficiency of human attentional processes, allowing only the most relevant cues to come into focus, saving resources for in-depth processing. In contrast is the influence of positive affect, leading towards a more general perception and visual attention. This visual attention is broadened to include further stimuli of neutral or positive valence, acting in a mood-conserving fashion. Additionally, a person's average mood is said to be neutral towards positive. This average mood leads to a generally and globally orientated visual perception. This is in balance between the extremes of negative and positive affective influence.

In addition to purely affect-based appraisal mechanisms, semantic priming can also affect attention and thus memory. Perceived affect or affective states can activate a network of semantic memory, subsequently altering cognition and thoughts. Not only is the simple affective response leading attention, motivational relevance and cognitive states are further influenced by these semantic networks of cognition. Depending on which semantic aspects are activated, attentional focus is altered accordingly.

The strength of semantic connections has also been shown in the context of memory, indicating that not only bi-modal but congruent – especially semantically congruent – bi-modal presentation can enhance memory encoding and retrieval. This strengthened encoding of memory has proven to be effective with bi-modal presentation that uses emotionally touching auditory stimuli. Auditory stimuli, especially music, guides perception in an emotionally congruent direction, as was shown regarding the perception of landscapes and film excerpts. The latter is also expected to be based on similarities between auditory and visual impressions. Similarities in neural coding are a possible reason for these effects of similar impressions.

Another factor contributing to attention is expectation. The reviewed literature suggests expectation to have a connection to semantic networks, by activating these networks. The mechanism through which attention is evoked by affective influences is conservative for both current mood and attentional resources. Activated semantic networks further direct general cognition in a mood-congruent way. Expectation for stimuli in a mood-congruent fashion could possibly strengthen attentional mechanisms.

Bi-modal stimulus presentation does not only guide perception and attention but strengthens the neural representation and subsequent memory due to enhanced encoding. Strengthened memory engrams through initial bi-modal presentation thus facilitate better consolidation of memory objects and better retrieval of those objects.

Memory performance can give indications about prior attentional focus, as generally objects attended to are better memorized than unattended objects. However, memory is susceptible to bias, as it is expected that arousal by an object such as an image can be the source of good memory representation, regardless of influence by a secondary stimulus.

The contextualized nature of memory encoding has shown that different stimulus qualities act upon possible retrieval times for memory and object encoding. Pleasant material is more readily retrieved from short-term memory, regardless of arousal, whereas in the long term, arousal is a driving force for enhanced memory. Thus, semantic networks have been found to enable better encoding in memory.

Without regarding the possibilities of bias in memory encoding, attention and memory share similar features. Enhanced attention and enhanced memory both occur in a mood-congruent fashion. It is assumed that perceived emotions can render similar results as induced emotions, with possible smaller effects.

Considering cross-modal correspondences, a review of recent research showed interesting connections. A semantically congruent fundamental connection seems to underpin human nature regarding auditory and visual feature perception. It seems that these connotations influence cognitive structures and associations. If this fundamental possibility does underlie perception, then this also indicates that attention could occur through semantic and affective influences.

2.7 Hypotheses and theoretical approach

It is hypothesized that visual attention is drawn to visual stimuli of equal valence and arousal as the concomitant auditory stimulus. This results in memory for these visual stimuli (H1).

Furthermore, it is expected that valence congruence has a greater effect on memory performance than does arousal congruence (H2).

Besides the described cross-modal effects, it is hypothesized that high-arousing visual stimuli are better remembered than others, regardless of the auditory stimuli present at the initial stimulus presentation (H3).

The review of recent research has shown significant advances in the field of attention and memory. Despite the diverse background and goals of these studies, the direct influence of one stimulus on the perception of a concomitant other has not received much attention to date. A study on the direct effects of a secondary stimulus on visual attention and memory is, therefore, needed to further understand immediate relationships and influences between these two modalities, and their perception and processing.

This study did not use an experimental control condition for memory performance without auditory stimulus, as this has mainly been the focus of a study by Bradley et al. (1992). Their study explored how image ratings varying along the dimensions of valence and arousal, influenced memory performance.

This research was conducted on the basis of affective qualities, as this allows analysis of influences at a cognitive level. Considering the amount of research in similar fields, the choice for an affect-based, cognitive approach for this research is beneficial, as the results can be compared to other research findings. Furthermore, the availability of tried and tested stimuli was useful as these have been specially designed for such research purposes. Additionally, the cognitive approach is easily contextualized and results are close to real-life cognitive processes. However, a psychoacoustic approach for auditory stimuli could have been beneficial to investigate this topic at a different level.

An important aspect to portray is to clarify what is meant by referring to emotions and affect in a person. Primarily, the distinction is made between emotion perception and emotion induction. Many scientific publications neglect this differentiation, and everyday language does not distinguish between the two concepts. However, when dealing with emotions in a research context, it should be clear which concept is referred to. This research uses the concept of perceived emotions, instead of relying on induced emotions, for two reasons. First, the lack of control of the experimental setting restricts reliable conclusions about whether an emotion can be or has been adequately induced, and whether this emotion induction has triggered the expected emotional response in the participant. Secondly, controlled conditions for emotion induction are highly personalized and can bias the results. Gabrielsson (2002), for example, stated that personally salient features of music can cause a heightened emotional response in a person. These features, however, might not cause any emotional response in another person. This bias could possibly be minimised through statistical analysis and an adequate sample size. Nonetheless, in this research, working with the concept of perceived emotions was the best choice. Basic emotional concepts and unambiguous ratings of stimuli were mainly used for this experiment. It was thus expected that all the emotions would be grasped by participants. In this experimental setting perceived emotions are still expected to deliver adequate results as compared to an experimental setting with reliably induced emotions. However effect sizes are expected to be smaller in this experimental setting. Perceived emotions are, however, still expected to trigger neural and cognitive processes, which underlie the focus of this study.

3 Method

3.1 Participants

The sample group of participants in the experiment included 27 women and 48 men (n =75) with a mean age of 33.05 years (*SD = 12 years*). Participants were recruited by inviting them through a student mailing list of the FG Audiokommunikation, TU Berlin. Furthermore, social media and leaflets were used to recruit participants.

The overall study took 11 months to complete. The experiment was online for 5 months. The recruitment of participants happened simultaneously to the experiment. Recruitment of participants started with initial activation of the online experiment and ended on the last day of the experiment.

3.2 Materials

To investigate the effects of auditory stimuli on attention and subsequent memory of concomitant visual stimuli in a bi-modal presentation setting, an online audiovisual memory experiment was conducted. This method was chosen because it allows for observing possible effects on memory of visual stimuli, and thus allows deductions to be made regarding the focus of visual attention.

The online audiovisual memory experiment presented participants with 12 collages of images, each with a concomitant auditory stimulus. A subsequent memory recall question followed stimulus presentation. Each cycle of audiovisual presentation included one collage and one auditory stimulus. The online experiment was conducted in the German language as the participants were mainly German-speaking.

3.2.1 Image stimuli

The image stimuli were taken from the open affective standardized image set (OASIS). (Kurdi et al., 2017) The OASIS set was used as it offers a large variety of images, which are provided with affective ratings in terms of valence and arousal. Furthermore, the OASIS stimulus set is free to use and has been specifically compiled to serve research in the affective sciences.

Each individual picture of a collage had a size of 500x400 pixels, and the final collages each had a size of 2000x1600 pixels. The final video had a size of 1280x720 pixels. However, because participants accessed the experiment via the internet-based software LimeSurvey, the actual displayed sizes of the final stimuli might have varied. At the beginning of the experiment, participants were instructed to set the size of the visual stimuli (video) to three-quarters of the display size, allowing the visual stimuli to be seen completely without having to scroll.

Each collage contained 16 individual images and these images were randomly assigned a place in the collage. The images in the collage could each be allocated to one quadrant in the circumplex model. Four images of similar affective rating represent one quadrant. All images selected were chosen from the outer third of each quadrant, so that each dimension was at its extreme values as much as possible – for example, high-arousal and high-valence, or the opposite. For further statistical analysis these similar ratings have been coded as high (1) or low (0). Arousal and valence, hence, each had two possible values, either (1) or (0). This was done to avoid ambiguous stimuli and to ensure comparability between stimuli. A large number (16) for the individual images in each collage was chosen to exhaust memory-chunk capacity and avoid possible bias in memory tasks (cf. Simon, 1974).

3.2.2 Auditory stimuli

The 12 auditory stimuli were film music excerpts and were taken from a study by Eerola and Vuoskoski (2011). This set of music excerpts was used as each auditory stimulus is already allocated to a discrete emotion. The discrete emotions thus enabled comparability to the image stimuli and were translatable to affective valence and arousal ratings in the affective space. The four discrete emotions used were happiness, sadness, tenderness and fear. Each discrete emotion was allocated to a part of each quadrant in the circumplex model, which allowed for correspondence with the affective ratings of the image stimuli. For comparison to the visual stimuli and for later statistical analysis, the discrete emotions of the auditory stimuli have been re-coded in terms of affective rating. The translation to an affective rating was done on the basis of the discrete emotion's position in the circumplex model. Each auditory stimulus was given a binary code for arousal and valence. These ratings have been coded as high (1) or low (0). Arousal and valence, hence, each had two possible values, either (1) or (0). Translation to either high or low was possible, as only auditory stimuli were used of which the conveyed emotion was rated as high. The original stimuli were classified with the emotion of fear, sadness, tenderness and happiness. See appendix C for details.

In total 192 images were used from the OASIS image set, and 12 music excerpts from Eerola and Vuoskoski (2011) were created. In total, 24 music/collage pairings were created. The 12 auditory stimuli were used in both stimulus groups. Each participant was assigned to one of two stimulus groups and each participant was presented with 12 pairings. The two stimulus groups contained different collages. However, all collages have been composed from the same collection of 192 images from the OASIS image set. The images in the second stimulus group have been mixed among the collages, thus keeping four images per quadrant of the affective space in each collage. The collage/music pairings were rendered as a video file and uploaded to an online streaming platform as private files and were then embedded in the survey.

3.3 Software

To conduct the audiovisual memory experiment, the online survey platform LimeSurvey (Version 2.64.7+170404) was used. The LimeSurvey platform was chosen as it allows different formats of questions to be incorporated and media to be embedded. It also provides well-structured data collection and customizable response schemes. (For screenshots of the survey refer to appendix A and appendix F)

3.4 Research design

The independent variable was the affective congruence between the auditory and visual stimuli in the two dimensions of valence and arousal. The independent variable had two possible values: *congruent (1) or not congruent (0).* The dependent variable in the experiment was *stimulus memorized (1) or stimulus not memorized (0).*

3.5 Procedure

3.5.1 Instructions to participants

On their arrival at the experiment's platform webpage, participants were instructed about the general aim of the experiment. They were told that the experiment investigated the effects of musical features on visual perception. Further information was held back until the end of the experiment. Prior to the main stimulus presentation, participants were presented with an audiovisual test stimulus and instructed to set the audio volume to a pleasant level, which allowed details in the music to be heard. Furthermore, participants were instructed to adjust the size or zoom of their on-screen material, so that the visual stimuli covered three-quarters of the physical screen, without the need to scroll.

3.5.2 Informed consent

Participants were presented with initial information about the experiment and that data obtained in the experiment would be solely used for research purposes. Furthermore, participants were informed that all data was collected anonymously and that no link can be made between data and individuals. By clicking to enter the experiment, participants agreed to the terms stated.

3.5.3 Assignment to experimental conditions

Participants were assigned to one of the two groups of stimuli. The assignment to a stimulus group was done automatically and randomly by the experiment software. The amount of participants assigned to each stimulus group was balanced.

3.5.4 Task

Participants in the online experiment accessed LimeSurvey via a link. After reading the introductory information and giving consent, participants were asked to answer basic demographic questions. They were then asked to rate their current mood, level of calmness and whether they felt drowsy or wide-awake, on a scale from 1 to 5 (with 1 being least). Additionally, participants were asked to fill out the short form of the big five inventory, the BFI-10 (Rammstedt, Kemper, Klein, Beierlein, & Kovaleva, 2013). The BFI-10 was included to allow possible conclusions about the data to participants' personalities. See appendix D for the questionnaire.

The main task consisted of 12 audiovisual presentations with subsequent memory recall. The audiovisual presentation started with an auditory stimulus while a fixation dot appeared in the centre of the display for 5 seconds, directly followed by the image collage. The image collage was then presented for 5 seconds. Each audiovisual presentation lasted 10 seconds in total. Participants were instructed to focus their vision on the dot at the beginning of each trial. Each stimulus presentation was followed by an open answer question, asking the participant which of the individual images were remembered. After completing the question, the next audiovisual stimulus pairing started. Each participant was presented with 12 audiovisual stimulus pairings with a subsequent question on images memorized. The last question before debriefing and dismissal, asked whether any of the music pieces was familiar.

3.5.5 Duration

Participants generally concluded the experiment within 10 to 15 minutes.

3.5.6 Debriefing and dismissal

Participants received written debriefing in the last slide of the online experiment. They were thanked for taking part in the experiment and a short paragraph of information appeared about the aim of the research. Further contact details were given, which participants could consult if any questions arose.

3.6 Statistical Analysis

The data collected in the online experiment was processed for statistical analysis. Participants' responses to the images they had memorized were restructured and recoded. This resulted in a binary response of "memorized" or "not memorized" (1 or 0) for each image. Valence and arousal ratings for both the auditory and the visual stimuli were re-coded to binary values (1-high and 0-low) (for details refer to stimulus description). Due to the binary nature of the data and to account for random effects that influenced the responses, generalized linear mixed models (GLMMs) were used to analyse the data. The GLMM analysis was performed using the lme4 package for R (Bates, Mächler, Bolker, & Walker, 2015).

To test H1, a GLMM was calculated using the full congruence variable, being both valence and arousal congruence as one variable, as a fixed effect. The full congruence variable indicated whether an audiovisual stimulus pairing was overall affectively congruent. To be overall affectively congruent, valence and arousal of both - the auditory and the visual stimulus - had to match in their binary rating. There were 16 values for congruence in each trial. The model for H1 contained the image ID of the individual images and participant ID as random effects (see model 1 in appendix B).

To test hypothesis H2, the same GLMM was used; however, the full congruence was replaced by two fixed effects, namely valence congruence and arousal congruence. The random effects were the same as for the GLMM in H1 (see model 2 in appendix B). H3 was tested with the same model as that used to test H1, but the fixed effect was changed to contain only arousal from the individual images (see model 3 in appendix B). To examine the possible influence of personality on memory performance and attention, a further GLMM was calculated, using the factors of the BFI as fixed effects (see model 4 in appendix B).

To assess the coefficient of determination, the R^2_{glmm} was used. This calculation provides two measures of R^2, namely R^2 marginal and R^2 conditional, specifically for use with GLMM (Nakagawa & Schielzeth, 2013). The R^2 conditional ($R^2_{glmm, (c)}$) was used in the data analysis, as it can be interpreted as the variance explained by the entire model. The delta value will be given.

4 Results

4.1 Effects on dependent variables

A significant positive influence of full congruence on memory performance was observed *($p = .043$)*, with an effect size for $R^2_{glmm, (c)}$ of 0.182 for model 1. The individual images as random effects influenced the model with a variance of $\sigma^2 = 1.187$. The individual participants, as a random effect only, had a variance of $\sigma^2 = 0.177$. See model 1 in appendix B for further details.

Model 2 indicated a significant effect of valence congruence on memory performance *($p < .001$),* whilst arousal congruence had no significant effect on memory performance *($p > .489$).* The effect size of this model is 0.178 for $R^2_{glmm, (c)}$. Valence congruence showed a higher estimate (0.294) (odds ratio = 1.341) than arousal congruence (0.054) (odds ratio = 1.055). The random effects of participant ID ($\sigma^2 = 0.177$) and the individual images ($\sigma^2 = 1.142$), remained close to that in model 1. See model 2 in appendix B for further details.

Arousal of image had a highly significant effect on memory performance *($p < .001$)*. The individual image as a random effect had a variance of $\sigma^2 = 1.070$, thus the individual participants showed a variance of $\sigma^2 = 0.177$. See model 3 in appendix B for further details.

4.2 Descriptive statistics

State of participant

The mean degree of alertness among the participants was $\mu = 3.720$. The mean degree of calmness was $\mu = 4$. The mean mood was $\mu = 3.853$. See descriptive statistics in appendix B for further details.

Frequencies for states of participants			
Level	Mood	Calmness	Alertness
5 - High	18	30	21
4	33	22	27
3	19	17	15
2	5	5	9
1 - Low	0	1	3

Table 1 Modes for states of participants.

Overall memorized images

The overall mean of images memorized, regardless of congruence, per participant and per collage was $\mu = 3.242$. The mean number of images memorized per affect quadrant, per participant and per collage, was as follows: (Arousal = 1, Valence = 1; $\mu = 1.107$); (Arousal = 0, Valence = 1; $\mu = 0.7$); (Arousal = 0, Valence = 0; $\mu = 0.552$); and (Arousal = 1, Valence = 0; $\mu = 0.882$). See descriptive statistics in appendix B for further details.

Musicality of participant

The musicality of participants did not have a significant effect on memory performance ($p > .995$). See model 5 in appendix B for further details.

Influence of Big Five Inventory parameters on memory performance

Openness had a significant influence on memory performance ($p < .002$), with the highest estimate (estimate = 0.170) (odds ratio = 1.185). The second highest influence was agreeableness ($p < .005$) (estimate = -0.203) (odds ratio = 0.816) in comparison to the other BFI parameters. The influence of the remaining BFI parameters was not significant: Extraversion ($p > .743$) (estimate = 0.017) (odds ratio = 1.017), conscientiousness ($p > .143$) (estimate = -0.088) (odds ratio = 0.915), emotional stability ($p > .792$) (estimate = -0.013) (odds ratio = 0.987). The effect size for this model was given by $R^2_{glmm, (c)} = 0.180$. See model 4 in appendix B for further details.

Further data analysis indicated that participants who scored higher on BFI openness score (openness > 3), memorized more images which were congruent to the auditory stimulus ($\mu = 7.826$ images of 48 possible congruent pairings in total) than participants scoring low. Participants scoring low on the BFI openness score (openness ≤ 3) memorized on average fewer images which were congruent to the auditory stimulus ($\mu = 1.74$ images of 48 possible congruent pairings in total). See influence of BFI data in appendix B for further details.

Frequencies for the BFI data of participants					
Level	Extra-version	Conscientiousness	Emotional instability	Open-ness	Agreeable-ness
5 - High	14	16	6	35	5
4	28	38	16	23	25
3	22	18	18	12	39
2	10	3	30	5	6
1 - Low	1	0	2	0	0

Table 2 Frequencies for the BFI data of participants.

5 Discussion

Based on the findings of recent research, it was hypothesized that in the experiment used in this study, visual attention will be guided towards images that are affectively congruent to the concomitant auditory stimulus.

The results show that congruence between the visual and auditory stimuli had a positive effect on memory performance. Even though the effect size was relatively small, H1 was verified. It was expected that effect sizes would be small due to the highly contextualized experiment and non-laboratory conditions. The experiment was conducted online, with the possibility of high variability in the surrounding circumstances for each participant. Little control was available over the participant taking part in the experiment and conditions were thus expected to strongly influence the results. Despite this high variability in conditions, significant results were noted nonetheless.

To further understand the detailed influence of multidimensional stimulus congruence on memory performance, the full congruence was broken down into two fixed effects, namely arousal congruence and valence congruence. The GLMM revealed that memory performance was more influenced by valence congruence than by arousal congruence. H2 was therefore verified.

Considering the concepts of emotion induction and emotion perception, the author expects that effect sizes for H1 and H2 would be higher for experimental conditions allowing reliable and controlled emotion induction in comparison to the experimental conditions in this study, using emotion perception. Based on the reviewed literature, it was assumed that neural and cognitive processes underlying the congruence effects are as well sufficiently triggered in situations of emotion perception. This assumption was derived from the results, indicating a correspondence between reviewed literature and the observed effects, as expected in the hypotheses.

The experimental situation can be seen as beneficial for the validity of the results. Despite the high variability of conducting an experiment in diverse situations of the participants, the results approximate everyday-life conditions. Cognitive states of participants before the experiment were not actively influenced by the researcher by inviting people to a laboratory session. This allowed the participant to start the experiment as unbiased as possible. Considering these uncontrolled experimental conditions it is interesting to still find a significant influence of congruence on memory. However the author expects that a more controlled experimental setting would likely render larger effect sizes.

Although initial mood was not experimentally influenced, the average mood was rated as neutral with a tendency towards the positive, whilst the mode of the ratings was 4. This is in accordance with findings by Gasper and Clore (2002), who stated that the average affective state is neutral with tendency towards positive. Therefore, initial attentional focus was balanced between globally versus locally oriented. These results indicate that on average, no extreme mood conditions prior to conducting the experiment occurred that might have biased the findings or especially participants memory performance.

Verifying H3, the analysed data show that the arousal of an image has a positive effect on memory performance. This effect was independent of the concomitant auditory stimulus. This is in accordance with earlier findings of Bradley et al. (1992). For example it can be seen, that the four images which were best memorized all have a high arousal rating and are balanced between low valence and high valence. The four least memorized images, except one image, show low arousal and are balanced between high valence and low valence. See list of the image variances in model 1 in appendix B of the images best memorized and the image files in the attached data storage in appendix F.

In this study, memory performance for images was used as an indication of which images were in focus of visual attention. The results indicate that memory and visual attention operate by similar mechanisms. However, highly arousing images might have biased this model by their memorability, regardless of the affective influence of a secondary stimulus.

Furthermore, the findings indicate that recent scientific work on models of attentional guidance appear valid in the context of bi-modal displays. Not only have prior models been validated, but can as well be extended to influences of auditory stimuli, at least in this context. This area of research focusing on human perception is complex and results tend to be highly contextualized, because many factors influence perception. “Contextualized” means that the nature of human perception is subject to many influences, which can have different impacts on perception depending on situations, cognitive states and experimental conditions. Therefore, this research only adds a little insight to understanding perceptual and attentional processes. As initially stated, a complete and unified model of human perception and intersensory influences in which context does not play a role is unlikely to be developed.

This research could also have been conducted using psycho-acoustic measures, such as roughness, loudness and spectral content. Finding physiological similarities could possibly better explain the similarity of neural codes between auditory and visual perception, which might underlie many of the congruence effects and cross-modal correspondences. However, this research used auditory stimuli in the form of music, and image stimuli in the form of photographs, to convey affective properties. This allowed for direct comparison using an affective scale, and close connections to observable cognitive processes. Furthermore, it is possible to embed this research in the context of affective sciences and psychological research on music.

Another benefit of the affective approach used in this study was that the stimuli might better facilitate semantic networks. This would enable the observation of congruence effects, considering the experimental setup. Semantic priming by auditory stimuli is assumed to be a driving factor in the congruence effects, due to the high influence of valence congruence and the moderate influence of arousal congruence. Multi-modal integration might enhance the activation of semantic networks in semantic memory, primarily guiding visual attention towards semantic aspects and in return strengthening memory for semantically congruent objects.

Concerning personality aspects, two factors from the BFI showed an influence on memory performance. Although the effect size was small, openness and agreeableness had significant influence on memory performance. Further data analysis indicated that Participants scoring high on the openness factor of the BFI on average memorized more fully congruent stimulus pairings from the pairings available than people who scored low on BFI openness. As the BFI factor of openness is related to curiosity and excitableness, possibly a higher susceptibility towards emotion perception and induction could result in greater attention being given to congruent objects by participants scoring high on openness; this could also result in better memory performance. This strong bi-modal receptiveness might be further enhanced by readily available semantic networks among people who scored high on BFI openness. Therefore, attentional guidance based on affective-feature congruence is assumed to be a more active mechanism in participants scoring high on openness than those who score low in openness. However, this might be the reason for openness in the first place. Above average active neural connections could underlie a heightened receptiveness for bi-modal congruent stimuli, leading to an increased openness and processing of sensory input. Enhanced neural representation of sensory input and more readily available networks in semantic memory might influence personality traits, such as openness, as openness is associated with fantasy and excitability.

To better interpret the findings on influences of the BFI openness parameter on memory performance, further literature was reviewed. Another interesting connection can be made to *sensory processing sensitivity*, which is defined as a deeper and more intense perception and processing of sensory input (Ahadi & Basharpoor, 2010). Although no significant relationship between ease of excitation and openness was found by Ahadi and Basharpoor (2010), a significant positive relationship between aesthetic sensitivity and openness was found. Ahadi and Basharpoor (2010) thus argue, that people scoring high on aesthetic sensitivity have rich inner experiences due to sensory input. In return, they suggest, that this leads to an increased level of openness. The argument that personality arises from manifested processing styles for sensory input (Ahadi & Basharpoor, 2010), could explain the significant susceptibility to congruent audio-visual stimuli by participants scoring high on BFI openness.

Using memory performance as an indicator for attentional focus is susceptible to bias. Arousal level of images having an effect on memory performance, regardless of the affective features of a secondary auditory stimulus, could thus affect the results of an experiment. Differences exist in memory-object processing with regard to arousal and valence levels, with valence performing best for short-term memory. Hence, accurate conclusions about auditory influences on attentional focus in regard to high arousing image stimuli, cannot be made in this study. The use of direct technologies such as eye-tracking devices could limit such bias, but come at the cost of technical effort and are restricted to laboratory sessions. Hence, for this research the choice had to be made between the control of a laboratory situation or a larger sample with uncontrolled settings.

6 Conclusion

This study investigated the influences of a secondary -auditory- stimulus on visual attention guidance. Aim was to find out how affective-feature congruence, in terms of valence and arousal, between both stimuli leads visual attention. In conclusion, in the context of music and image perception it appears that affective congruence (between the images and music) facilitated visual attention towards congruent features. Valence congruence seems to be the driving factor in visual attention guidance, whereas arousal of visual objects could bias such guidance. Although this research portrayed a specialized context using music and photographs, the author assumes that the proposed model of visual attention guidance to hold true in a general non-musical context, at least to some extent. Further research could facilitate non-musical auditory stimuli paired with visual environmental surroundings; this would allow testing the validity of the attention-guidance model. Future research extending this investigation could also use eye-tracking to avoid biases through memory tasks. Additionally, high-arousal baseline conditions could be created to further explore the mechanism when arousing stimuli are presented. Future studies could also include more detailed questionnaires about personality and sensory processing sensitivity, to better account for individual differences in sensory processing.

References

Ahadi, B., & Basharpoor, S. (2010). Relationship between sensory processing sensitivity, personality dimensions and mental health. *Journal of Applied Sciences*, *10*(7), 570–574.

Amezcua, C., Guevara, M. A., & Ramos-Loyo, J. (2005). Effects of musical tempi on visual attention ERPs. *The International Journal of Neuroscience*, *115*(2), 193–206. https://doi.org/10.1080/00207450590519094

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes1. In *Psychology of learning and motivation* (pp. 89–195). Elsevier.

Baddeley, A. (2010). Working memory. *Current Biology : CB*, *20*(4), R136-40. https://doi.org/10.1016/j.cub.2009.12.014

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software.* Advance online publication. https://doi.org/10.18637/jss.v067.i01

Bishop, D. T., Karageorghis, C. I., & Kinrade, N. P. (2009). Effects of Musically-Induced Emotions on Choice Reaction Time Performance. *The Sport Psychologist*, *23*(1), 59–76. https://doi.org/10.1123/tsp.23.1.59

Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception: an Interdisciplinary Journal*, *18*(4), 427–454.

Boltz, M. G., Ebendorf, B., & Field, B. (2009). Audiovisual Interactions: the Impact of Visual Information on Music Perception and Memory. *Music Perception*, *27*(1), 43–59. https://doi.org/10.1525/mp.2009.27.1.43

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 379–390. https://doi.org/10.1037//0278-7393.18.2.379

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review*, *64*(3), 205.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, *93*(24), 13494–13499. https://doi.org/10.1073/pnas.93.24.13494

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1319–1331. https://doi.org/10.1098/rstb.1998.0286

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*(3), 183.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49. https://doi.org/10.1177/0305735610362821

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6.1-12. https://doi.org/10.1167/10.1.6

Frijda, N. H. (1986). *The emotions*: Cambridge University Press.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different? *Musicae Scienta*, *2001-2002*, 123–147.

Gasper, K., & Clore, G. L. (2002). Attending to the Big Picture: Mood and Global Versus Local Processing of Visual Information. *Psychological Science*, *13*(1), 34–40. https://doi.org/10.1111/1467-9280.00406

Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research*, *44*(5), 501–521. https://doi.org/10.1016/j.visres.2003.09.033

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of Musical Tempo and Mode on Arousal, Mood, and Spatial Abilities. *Music Perception*, *20*(2), 151–171. https://doi.org/10.1525/mp.2002.20.2.151

Irtel, H., & Goldstein, E. B. (Eds.). (2011). *Wahrnehmungspsychologie: Der Grundkurs* (G. Plata, Trans.) (7. Aufl., Nachdr). Berlin: Spektrum Akad. Verl.

Iwamiya, S.-i. (1997). Interaction between Auditory and Visual Processing in Car Audio: Simulation Experiment using Video Reproduction. *APPLIED HUMAN SCIENCE Journal of Physiological Anthropology*, *16*(3), 115–119. https://doi.org/10.2114/jpa.16.115

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*(4), 575–591. https://doi.org/10.1016/j.jml.2006.05.004

Kitayama, S., & Niedenthal, P. M. (1994). *The Heart's Eye: Emotional influences in Perception and Attention*: Academic Press.

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, *49*(2), 457–470. https://doi.org/10.3758/s13428-016-0715-3

Lane, R., Chua, P., & Dolan, R. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*. (37), 989–997.

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, *24*(2), 326–334. https://doi.org/10.1016/j.cogbrainres.2005.02.005

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, *16*(3), 468–478. https://doi.org/10.1016/S0926-6410(03)00074-0

Mather, M., & Sutherland, M. (2009). Disentangling the Effects of Arousal and Valence on Memory for Intrinsic Details. *Emotion Review*, *1*(2), 118–119. https://doi.org/10.1177/1754073908100435

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural Substrates of Perceptual Enhancement by Cross-Modal Spatial Attention. *Journal of Cognitive Neuroscience*, *15*(1), 10–19. https://doi.org/10.1162/089892903321107783

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133–136. https://doi.org/10.3758/CABN.4.2.133

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Proverbio, A. M., Mado Proverbio, C. A. A., Lozano Nasi, V., Alessandra Arcari, L., Benedetto, F. de, Guardamagna, M., . . . Zani, A. (2015). The effect of background music on episodic memory and autonomic responses: listening to emotionally touching music enhances facial memory capacity. *Scientific Reports*, *5*, 15219. https://doi.org/10.1038/srep15219

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2012). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI-10).

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10). Advance online publication. https://doi.org/10.12758/MDA.2013.013

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Scherer, K. R. (2004). Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them? *Journal of New Music Research*, *33*(3), 239–251. https://doi.org/10.1080/0929821042000317822

Schermer, F. J. (1998). *Lernen und Gedächtnis* (2., überarb. u. erw. Aufl.). *Grundriss der Psychologie: Vol. 10*. Stuttgart, Berlin, Köln: Kohlhammer.

Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, *14*(4), 325–345.

Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental Psychology*, *55*(2), 121–132. https://doi.org/10.1027/1618-3169.55.2.121

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*(3), 269–284. https://doi.org/10.1016/j.plrev.2010.04.006

Simon, H. A. (1974). How Big Is a Chunk?: By combining data from several experiments, a basic human memory unit can be identified and measured. *Science (New York, N.Y.)*, *183*(4124), 482–488. https://doi.org/10.1126/science.183.4124.482

Steffens, J., Steele, D., & Guastavino, C. (2016). Music influences the perception of our acoustic and visual environment, *1*.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*, *1*, 381–403.

Wadlinger, H. A., & Isaacowitz, D. M. (2006). Positive mood broadens visual attention to positive stimuli. *Motivation and Emotion*, *30*(1), 89–101. https://doi.org/10.1007/s11031-006-9021-1

Welch, R. B., & Warren, D. H. (1986). *Intersensory Interactions: Handbook of perception and human performance*. *A Wiley-Interscience publication: Vol. 1*. New York: Wiley.

Yamasaki, T., Yamada, K., & Laukka, P. (2015). Viewing the world through the prism of music: Effects of music on perceptions of the environment. *Psychology of Music*, *43*(1), 61–74. https://doi.org/10.1177/0305735613493954

Appendix A - Screenshots Of Stimulus Collage

For the complete set of screenshots please refer to the data storage attached.

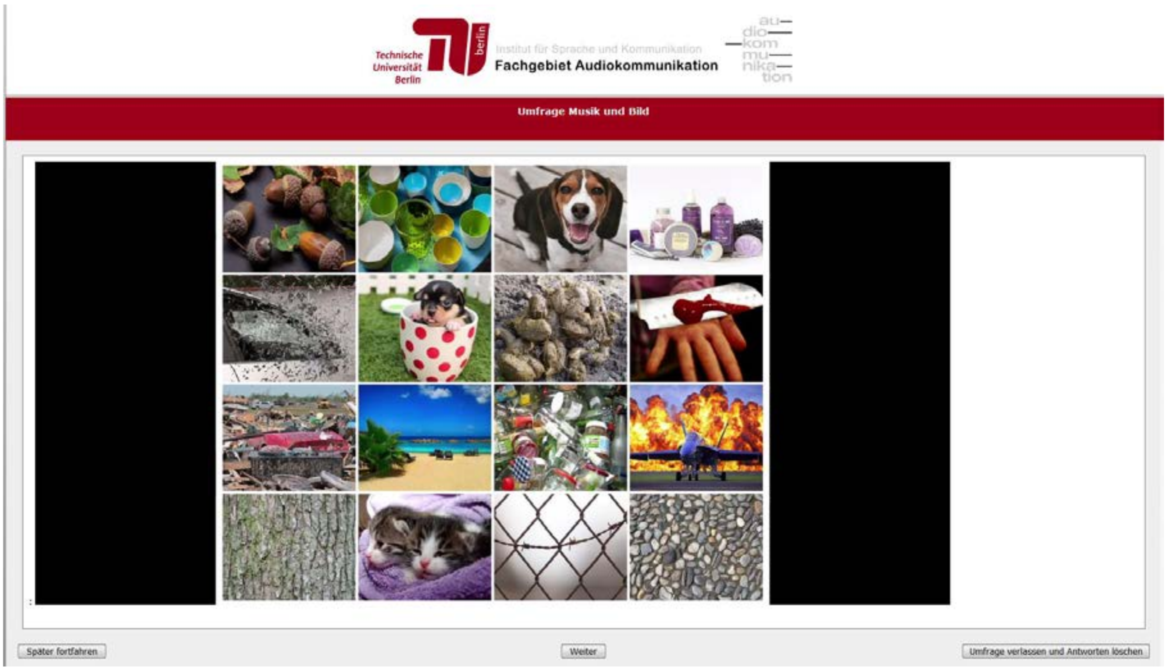

Figure 4 Screenshot of an image collage embedded into the LimeSurvey platform.

Figure 5 Open ended question. Participants were asked to write down the images they memorized. This question followed each audiovisual stimulus presentation.

Appendix B - Statistical Data Output

R script with all statistical calculations and outputs of calculations.

```
> library("lme4", lib.loc="~/R/win-library/3.5")
Lade nötiges Paket: Matrix
> library("haven", lib.loc="~/R/win-library/3.5")
> library("rcompanion", lib.loc="~/R/win-library/3.5")
Error: package or namespace load failed for 'rcompanion' in loadNamespace(j <- i[[1L]], c(lib.loc,
.libPaths()), versionCheck = vI[[j]]):
 es gibt kein Paket namens 'multcomp'
> library("DescTools", lib.loc="~/R/win-library/3.5")
> library("MuMIn", lib.loc="~/R/win-library/3.5")
>
> d <- read_sav("Umfragedaten_David_Du_Bruyn.sav")              #Einlesen des Datensatzes
>
>
>
##########################################################################################
##################################
>
> # model1: Erstes Modell GLMER. Hier Vollkongruenz als Fixed Effect fuer These H1.
>
> model1<- glmer(Gemerkt_gesamt ~ Kongruenz_voll_gesamt +
+                 (1|Teilnehmer_id) + (1|Image_ID_gesamt_einheitlich), data=d,family='binomial')
>
> # Output: Auflistung der Ergebnisse des GLMER(model1)
> summary(model1)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Gemerkt_gesamt ~ Kongruenz_voll_gesamt + (1 | Teilnehmer_id)
 +   (1 |Image_ID_gesamt_einheitlich)
   Data: d

    AIC      BIC       logLik     deviance    df.resid
 12815.8  12846.1      -6403.9     12807.8     14396

Scaled residuals:
    Min         1Q       Median     3Q          Max
-2.0783     -0.4979     -0.3290  -0.1832        8.0752

Random effects:
 Groups                          Name               Variance              Std.Dev.
 Image_ID_gesamt_einheitlich    (Intercept)         1.1879                1.0899
 Teilnehmer_id                  (Intercept)          0.1776                0.4214
Number of obs: 14400, groups:  Image_ID_gesamt_einheitlich, 190; Teilnehmer_id, 75

Fixed effects:
                              Estimate   Std. Error            z value     Pr(>|z|)
(Intercept)                   -1.74980    0.09902              -17.671     <2e-16 ***
Kongruenz_voll_gesamt         0.15188     0.07522              2.019       0.0435 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr)
Kngrnz_vll_ -0.199
> # Effektstaerke Ueber das R²
> r.squaredGLMM(model1)
                R2m                   R2c
theoretical  0.0009275595          0.2939787
delta        0.0005746206          0.1821190
>
```

> #Output: Zeigt hier welches Bild besonders gut gemerkt wurde. (vgl. ID mit Stimulus Datenbank auf Datenträger)
> ranef(model1)
$Image_ID_gesamt_einheitlich
(Intercept)

1 0.28579019	65 1.78833905	129 0.36830058
2 -0.73975359	66 -1.19460949	130 1.02782109
3 -0.27529862	67 -0.06427610	131 1.94704666
4 1.24578497	68 0.36830058	132 0.20792799
5 -1.48297155	69 -0.14576023	133 0.96887492
6 2.16286339	70 -0.35593465	134 -0.66136939
7 -1.26812016	71 0.57165357	135 -0.52089580
8 -0.89459584	72 -0.05029051	136 0.28986613
9 -0.58033988	73 1.05631012	137 -0.20358764
10 1.84107620	74 0.88541624	138 0.06673661
11 2.62798402	75 0.22819744	139 -0.20358764
12 1.02221021	76 0.82660426	140 -1.12130939
13 -1.77996703	77 -0.81741065	141 0.44364285
14 -0.27340687	78 -0.16618614	142 0.65439464
15 -0.14576023	79 -1.19460949	143 -1.71346738
16 -0.47301024	80 -0.06427610	144 0.20792799
17 -0.52089580	81 -0.71845772	145 1.07920274
18 0.12199947	82 0.96622402	146 2.38266221
19 0.38096636	83 2.04650373	147 1.19090005
20 0.72043616	84 -0.87357409	148 -0.87513249
21 -0.87513249	85 -0.81741065	149 -0.14576023
22 0.78473442	86 -0.06427610	150 -1.50094304
23 -0.45302138	87 -1.19460949	151 1.33577172
24 0.72043616	88 -0.39295943	152 0.28579019
25 1.02221021	89 -0.39295943	153 0.28579019
26 -0.06427610	90 -1.43032891	154 -0.73975359
27 -0.35040789	91 -1.19460949	155 -1.50094304
28 -0.99318811	92 0.72043616	156 -1.77996703
29 -0.27529862	94 -1.43032891	157 -1.19460949
30 0.20792799	95 -1.71346738	158 -0.06427610
31 0.58640728	96 0.51624424	159 0.84746923
32 0.36830058	97 0.20792799	161 1.08575886
33 -1.06880276	98 0.36830058	162 -0.39295943
34 -0.47301024	99 -0.52089580	163 1.02782109
35 1.22726762	100 0.21447399	164 -0.66136939
36 1.49585980	101 -0.06427610	165 -0.52089580
37 -1.71346738	102 -0.06427610	166 0.36830058
38 1.36387159	103 0.78473442	167 0.03149905
39 0.12199947	104 0.20792799	168 0.20792799
40 0.20792799	105 1.41779528	169 1.52456153
41 1.41779528	106 0.84746923	170 1.30948990
42 0.20792799	107 -0.52089580	171 1.19903751
43 -0.39295943	108 -1.19460949	172 -0.99318811
44 2.10824766	109 -1.31904039	173 -1.48297155
45 -1.82588456	110 2.01800670	174 -0.87513249
46 -1.12130939	111 -1.54975936	175 0.65729563
47 -2.16673919	112 -0.20358764	176 -0.45302138
48 -0.52926291	113 0.70198975	177 -0.27529862
49 1.49585980	114 0.70198975	178 -0.27529862
50 1.00321780	115 0.12569541	179 0.65439464
51 -0.73975359	116 0.43357446	180 0.58640728
52 0.28579019	117 -0.87513249	181 -0.72005071
53 0.65729563	118 -1.48297155	182 -0.72005071

54 1.02221021 55 -0.58033988 56 1.99033501 57 0.03149905 58 -0.99318811 59 0.12199947 60 1.84107620 61 1.19717547 62 -0.66282704 63 -0.99456988 64 -0.81883347	119 -0.03015925 120 0.45343112 121 0.84746923 122 1.14279743 123 0.72043616 124 0.44364285 125 -1.71346738 126 -1.43032891 127 -0.06427610 128 -0.06427610	183 -2.10947736 184 0.52347255 185 0.96887492 186 -1.29783369 187 -0.81741065 188 1.57754350 189 0.96887492 190 0.65439464 191 -0.06427610 192 0.58640728
with conditional variances for "Image_ID_gesamt_einheitlich" "		

>
##
###################################
>
> # **model2**: Zweites Modell GLMER. Hier Arousal_Kongruenz und Valenz_Kongruenz als fixed effects. Fuer These H2.
>
> model2<- glmer(Gemerkt_gesamt ~ Arousal_Kongruenz + Valenz_Kongruenz +
\+ (1|Teilnehmer_id) + (1|Image_ID_gesamt_einheitlich), data=d,family='binomial')
>
> # Output: Auflistung der Ergebnisse des GLMER(model2)
> summary(model2)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: binomial (logit)
Formula: Gemerkt_gesamt ~ Arousal_Kongruenz + Valenz_Kongruenz + (1 |
Teilnehmer_id) + (1 | Image_ID_gesamt_einheitlich)
Data: d

AIC	BIC	logLik	deviance	df resid
12809.5	12847.4	-6399.7	12799.5	14395

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.1292	-0.4983	-0.3284	-0.1839	7.9975

Random effects:

Groups	Name	Variance	Std.Dev.
Image_ID_gesamt_einheitlich	(Intercept)	1.1423	1.0688
Teilnehmer_id	(Intercept)	0.1774	0.4211

Number of obs: 14400, groups: Image_ID_gesamt_einheitlich, 190; Teilnehmer_id, 75

model 2 continued

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.79945	0.10040	-17.923	< 2e-16 ***
Arousal_Kongruenz	0.05481	0.07933	0.691	0.489659
Valenz_Kongruenz	0.29430	0.08418	3.496	0.000472 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Correlation of Fixed Effects:
            (Intr) Arsl_K
Arsl_Kngrnz -0.204
Vlnz_Kngrnz -0.225  0.022
>
> # Effektstaerke ueber das R²
> r.squaredGLMM(model2)
                 R2m                 R2c
theoretical  0.003632301             0.2888778
delta        0.002244032             0.1784684
Warning message:
The null model is correct only if all variables used by the original model remain unchanged.
>
>########################################################################################
###################################
>
> # moodel3: Drittes Modell GLMER. Hier Arousal_image_gesamt als fixed effects fuer These H3.
>
> model3<- glmer(Gemerkt_gesamt ~ Arousal_image_gesamt +
+                  (1|Teilnehmer_id) + (1|Image_ID_gesamt_einheitlich), data=d,family='binomial')
>
> # Output: Auflistung der Ergebnisse des GLMER(model3)
> summary(model3)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Gemerkt_gesamt ~ Arousal_image_gesamt + (1 | Teilnehmer_id) +      (1 |
Image_ID_gesamt_einheitlich)
   Data: d

     AIC       BIC       logLik      deviance   df resid
 12804.5    12834.8     -6398.2     12796.5    14396

Scaled residuals:
    Min      1Q       Median          3Q        Max
-2.1518 -0.4984      -0.3295        -0.1827    7.6255

Random effects:
 Groups                          Name                  Variance               Std.Dev.
 Image_ID_gesamt_einheitlich     (Intercept)           1.0707                  1.035
 Teilnehmer_id                   (Intercept)           0.1772                 0.421
Number of obs: 14400, groups:  Image_ID_gesamt_einheitlich, 190; Teilnehmer_id, 75

Fixed effects:
                                 Estimate  Std. Error  z value    Pr(>|z|)
(Intercept)                      -2.0313    0.1246   -16.300     < 2e-16 ***
Arousal_image_gesamt              0.6368    0.1591    4.002       6.29e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr)
Arsl_mg_gsm -0.659
>
> # Effektstaerke ueber das R²
> r.squaredGLMM(model3)
                R2m                 R2c
theoretical  0.02185747             0.2908488
delta        0.01351784             0.1798767
>
```

```
>
#########################################################################################
> # model4: Einfluss der BFI-Daten
>
> model4<- glmer(Gemerkt_gesamt ~  BFI_Extraversion + BFI_Offenheit + BFI_Gewissenhaftigkeit +
BFI_Verträglichkeit + BFI_Neurotizismus
+              +                  (1|Teilnehmer_id) + (1|Image_ID_gesamt_einheitlich), data=d,family='binomial')
>
> # Output: Auflistung der Ergebnisse des GLMER(model4)
> summary(model4)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Gemerkt_gesamt ~ BFI_Extraversion + BFI_Offenheit + BFI_Gewissenhaftigkeit +
    BFI_Verträglichkeit + BFI_Neurotizismus + (1 | Teilnehmer_id) +      (1 | Image_ID_gesamt_einheitlich)
   Data: d

    AIC      BIC           logLik      deviance    df resid
 12810.9  12871.5        -6397.5     12794.9    14392

Scaled residuals:
    Min        1Q          Median       3Q          Max
-2.1232     -0.4975     -0.3282     -0.1840      7.9594

Random effects:
 Groups                     Name        Variance Std.Dev.
 Image_ID_gesamt_einheitlich (Intercept) 1.174    1.0835
 Teilnehmer_id              (Intercept) 0.133    0.3647
Number of obs: 14400, groups:  Image_ID_gesamt_einheitlich, 190; Teilnehmer_id, 75

Fixed effects:
                           Estimate        Std. Error       z value        Pr(>|z|)
(Intercept)                 -1.43326        0.50328          -2.848         0.00440 **
BFI_Extraversion           0.01770         0.05417           0.327         0.74392
BFI_Offenheit               0.17026         0.05402          3.152          0.00162 **
BFI_Gewissenhaftigkeit -0.08833            0.06044          -1.461         0.14389
BFI_Verträglichkeit     -0.20356            0.07223          -2.818         0.00483 **
BFI_Neurotizismus       -0.01366            0.05193           -0.263         0.79246
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
                       (Intr)      BFI_Ex      BFI_Of      BFI_Gw    BFI_Vr
BFI_Extrvrs           -0.365
BFI_Offenht           -0.411      -0.035
BFI_Gwssnhf           -0.568      -0.030      -0.070
BFI_Vrtrglc           -0.570      -0.081      0.028       0.235
BFI_Nrtzsms           -0.526      0.249       0.043       0.220       0.095
Warning messages:
>
> # Output: Effektstaerke Ueber das R²
> r.squaredGLMM(model4)
              R2m                     R2c
theoretical  0.009564048             0.2911772
delta        0.005915970             0.1801116
>
#####################
# Einfluss der BFI Daten 2
>
> #Wie viel Kongruente Paarungen wurden gemerkt bei hohem bzw. niedrigem BFI Offenheits Score
>
```

```
> table(d$Offenheit_hoch_kongruenz_gemerkt) #Offenheit_hoch

    0     1 
13813   587 
> table(d$Offenheit_niedrig_kongruenz_gemerkt)  #Offenheit_niedrig

    0     1 
14269   131 
> 587/75                # 7.82 mean gemerkte kongruente Paarungen pro Teilnehmer Offenheit hoch
[1] 7.826667
> 131/75                # 1.74 mean gemerkte kongruente Paarungen pro Teilnehmer Offenheit niedrig
[1] 1.746667
> 
####################################################################################################################################
> #### Deskriptive Statistiken ####
> 
> # Zustand des Teilnehmers
> 
> summary(d$Aktueller_Wachheitsgrad)                  # Aktueller Wachheitsgrad (1-müde bis 5-wach)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   1.00    3.00    4.00    3.72    5.00    5.00 
> summary(d$Aktueller_Ruhegrad)                       # Aktueller Ruhegrad der Teilnehmer (1-unruhig
                                                        bis 5-ruhig)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      1       3       4       4       5       5 
> summary(d$Aktuelle_Stimmung)                        # Aktuelle Stimmung der Teilnehmer (1-schlechte
Stimmung bis 5-gute Stimmung)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  2.000   3.000   4.000   3.853   4.000   5.000 
> 
> # Durchschnittlich pro Trial gemerkte Bilder
> 
> table(d$Gemerkt_gesamt)  # divide ans by 75 participants -> divide by 12 trials => 3.242223 images
memorized

    0     1 
11482  2918             # per trial
> > 
> ##########
> 
> # Durschnittlich gemerkte Bilder pro Affektquadrant
> 
> # divide each answer by 75 participants -> divide by 12 trials
> table(d$Gemerkt_Val_Arous_1)              #Valence=1 Arousal=1

    0     1 
13403   997 
> table(d$Gemerkt_Val_Arous_0)              #Valence=0 Arousal=0

    0     1 
13903   497 
> table(d$Gemerkt_Val_1_Arous_0)            #Valence=1 Arousal=0

    0     1 
13770   630 
> table(d$Gemerkt_Val_0_Arous_1)            #Valence=0 Arousal=1

    0     1 
13606   794 
```

```
> > ##########
>
> # model 5 Einfluss der Musikalität auf die Gedächtnisleistung
>
> model5<- glmer(Gemerkt_gesamt ~  Mus_J_N +
+                   +(1|Teilnehmer_id) + (1|Image_ID_gesamt_einheitlich), data=d,family='binomial')
>
> # Output: Auflistung der Ergebnisse des GLMER(model5)
> summary(model5)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Gemerkt_gesamt ~ Mus_J_N + +(1 | Teilnehmer_id) + (1 | Image_ID_gesamt_einheitlich)
   Data: d

    AIC      BIC           logLik      deviance    df resid
 12819.8  12850.1       -6405.9      12811.8    14396

Scaled residuals:
   Min        1Q          Median      3Q            Max
-2.1456     -0.4987     -0.3290      -0.1847  8.0153

Random effects:
 Groups                  Name        Variance Std.Dev.
 Image_ID_gesamt_einheitlich (Intercept) 1.1748   1.084
 Teilnehmer_id              (Intercept) 0.1773   0.421
Number of obs: 14400, groups:  Image_ID_gesamt_einheitlich, 190; Teilnehmer_id, 75

Fixed effects:
              Estimate                  Std. Error                 z value      Pr(>|z|)
(Intercept) -1.7109703                  0.1089517                 -15.704      <2e-16 ***
Mus_J_N      0.0005873                 0.1076162                  0.005        0.996
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr)
Mus_J_N -0.462
>
> # Output: Effektstaerke Ueber das R²
> r.squaredGLMM(model5)
                R2m                     R2c
theoretical  1.849717e-08           0.2912679
delta           1.144223e-08           0.1801765
##########################################################################################
############

# Deskriptive Statistik BFI Daten
>
> mean(d$BFI_Extraversion)
[1] 3.36
> mean(d$BFI_Gewissenhaftigkeit)
[1] 3.666667
> mean(d$BFI_Offenheit)
[1] 3.973333
> mean(d$BFI_Verträglichkeit)
[1] 3.206667
> mean(d$BFI_Neurotizismus)
[1] 2.693333
```

Appendix C - Music titles used for online experiment

Music titles used for online experiment. A 10 second sample was taken out of the timeframe stated. (cf. Eerola & Vuoskoski, 2011)

Emotion & level	**Album name**	**Track**	**Min:Sec**
Fear High	Batman Returns	5	00:09–00:25
Happy High	Batman	18	00:55–01:15
Sad High	The English Patient	18	00:07–00:32
Tender High	Shine	10	01:28–01:48
Fear High	JFK	8	01:26–01:40
Happy High	Man Of Galilee	2	03:02–03:18
Sad High	Portrait Of A Lady	9	00:00–00:22
Tender High	Pride And Prejudice	1	00:10–00:26
Fear High	Alien Trilogy	5	00:26–00:41
Happy High	Oliver Twist	1	00:17–00:34
Sad High	Big Fish	15	00:55–01:11
Tender High	Dance With Wolves	4	01:31–01:48

Appendix D - BFI Questionnaire

The Big Five Inventory questionnaire. Taken from Rammstedt, Kemper, Klein, Beierlein, and Kovaleva (2012)

BFI-10

Inwieweit treffen die folgenden Aussagen auf Sie zu?

	trifft überhaupt nicht zu	trifft eher nicht zu	weder noch	eher zutreffend	trifft voll und ganz zu
(1) Ich bin eher zurückhaltend, reserviert.	□1	□2	□3	□4	□5
(2) Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	□1	□2	□3	□4	□5
(3) Ich bin bequem, neige zur Faulheit.	□1	□2	□3	□4	□5
(4) Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	□1	□2	□3	□4	□5
(5) Ich habe nur wenig künstlerisches Interesse.	□1	□2	□3	□4	□5
(6) Ich gehe aus mir heraus, bin gesellig.	□1	□2	□3	□4	□5
(7) Ich neige dazu, andere zu kritisieren.	□1	□2	□3	□4	□5
(8) Ich erledige Aufgaben gründlich.	□1	□2	□3	□4	□5
(9) Ich werde leicht nervös und unsicher.	□1	□2	□3	□4	□5
(10) Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	□1	□2	□3	□4	□5

Appendix E - Auditory And Visual Stimuli

This table shows the all the visual and auditory stimuli used in this experiment with valence and arousal values translated to a binary value for this experiment. The stimuli in this table simultaneously present the contents of each collage and the final stimulus pairings for experiment group 1. See the attached data storage (appendix F) for the complete table for both groups. See appendix F (attached data storage) for stimuli files. Note the same stimuli have been used for group 2, but have been mixed (refer to section Materials for details).

ID	Array	Name Bild	Valence	Arousal	Name Audio
1	1	Acorns 3	1	0	011_FearHigh_BatmanReturns_5
2	1	Bark 3	1	0	011_FearHigh_BatmanReturns_5
3	1	Beach 1	1	1	011_FearHigh_BatmanReturns_5
4	1	Bloody knife 2	0	1	011_FearHigh_BatmanReturns_5
5	1	Car Accident 3	0	1	011_FearHigh_BatmanReturns_5
6	1	Cat 5	1	1	011_FearHigh_BatmanReturns_5
7	1	Clean 1	1	0	011_FearHigh_BatmanReturns_5
8	1	Cups 3	1	0	011_FearHigh_BatmanReturns_5
9	1	Destruction 8	0	1	011_FearHigh_BatmanReturns_5
10	1	Dog 6	1	1	011_FearHigh_BatmanReturns_5
11	1	Dog 12	1	1	011_FearHigh_BatmanReturns_5
12	1	Explosion 1	0	1	011_FearHigh_BatmanReturns_5
13	1	Feces 1	0	0	011_FearHigh_BatmanReturns_5
14	1	Fence 4	0	0	011_FearHigh_BatmanReturns_5
15	1	Garbage dump 7	0	0	011_FearHigh_BatmanReturns_5
16	1	Sidewalk 6	0	0	011_FearHigh_BatmanReturns_5
17	2	Fence 2	1	0	022_HappyHigh_Batman_18
18	2	Fire 9	0	1	022_HappyHigh_Batman_18
19	2	Fireworks 2	1	1	022_HappyHigh_Batman_18
20	2	Flowers 6	1	0	022_HappyHigh_Batman_18
21	2	Galaxy 7	1	1	022_HappyHigh_Batman_18
22	2	Grass 4	1	0	022_HappyHigh_Batman_18
23	2	Lake 12	1	1	022_HappyHigh_Batman_18
24	2	Pinecone 3	1	0	022_HappyHigh_Batman_18
25	2	Rollercoaster 2	1	1	022_HappyHigh_Batman_18
26	2	Snow 4	0	0	022_HappyHigh_Batman_18
27	2	Toilet 4	0	0	022_HappyHigh_Batman_18
28	2	Tornado 1	0	1	022_HappyHigh_Batman_18
29	2	War 2	0	1	022_HappyHigh_Batman_18
30	2	War 8	0	1	022_HappyHigh_Batman_18
31	2	Windmill 1	0	0	022_HappyHigh_Batman_18
32	2	Yarn 1	0	0	022_HappyHigh_Batman_18
33	3	Galaxy 2	1	1	031_SadHigh_TheEnglishPatient_18
34	3	Lake 15	1	1	031_SadHigh_TheEnglishPatient_18
35	3	Nude couple 5	1	1	031_SadHigh_TheEnglishPatient_18
36	3	Nude woman 15	1	1	031_SadHigh_TheEnglishPatient_18
37	3	Bird 1	1	0	031_SadHigh_TheEnglishPatient_18
38	3	Food 3	1	0	031_SadHigh_TheEnglishPatient_18
39	3	Pinecone 1	1	0	031_SadHigh_TheEnglishPatient_18
40	3	Present 1	1	0	031_SadHigh_TheEnglishPatient_18
41	3	Dog 26	0	1	031_SadHigh_TheEnglishPatient_18
42	3	Dog attack 3	0	1	031_SadHigh_TheEnglishPatient_18
43	3	Fire 11	0	1	031_SadHigh_TheEnglishPatient_18

44	3	Shot 1	0	1	031_SadHigh_TheEnglishPatient_18
45	3	Destruction 2	0	0	031_SadHigh_TheEnglishPatient_18
46	3	Dirt 5	0	0	031_SadHigh_TheEnglishPatient_18
47	3	Garbage dump 3	0	0	031_SadHigh_TheEnglishPatient_18
48	3	Roofing 4	0	0	031_SadHigh_TheEnglishPatient_18
49	4	Cat 6	1	1	041_TenderHigh_Shine_10
50	4	Elephant 1	1	1	041_TenderHigh_Shine_10
51	4	Sailing 2	1	1	041_TenderHigh_Shine_10
52	4	Skydiving 1	1	1	041_TenderHigh_Shine_10
53	4	Cold 2	1	0	041_TenderHigh_Shine_10
54	4	Cold 4	1	0	041_TenderHigh_Shine_10
55	4	Lake 3	1	0	041_TenderHigh_Shine_10
56	4	Pig 1	1	0	041_TenderHigh_Shine_10
57	4	Car crash 3	0	1	041_TenderHigh_Shine_10
58	4	Flood 3	0	1	041_TenderHigh_Shine_10
59	4	Shark 1	0	1	041_TenderHigh_Shine_10
60	4	Snake 2	0	1	041_TenderHigh_Shine_10
61	4	Ambulance 2	0	0	041_TenderHigh_Shine_10
62	4	Destruction 10	0	0	041_TenderHigh_Shine_10
63	4	Garbage dump 8	0	0	041_TenderHigh_Shine_10
64	4	Jail 4	0	0	041_TenderHigh_Shine_10
65	5	Dog 8	1	1	012_FearHigh_JFK_8
66	5	Galaxy 4	1	1	012_FearHigh_JFK_8
67	5	Lightning 1	1	1	012_FearHigh_JFK_8
68	5	Scared cat 1	1	1	012_FearHigh_JFK_8
69	5	Bark 2	1	0	012_FearHigh_JFK_8
70	5	Dessert 7	1	0	012_FearHigh_JFK_8
71	5	Paperclips 4	1	0	012_FearHigh_JFK_8
72	5	Path 1	1	0	012_FearHigh_JFK_8
73	5	Animal carcass 6	0	1	012_FearHigh_JFK_8
74	5	Dog 24	0	1	012_FearHigh_JFK_8
75	5	Explosion 3	0	1	012_FearHigh_JFK_8
76	5	Ferret 1	0	1	012_FearHigh_JFK_8
77	5	Destruction 5	0	0	012_FearHigh_JFK_8
78	5	Feces 2	0	0	012_FearHigh_JFK_8
79	5	Hallway 1	0	0	012_FearHigh_JFK_8
80	5	Prison 2	0	0	012_FearHigh_JFK_8
81	6	Beach 7	1	1	024_HappyHigh_ManOfGalileeCD1_2
82	6	Bear 3	1	1	024_HappyHigh_ManOfGalileeCD1_2
83	6	Dog 13	1	1	024_HappyHigh_ManOfGalileeCD1_2
84	6	Sunset 4	1	1	024_HappyHigh_ManOfGalileeCD1_2
85	6	Bubble 2	1	0	024_HappyHigh_ManOfGalileeCD1_2
86	6	Pumpkin 1	1	0	024_HappyHigh_ManOfGalileeCD1_2
87	6	Street 4	1	0	024_HappyHigh_ManOfGalileeCD1_2

88	6	Wedding 12	1	0	024_HappyHigh_ManOfGalileeCD1_2
89	6	Ambulance 1	0	1	024_HappyHigh_ManOfGalileeCD1_2
90	6	Animal carcass 2	0	1	024_HappyHigh_ManOfGalileeCD1_2
91	6	Destruction 6	0	1	024_HappyHigh_ManOfGalileeCD1_2
92	6	Gun 9	0	1	024_HappyHigh_ManOfGalileeCD1_2
93	6	Fence 4	0	0	024_HappyHigh_ManOfGalileeCD1_2
94	6	Garbage dump 6	0	0	024_HappyHigh_ManOfGalileeCD1_2
95	6	Graveyard 1	0	0	024_HappyHigh_ManOfGalileeCD1_2
96	6	Shot 4	0	0	024_HappyHigh_ManOfGalileeCD1_2
97	7	Rollercoaster 1	1	1	033_SadHigh_PortraitOfALady_9
98	7	Shark 6	1	1	033_SadHigh_PortraitOfALady_9
99	7	Soldiers 4	1	1	033_SadHigh_PortraitOfALady_9
100	7	Thunderstorm 10	1	1	033_SadHigh_PortraitOfALady_9
101	7	Bar 3	1	0	033_SadHigh_PortraitOfALady_9
102	7	Boat 1	1	0	033_SadHigh_PortraitOfALady_9
103	7	Paperclips 2	1	0	033_SadHigh_PortraitOfALady_9
104	7	Wedding 4	1	0	033_SadHigh_PortraitOfALady_9
105	7	Cockroach 2	0	1	033_SadHigh_PortraitOfALady_9
106	7	Fire 2	0	1	033_SadHigh_PortraitOfALady_9
107	7	Flood 1	0	1	033_SadHigh_PortraitOfALady_9
108	7	Thunderstoorm 4	0	1	033_SadHigh_PortraitOfALady_9
109	7	Bored pose 1	0	0	033_SadHigh_PortraitOfALady_9
110	7	Bored pose 2	0	0	033_SadHigh_PortraitOfALady_9
111	7	Garbage dump 5	0	0	033_SadHigh_PortraitOfALady_9
112	7	Miserable pose 2	0	0	033_SadHigh_PortraitOfALady_9
113	8	Dog 5	1	1	042_TenderHigh_PrideAndPredjudice_1
114	8	Father 1	1	1	042_TenderHigh_PrideAndPredjudice_1
115	8	Nude couple 1	1	1	042_TenderHigh_PrideAndPredjudice_1
116	8	Parachuting 3	1	1	042_TenderHigh_PrideAndPredjudice_1
117	8	Food 2	1	0	042_TenderHigh_PrideAndPredjudice_1
118	8	Rugby 2	1	0	042_TenderHigh_PrideAndPredjudice_1
119	8	School 6	1	0	042_TenderHigh_PrideAndPredjudice_1
120	8	Solar panel 1	1	0	042_TenderHigh_PrideAndPredjudice_1
121	8	Explosion 6	0	1	042_TenderHigh_PrideAndPredjudice_1
122	8	Gun 2	0	1	042_TenderHigh_PrideAndPredjudice_1
123	8	Police 2	0	1	042_TenderHigh_PrideAndPredjudice_1
124	8	War 6	0	1	042_TenderHigh_PrideAndPredjudice_1
125	8	Angry face 1	0	0	042_TenderHigh_PrideAndPredjudice_1
126	8	Depressed pose 2	0	0	042_TenderHigh_PrideAndPredjudice_1
127	8	Depressed pose 4	0	0	042_TenderHigh_PrideAndPredjudice_1
128	8	Hangover 1	0	0	042_TenderHigh_PrideAndPredjudice_1
129	9	Galaxy 3	1	1	014_FearHigh_AlienTrilogy_5
130	9	Lightning 6	1	1	014_FearHigh_AlienTrilogy_5
131	9	Mother 7	1	1	014_FearHigh_AlienTrilogy_5

132	9	Skydiving 3	1	1	014_FearHigh_AlienTrilogy_5
133	9	Cups 1	1	0	014_FearHigh_AlienTrilogy_5
134	9	Fire hydrant 4	1	0	014_FearHigh_AlienTrilogy_5
135	9	Food 4	1	0	014 FearHigh AlienTrilogy 5
136	9	Rocks 5	1	0	014_FearHigh_AlienTrilogy_5
137	9	Gargoyle 2	0	1	014_FearHigh_AlienTrilogy_5
138	9	KKK rally 2	0	1	014_FearHigh_AlienTrilogy_5
139	9	Snake 3	0	1	014 FearHigh AlienTrilogy 5
140	9	Soldiers 10	0	1	014_FearHigh_AlienTrilogy_5
141	9	Bored pose 3	0	0	014_FearHigh_AlienTrilogy_5
142	9	Depressed pose 1	0	0	014_FearHigh_AlienTrilogy_5
143	9	Garbage dump 1	0	0	014 FearHigh AlienTrilogy 5
144	9	Sleepy pose 1	0	0	014_FearHigh_AlienTrilogy_5
145	10	Car 1	1	1	025_HappyHigh_OliverTwist_1
146	10	Dog 20	1	1	025_HappyHigh_OliverTwist_1
147	10	Flowers 3	1	1	025 HappyHigh OliverTwist 1
148	10	Rafting 5	1	1	025_HappyHigh_OliverTwist_1
149	10	Camping 7	1	0	025_HappyHigh_OliverTwist_1
150	10	Cold 5	1	0	025_HappyHigh_OliverTwist_1
151	10	Goat 2	1	0	025 HappyHigh OliverTwist 1
152	10	Ornament 1	1	0	025_HappyHigh_OliverTwist_1
153	10	Fire 3	0	1	025_HappyHigh_OliverTwist_1
154	10	Garbage dump 2	0	1	025_HappyHigh_OliverTwist_1
155	10	Plane crash 3	0	1	025 HappyHigh OliverTwist 1
156	10	Tornado 3	0	1	025_HappyHigh_OliverTwist_1
157	10	Destruction 3	0	0	025_HappyHigh_OliverTwist_1
158	10	Frustrated pose 7	0	0	025_HappyHigh_OliverTwist_1
159	10	Skinhead 1	0	0	025 HappyHigh OliverTwist 1
160	10	Toilet 4	0	0	025_HappyHigh_OliverTwist_1
161	11	Orangutan 1	1	1	034_SadHigh_BigFish_15
162	11	Parasailing 2	1	1	034_SadHigh_BigFish_15
163	11	Rainbow 1	1	1	034 SadHigh BigFish 15
164	11	Rock climbing 2	1	1	034_SadHigh_BigFish_15
165	11	Bricks 1	1	0	034_SadHigh_BigFish_15
166	11	Camping 1	1	0	034_SadHigh_BigFish_15
167	11	Cold 3	1	0	034 SadHigh BigFish 15
168	11	Paper 5	1	0	034_SadHigh_BigFish_15
169	11	Sad face 9	0	1	034_SadHigh_BigFish_15
170	11	Shot 3	0	1	034_SadHigh_BigFish_15
171	11	Snake 4	0	1	034 SadHigh BigFish 15
172	11	Weapon 1	0	1	034_SadHigh_BigFish_15
173	11	Bored pose 4	0	0	034_SadHigh_BigFish_15
174	11	Nude man 9	0	0	034_SadHigh_BigFish_15
175	11	Scared face 4	0	0	034 SadHigh BigFish 15

176	11	Yoga 5	0	0	034_SadHigh_BigFish_15
177	12	Rafting 4	1	1	043_TenderHigh_DanceWithWolves_4
178	12	Rollercoaster 3	1	1	043_TenderHigh_DanceWithWolves_4
179	12	Sailing 3	1	1	043_TenderHigh_DanceWithWolves_4
180	12	Sunset 2	1	1	043_TenderHigh_DanceWithWolves_4
181	12	Street 5	1	0	043_TenderHigh_DanceWithWolves_4
182	12	Timber 1	1	0	043_TenderHigh_DanceWithWolves_4
183	12	Wall 1	1	0	043_TenderHigh_DanceWithWolves_4
184	12	Yarn 4	1	0	043_TenderHigh_DanceWithWolves_4
185	12	Shark 4	0	1	043_TenderHigh_DanceWithWolves_4
186	12	Thunderstorm 10	0	1	043_TenderHigh_DanceWithWolves_4
187	12	War 5	0	1	043_TenderHigh_DanceWithWolves_4
188	12	Wolf 2	0	1	043_TenderHigh_DanceWithWolves_4
189	12	Dog 25	0	0	043_TenderHigh_DanceWithWolves_4
190	12	Frustrated pose 4	0	0	043_TenderHigh_DanceWithWolves_4
191	12	Miserable pose 5	0	0	043_TenderHigh_DanceWithWolves_4
192	12	Yoga 2	0	0	043_TenderHigh_DanceWithWolves_4

Appendix F - Data Storage

Contents of attached CD:

- This document as PDF file
- Statistical analysis as an R script
- Dataset of collected data
- Literature sources used in this research as PDF file
- A stimulus database spreadsheet with all information about the stimuli used
- All auditory, visual and audiovisual stimuli used for this research