Automaticity of Evaluations and Emotion–Attention Interactions in the Auditory Modality

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultät III
der Universität des Saarlandes

vorgelegt von

Tímea Folyi
aus Kapuvár, Ungarn

Saarbrücken

2016

Dekan:

Prof. Dr. Roland Brünken

Berichterstatter:

Prof. Dr. Dirk Wentura

Prof. Dr. Hubert Zimmer

Tag der Disputation: 20.04.2016.

“Now I will do nothing but listen,
To accrue what I hear into this song, to let sounds contribute toward it.

I hear bravuras of birds, bustle of growing wheat, gossip of flames, clack of sticks cooking my meals,
I hear the sound I love, the sound of the human voice,
I hear all sounds running together, combined, fused or following,
Sounds of the city and sounds out of the city, sounds of the day and night...”

Walt Whitman, Song of Myself

“By the power of perceiver and perceived
All kinds of things are born;
They soon pass away, not staying,
Dying out instant to instant.”

Avatamsaka Sutra 10

Author Note

Several parts of this work are also included in articles already published or are in preparation as drafts before submission for publication. Several parts are included in the published article of Folyi and Wentura (2015) including Experiments 1-2, and in the published article of Folyi, Liesefeld, and Wentura (2016) including Experiment 3. Furthermore, Experiment 4 and Experiments 6-7 are included in two drafts, respectively (Folyi & Wentura, in preparation, a, b). For these manuscripts, I am the first author. In keeping with the practice of these manuscripts, I use the term "we" instead of "I" throughout my thesis. The respective chapters that constitute the main part of these manuscripts are indicated with footnotes; however, all specific passages are not highlighted separately in the continuous text in order to facilitate smooth reading.

Acknowledgements

I would like to express my sincere gratitude to all the people who directly or indirectly contributed to this work. Especially, I would like to deeply thank my supervisor, Prof. Dr. Dirk Wentura, not only for the fruitful discussions and his thoughtful guidance, but also for his unceasing care, encouragement, and patience. Above all, I would like to thank him for always pointing to the brightest sides of situations and having an open ear for me. I owe many thanks to Dr. Heinrich René Liesefeld for the helpful discussions concerning ERP-methodology, and to Prof. Dr. Hubert Zimmer for being my referee. Furthermore, I wish to acknowledge my colleagues, for each of them being a unique source of inspiration. My special thanks go to my PhD mates, who made the past years cheerful and memorable. I am grateful for having shared these years with you. I wish to sincerely thank those special persons who made enormous contributions to this work indirectly: My family for their unconditional support, my friends, who felt to be always close even from thousands of kilometers distance, and my partner, who always believed in me. Last but not least, I thank the student assistants for their help in data collection, and all participants who took part in the experiments and listened to tremendous amount of beep tones and noises with remarkable patience. Without their contribution, this work could not have been realized.

Abstract

Without any doubt, the auditory modality has a crucial role in representing our emotional world. It is enough to think about the euphoria when listening to a favorite piece of music, the warm feeling when hearing the voice of a beloved person, or, at the other end of the spectrum, how unbearable it can be to hear a painful scream or even a simple whir of a dentist drill. Converging evidence suggests that people automatically (e.g., in the absence of conscious intention or outside of the focus of voluntary attention) encode the affective content of an incoming stimulus. However, this knowledge is dominantly based on visual affective research. The relative neglect of auditory modality by affective research is astonishing given the unique operating and organizing principles of our auditory system and its vital role in monitoring our environment. The present work aimed to get a more complete understanding of sound evaluation and its interaction with auditory attention concerning different aspects of automaticity (like fast and unintentional affective processes), while highlighting unique features of our auditory perception and attention, first of all, the basic importance of the temporal dimension in the auditory domain.

Specifically, our first research route investigated the boundaries for evaluation of complex, natural emotional sounds concerning the available time and the intentionality of the evaluation process. Our results indicated that natural emotional sounds can be evaluated rapidly (i.e., within a few hundred milliseconds) with a remarkable precision by the means of explicit evaluations, despite the apparent drawback of temporally extended information conveyed by natural sounds. Furthermore, we demonstrated that natural emotional sounds can be evaluated rapidly and unintentionally in a speeded, reaction time-based paradigm.

Beyond question, fast and efficient selection of affectively significant auditory stimuli from our complex and rapidly changing acoustic environment is of vital importance for our survival and well-being. Our second research route targeted the ques-

tion whether valenced (i.e., positive and/or negative) tones can be prioritized relative to neutral ones already at an early, perceptual stage of auditory processing. Our results from three event-related potential experiments suggest that (i) preferential attention to valenced tones can operate already at perceptual level as reflected in a differential enhancement of the auditory N1 for valenced relative to neutral tones within about 100 ms following tone onset; (ii) it is the general relevance of the valenced tones that governs early attentional processes rather than a specific valence category; (iii) early attentional enhancement for the task-irrelevant valenced tones can occur outside of the focus of voluntary attention; and (iv) this early attentional effect was not moderated by the affective-motivational context of anticipating positive or negative outcomes. Both its early temporal locus and relative independence from the characteristics of the concomitant task suggest that the preferential attentional enhancement for task-irrelevant valenced tones occurs automatically in the sense of involuntary attentional capture.

Given the special importance of the temporal aspect in the auditory modality, our third research route investigated whether valenced tones can be selected preferentially from temporally distributed auditory patterns under the circumstances when auditory attention is limited in time. A performance deficit termed as attentional blink is a well-established result in the visual modality during multiple target detection in rapid stimulus sequences. Importantly, affectively significant visual targets are relatively preserved from this temporal constraint of visual attention, thereby signaling their special attentional status. We examined whether target tones endowed with affective valence can remain relatively resistant to the multiple target detection deficit in rapid sound sequences. We found a weak indication of an affective bias that appeared as a preferential weighting of valenced targets over neutral ones in a short-range competition between the targets. We discussed the possible distinctiveness of the auditory dual-target deficit from the visual attentional blink concerning its temporal pattern and moderation by affective valence. Limitations of our findings have also been considered and discussed.

Taken together, the present work highlights that auditory perception, attention, and affective processing are strongly connected: Our auditory system allows for rapid sound evaluation even unintentionally, and affective information of sounds can bias auditory perception and attention with an early locus during sound processing. However, on the other side of the coin, the investigated affective processes appeared to be relatively "immune" to concomitant contextual changes (such as competing motivational and task demands), further supporting our claim about their automatic nature.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

% .... Percent

€.......... Euro

α.......... Rate of type I error

β.......... Rate of type II error

η_p^2.......... Partial Eta

χ^2.......... Test statistic from chi-squared distribution

µV.......... Microvolt

ACC .... Accuracy rate

ANOVA .... Analysis of Variance

AST .... Affective Simon Task

B .... Regression coefficient

cf.......... confer

d_z.......... Cohen's d_z

dB .... Decibel

dB SL .... Decibel relative to Sensation Level

dB(A) .... A-weighted Decibel

df.......... Degree of Freedom

EEG.......... Electroencephalography

e.g. .... For example

ER.......... Error Rate

ERP .... Event-Related Potential

F .... Test statistic from *F*-distribution

FAR.......... False Alarm Rate

fMRI.......... Functional Magnetic Resonance Imaging

FSOA .... Feature Start Onset Asynchrony

Hz.......... Hertz

HR .... Hit Rate

IADS .... International Affective Digitized Sounds

ICC .... Intraclass-Correlation

i.e .... That is

ISI .... Interstimulus Interval

M .... Mean

MANOVA .... Multivariate Analysis of Variance

Mdn .... Median

ms .... Millisecond

N .... Sample size

n.s. .... Non-Significant

p .... Probability of equally extreme test statistic, given null hypothesis is true

r .... Pearson‘s correlation coefficient

RASP .... Rapid Auditory Serial Presentation

ROI .... Region of Interest

RSVP .... Rapid Serial Visual Presentation

RT .... Reaction Time

s .... Second

SD .... Standard Deviation

SE .... Standard Error

SOA .... Stimulus Onset Asynchrony

t .... Test statistic from *t*-distribution

T1 .... First Target

T2 .... Second Target

vs .... Versus

Introduction: Automatic Evaluations and Affective Attention in the Auditory Modality

Some stimuli are highly relevant for our survival and well-being, including signals warning of potential danger or, at the opposite end of the spectrum, signals indicating potential benefits. Efficient selection of these affectively significant stimuli from the rapidly changing, multisensory environment is essential for fast and adaptive reactions to the challenges in our surroundings. Despite a growing research interest toward affective processing, remarkably, some aspects have been relatively neglected so far. Most importantly, we live and act in a multisensory environment. However, compared to the large body of research investigating the processing of evaluative stimuli in the visual domain (which encompasses hundreds and hundreds of published studies), similar attempts are relatively sparse in other stimulus modalities such as the auditory domain. The relative neglect of auditory modality by affective research is astonishing given that our auditory system has unique operating and organizing principles that allow it to take a vital role in monitoring our environment. Accordingly, unique features of auditory modality – which can be potentially relevant concerning affective processing – were scarcely taken into consideration so far (for exceptions, see, e.g., Czigler, Cox, Gyimesi, & Horváth, 2007; Schirmer & Escoffier, 2010). In the present work, we will argue that we need a clear understanding of the relevant specificities of auditory perception and attention, and we need to explore whether and how the modality-specific factors play a role in affective processing.

Despite the relative scarcity of auditory affective research compared with that in the visual domain, there are remarkable lines of investigations on sound evaluation employing various approaches. So far, auditory affective research has mainly addressed processing of emotional vocalizations (e.g., Bestelmeyer, Maurage, Rouger, Latinus, & Belin, 2014; Frühholz & Grandjean, 2013; Grandjean et al., 2005; Kotz, Kalberlah, Bahlmann, Friederici, & Haynes, 2013; Kotz et al., 2003; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; for reviews, see Kotz & Paulmann, 2011; Schirmer & Kotz, 2006), and emotions conveyed by music (e.g., Armony, Aubé, Angulo-Perkins, Peretz, & Concha, 2015; Juslin & Västfjäll, 2008; Khalfa, Schon, Anton, & Liégeois-Chauvel, 2005; Koelsch, Fritz, von Cramon, Müller, & Friederici, 2006; Vieillard et al., 2008; for a review, see Koelsch, 2014). Furthermore, there are notable representatives of exploring event-related brain potential correlates of complex emotional sounds (e.g., Czigler et al., 2007; Paulmann, Bleichner, & Kotz, 2013; Sauter & Eimer, 2009; Schirmer & Escoffier, 2010; Thierry & Roberts, 2007), preferential processing of conditioned valence of sounds (Bröckelmann et al., 2011, 2013), psychophysiological reactions (facial muscle reactions, startle reflex, heart rate change) to natural emotional sounds (e.g., Bradley & Lang, 2000; Hawk, Fischer, & Van Kleef, 2012), and non-symbolic, low-level acoustic features that can contribute to the evaluation of a wide range of sounds by using the approach of computational modelling (e.g., Weninger, Eyben, Schuller, Mortillaro, & Scherer, 2013).

However, our understanding on automatic affective processes in the auditory modality is sparse, especially while considering modality-specific aspects. The present work aims to contribute to the growing field of auditory affective research by systematic investigation of automatic sound evaluations and preferential attentional effects for affectively significant tones, while taking into consideration auditory modality-specific factors. As the present work is rather exploratory, we will limit the scope of investigations by focusing particularly on one of the most conspicuous characteristics of auditory modality, that is, the importance of the temporal dimension, correspondingly to the transient nature and complex temporal structure of sounds. Furthermore, we will focus primarily on certain aspects of automaticity such as fast and unintentional affective processes, and, related to the question of automaticity, we will investigate involuntary preferential attentional effects for affectively significant tones. Additionally, when investigating affective evaluations and prioritization of affective stimuli in the auditory modality, we will concern ourselves with the basic positive versus negative affective valence dimension of evaluative stimuli that can be considered as a core aspect of affective evaluations (while its subjective affective experience often labeled as pleasant to unpleasant, e.g., Barrett, 2006). Given this backdrop, we will target the lingering question of emotional attention explicitly first time in the auditory modality, that is, whether basic affective biases in attentional processes are driven by a specific valence category (especially by negative valence), or by the general relevance of affectively significant information independently from the direction of valence. As a further but not exclusive possibility, we will investigate whether affective attentional biases vary flexibly regarding the challenges of different affective-motivational contexts. To target these questions, we will employ not only the paradigms of affective research that have been grounded dominantly in visual investigations, but we will also utilize methodology adapted from basic auditory attentional research. In the following, we will summarize the most relevant theoretical background for our work, which is related to basic auditory and – dominantly visual – affective research. After that, we will delineate our specific research questions that will be targeted by three research routes and presented in three main sections of this thesis.

The Auditory World

Our auditory system has unique operating and organizing principles (see, e.g., King & Nelken, 2009). Accordingly, we emphasize in the present work that we need to take into account these unique features when investigating sound evaluation and preferential attention for affective sounds. In the following, we will give an overview of the specificities of auditory perception that can be relevant for auditory affective research, and on how the selection of significant as opposed to irrelevant information accomplished by auditory selective attention.

Unique features of the auditory modality

Our auditory perception is in several aspects essentially different from vision, which is predominantly considered in the context of affective research. While vision has a limited spatial extent, auditory perception is omnidirectional (i.e., it covers 360-degree in space). Compared with vision, auditory perception is less dependent on the spatial distance of the source, and sounds are perceived even if the sound source is obscured. Moreover, hearing is more insistent, in the sense that we cannot control at the sensory periphery easily whether receiving a sensory input or not (in other words, while we can shut our eyes or look away, we cannot “shut our ears” and “listen away” in a comparable way, Hughes & Jones, 2001). Compared with vision, auditory perception is more “obligatory”: Auditory information is processed and organized more extensively already before it reaches the cortex (e.g., King & Nelken, 2009), and pre-attentively in sensory-specific cortices (e.g., Näätänen, Astikainen, Ruusuvirta, & Huotilainen, 2010), accounting for many complex and refined auditory abilities. Furthermore, it is characterized by an extensive pre-attentive change detection system that allows for rapid detection of significant changes in the auditory input that deviates from an internal, predictive model of the acoustic environment (e.g., Bendixen, SanMiguel, & Schröger, 2012; Näätänen, Paavilainen, Rinne, & Alho, 2007). In line with these characteristics, without doubt, the auditory modality has a great importance in monitoring our environment.

Moreover, as it will be highlighted throughout the present work regarding different aspects of auditory perception and attention, time plays a central role in the auditory modality. While the visual input can be described as spatial patterns of activation on the retina, auditory input reaches sensory periphery as variations in pressure distributed in time that is transmitted as mechanical vibrations to the eardrum. In line with the different nature of the sensory inputs, it is often claimed that while space is retained as an essential organizing principle of visual stimulus representation, temporal information (besides tone-frequency) remains a similarly important organizing principle of auditory stimulus representation (for reviews, see, B. C. J. Moore, 2012; Näätänen & Winkler, 1999). Accordingly, one of the most conspicuous characteristics of our auditory world is its transient and also “temporally extended” nature. First, sounds are characterized by complex and rapid changes in frequency and amplitude, and, moreover, in meaningful sounds (such as environmental sounds, speech, or music) information is conveyed rather by these changes themselves than the relatively static characteristics (B. C. J. Moore, 2012). Second, in natural environments sounds typically occur as several simultaneous and successive acoustic events. In our perception, multiple acoustic events can link together in time, and these temporally extended representations are often considered as the fundamental building blocks of our auditory world (for comprehensive discussions on temporally extended auditory object formation, see e.g., Bregman, 1990; Griffiths & Warren, 2004; Winkler, Denham, & Nelken, 2009; Winkler & Schröger, 2015). Correspondingly to the basic importance of the temporal dimension, the auditory system is characterized by high temporal resolution and comprehensive temporal analysis of the sensory input (e.g., Boemio, Fromm, Braun, & Poeppel, 2005; Repp & Penel, 2002; Santoro et al., 2014; Zatorre & Belin, 2001; for a review, see B. C. J. Moore, 2012).

Taking into consideration the vital role of hearing in monitoring our environment, the importance of the temporal aspect and, consequently, the refined temporal abilities in the auditory modality, it seems straightforward that our auditory system should allow for fast and efficient evaluation and selection of significant sounds from our ever-changing surroundings. On the other hand, gradual evolving of auditory information due to the complex and extended temporal structure of sounds would predict rather slow auditory affective processes. In the present work, we will investigate evaluation and preferential attentional effects for valenced sounds with a special focus on the temporal aspect of auditory perception and attention.

AUDITORY SELECTIVE ATTENTION

Selection of significant cues that are highly relevant for our survival and well-being at the expense of less relevant information from the rapidly changing and vast acoustic input is vital for adaptive reactions to the challenges of our environment. Selective attention is one of the most dominant concepts in the fields of both attention-emotion interactions and basic auditory research (for reviews, see Fritz, Elhilali, David, & Shamma, 2007; Giard, Fort, Mouchetant-Rostaing, & Pernier, 2000; Yiend, 2010). The notion refers to the mental ability that allows selecting relevant information as opposed to irrelevant information, thereby only a subset of the vast sensory environment receives profound analysis (e.g., Broadbent, 1958; Lakatos et al., 2013).

Theoretical questions of auditory selective attention were examined extensively by using event-related potential (ERP) methodology. Early findings led to a lingering debate about at what point of auditory processing, and by which mechanism a differential processing of relevant relative to irrelevant auditory stimuli can occur. These investigations gave rise to two influential physiological theories on auditory selective attention: the gain theory of attention (Hillyard, Hink, Schwent, & Picton, 1973; Hillyard, Mangun, Woldorff, & Luck, 1995) and the attentional trace theory (Näätänen, Gaillard, & Mäntysalo, 1978; for a review, see Giard et al., 2000). The gain theory of attention assumes that attended stimuli can receive relative enhancement compared with unattended stimuli during "obligatory" auditory analysis via sensory gain control (e.g., Hillyard et al., 1973; Hillyard et al., 1995; Hillyard, Teder-Sälejärvi, & Münte, 1998). Accordingly, auditory attention has been found to selectively enhance sensory analysis of the attended compared with unattended auditory stimuli (especially reflected in the well-established relative enhancement of the "obligatory" auditory N1 component that is functionally linked to initial sound detection, and it, at least partly, originates from the auditory cortex), suggesting that selective attention can amplify the same sensory processes that are involved in analyzing a certain sound even without attention (e.g., Bidet-Caulet et al., 2007; Hillyard et al., 1973; Schröger, Marzecová, & SanMiguel, 2015; Woldorff et al., 1993; Woldorff & Hillyard, 1991; for a review, see Giard et al., 2000). The attentional trace account of auditory selective attention emerged as a reinterpretation of the same negative-going shift in the ERPs to the attended relative to unattended tones as found by Hillyard et al. (1973). The attentional trace theory explained the observed early attentional effects as correlates of a dedicated "endogenous" attentional mechanism that compares the sensory input with an attentional trace (i.e., representation about the distinguishing features of the relevant stimuli). Thus, the observed attentional difference emerges not as the result of the selection process (i.e., enhanced sensory analysis) but it represents the selection process itself. However, this "endogenous" attentional effect can appear with an early onset (i.e., in the time window considered as relating to "obligatory" sensory analysis) when sounds are presented at a rapid rate (e.g., Alho, Töttölä, Reinikainen, Sams, & Näätänen, 1987; Näätänen, 1982; Näätänen et al., 1978).

To resolve the long-lasting conflict of these competing theories that inspired numerous studies (e.g., Alho et al., 1987; Bidet-Caulet et al., 2007; Jääskeläinen & Ahveninen, 2014; Lehmann & Schönwiesner, 2014; Näätänen, 1982; Ross, Hillyard, & Picton, 2010; Woldorff et al., 1993; Woldorff & Hillyard, 1991), a current agreement proposes at least two mechanisms of auditory selective attention as reflected in the ERPs to attended compared with unattended auditory stimuli: an early relative enhancement of "obligatory" sensory analysis by selective attention and a later "endogenous" mechanism possibly reflecting attentional control regarding the attentional trace (for reviews, see Giard et al., 2000; Schröger et al., 2015). Moreover, a current view of auditory selective attention suggests a highly adaptive mechanism that can selectively enhance representations of important sounds at the level of processing at which a certain representation is of most relevance, thus, it can take effect at multiple stages including early auditory processing (e.g., Giard, Collet, Bouchet, & Pernier, 1994; Kauramäki et al., 2012; Lehmann & Schönwiesner, 2014; for reviews, see Fritz et al., 2007; Giard et al., 2000; Schröger et al., 2015).

Beforehand, with the notion of selective attention we did not differentiate between voluntary (i.e., "top-down", "endogenous", or task-relevant) and involuntary (i.e., “bottom-up”, “exogenous”, or relatively automatic) attentional selection, nonetheless, the above described classical investigations into the neurophysiological basis of auditory selective attention mainly concerned with the former case (thus, participants were instructed to attend selectively a certain auditory stimulation and ignore another). However, unexpected and perceptually salient sounds can also trigger auditory attention involuntarily, and, moreover, they can enhance responses already in the time range of the N1 component that is associated with transient-detection (see, e.g., Escera, Alho, Winkler, & Näätänen, 1998; Horváth, Winkler, & Bendixen, 2008).

Importantly, preferential attentional effects to task-irrelevant but affectively significant stimuli have often been observed in various paradigms in the visual modality, and it was also reported in the auditory modality (for behavioral demonstrations in the auditory modality, see, e.g., Bertels, Kolinsky, & Morais, 2012; Nielsen & Sarason, 1981; for a review of dominantly visual studies, see Yiend, 2010). Selective attention theories often resolve the preferential attentional effects for affectively significant cues by assuming that affectively significant stimuli constitute a special class of high-salience stimuli relative to neutral ones, that in turn can bias attentional selection in a relatively automatic way (for a discussion of preferential visual attention to affective stimuli, see Yiend, 2010). Given the vital importance of detecting auditory cues that can be relevant for our survival or well-being, one of the main questions raised by the present work is whether affectively significant tones can be selected preferentially relative to neutral ones relatively “automatically”. We will target this question by two approaches: (1) We will investigate whether valenced tones can receive preferential attention relative to neutral tones already at an early, perceptual level of sound encoding, and (2) we will examine whether valenced tones can be detected preferentially in rapid sound sequences, in both cases under the conditions of limited voluntary attention. In the following, we will define how we use the term of automaticity in the present work, and what aspects of automaticity will be investigated concerning affective evaluations and preferential attentional effects.

A Decompositional Approach of Automaticity and the Aspects of Interest

Despite its central role in cognitive and affective research, the concept of automaticity is not used consequently in the literature, and a consensus has not emerged regarding its definition. Traditional views on automaticity has considered it as an all-or-none phenomenon (i.e., as a process being either non-automatic and thus requiring limited processing capacity, or automatic that is only minimally dependent on such limits) or as a single continuum from non-automatic to automatic processing (for the capacity view of automaticity, see e.g., Hasher & Zacks, 1979; Logan, 1979; Shiffrin & Schneider, 1977). However, unitary concepts of automaticity have been challenged by evidence suggesting that that there is no perfect coherence among the automaticity features or ingredients (such as efficient, unintentional, uncontrollable, and unconscious, Bargh, 1992, 1994). More recent differentiated analyses support this decompositional view, suggesting that the term of automaticity rather refers to a bundle of only loosely related features or factors that can influence processes, such as affective evaluations or orienting of attention (Moors, 2015; Moors & De Houwer, 2006). Therefore, in this alternative picture, processes are not considered as automatic or non-automatic in an all-or-none manner, but rather a relevant descriptor of a specific process can be whether it is relatively automatic along certain automaticity factors or features compared with other processes. Moreover, in a recent differentiated view of automaticity, factors determining the relative automaticity of a process are assumed to compensate each other in an additive way, at least to some extent, thereby influencing the occurrence of processes such as evaluations. For example, brief stimulus duration may not allow for certain processes (e.g., evaluations) to occur, but this brief duration can be compensated by preactivation in line with goals or expectations (Moors, 2015).

In line with the approach of decompositional views on automaticity, in the present work, we will focus particularly on the following features: unintentional (i.e., uncontrolled in a promoting sense; when an act is not caused by an intention, thus, by the goal to engage into the certain act), efficient (i.e., does not require substantial amount of attentional resources), and fast affective processes (Moors, 2015; Moors & De Houwer, 2006). Furthermore, we refer an attentional enhancement as involuntary attentional capture if it occurs rapidly, unintentionally, and relatively independently of voluntary attentional processes (thus, it can occur in the absence of less automatic forms of attention; for reviews, see Pourtois, Schettino, & Vuilleumier, 2013; Vuilleumier, 2005). Note that this is a rather strict criterion for specifying the automaticity of affective attention as compared with mere task-irrelevance (Okon-Singer, Lichtenstein-Vidne, & Cohen, 2013; Pessoa, 2005).

Attentional Biases towards Valent Stimuli

In the present work, we will target first time in the auditory modality systematically the pervasive question of emotional attention, that is, whether basic affective biases in attentional processes are driven by a specific valence category or by the general relevance of valenced information independently from the direction of valence. The most influential representative of the former assumption is the evolutionary-based negativity bias hypothesis (e.g., Öhman & Mineka, 2001) that predicts preferential attentional effects for negative, especially for threat-related stimuli (e.g., angry or fearful faces). However, numerous empirical evidences indicate that preferential attentional effects can be triggered by positive stimuli as well (e.g., Brosch, Sander, Pourtois, & Scherer, 2008; Müller, Rothermund, & Wentura, 2015; Wentura, Müller, & Rothermund, 2014), opening the way to accounts that do not restrict affective attentional biases to negative valence category, but extend it to stimuli that are relevant concerning our needs and goals in general irrespective of the direction of valence. We will give an overview of these assumptions, and additionally, a comprehensive account that proposes that fixed affective attentional biases (either to a specific valence category or valenced stimuli in general) reflect an oversimplified view, and, instead, it suggests that attentional biases can vary flexibly according to the challenges of different motivational-emotional contexts (Rothermund, Voss, & Wentura, 2008).

NEGATIVITY BIAS HYPOTHESIS

The negativity bias hypothesis (e.g., Hansen & Hansen, 1988; Öhman & Mineka, 2001) is a far-reaching assumption in the field of cognition and emotion, with the central prediction that our attentional system is tuned to prioritize negative (or more specifically threatening) information over neutral and positive information. Öhman and Mineka (2001) proposed a "hard-wired" threat-detector module that provides preferential sensitivity to cues that signaled threat in our evolutionary past (e.g., dangerous predators), presumably via an amygdala-mediated mechanism. This sensitivity is assumed to have evolved as detection of dangers has a higher survival value relative to detecting neutral and even positive cues: While missing to detect a danger can lead to severe harm, missing an opportunity has less severe consequences, at least in the short term. In line with this assumption, negative information has often been found to weigh more heavily than neutral and positive information on engaging and/or holding attention (in the visual modality, see, e.g., Armony & Dolan, 2002; Carlson et al., 2012; Carretié, Mercado, Tapia, & Hinojosa, 2001; Hajcak & Olvet, 2008; Öhman, Flykt, & Esteves, 2001; Öhman, Soares, Juth, Lindström, & Esteves, 2012; Pourtois, Grandjean, Sander, & Vuilleumier, 2004; Pratto & John, 1991; Schupp et al., 2004; N. K. Smith, Cacioppo, Larsen, & Chartrand, 2003). However, studies investigating preferential attention to negative stimuli often employed only negative valence category without any comparison with positive valence (in the visual modality, see, e.g., Armony & Dolan, 2002; Carlson et al., 2012; Öhman et al., 2001). Moreover, it is suggested that the negativity bias can depend heavily on the employed paradigm and stimulus material (e.g., Hahn & Gronlund, 2007; Hilgard, Weinberg, Hajcak Proudfit, & Bartholow, 2014; Weinberg & Hajcak, 2010), experimental situation (e.g., L. Rohr & Abdel Rahman, 2015; N. K. Smith et al., 2006), and personal factors (e.g., high anxiety, Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007), thereby undermining the view of an "obligatory" and unitary negativity bias.

General relevance principle in affective attentional biases

One can argue that a fixed negativity bias can also be maladaptive as not only overlooking dangers but also missing opportunities can have severe consequences on a longer time scale. In line with this assumption, a remarkable line of evidence suggests that our attention can be biased towards positive stimuli in a similar or even more pronounced way as towards negative stimuli, thereby challenging the view of a solid negativity-dominance in affective attentional biases (in the visual modality, for evidence of attentional capture by natural positive stimuli, see, e.g., Brosch, Sander, & Scherer, 2007; and by reward-associated visual stimuli, see B. A. Anderson, Laurent, & Yantis, 2011a, 2011b; and B. A. Anderson & Yantis, 2013; for evidence of positivity bias, see, e.g., Hodsoll, Viding, & Lavie, 2011; Raymond & O'Brien, 2009; Oca, Villa, Cervantes, & Welbourne, 2012; for comparable preferential attentional effects for negative and positive valence, see, e.g., Brosch et al., 2008; Keil & Ihssen, 2004; Müller et al., 2015; Nummenmaa, Hyönä, & Calvo, 2009; Schupp, Junghöfer, Weike, & Hamm, 2003; Wentura et al., 2014; for an auditory evidence, see Bröckelmann et al., 2011). Moreover, amygdala reaction was also found repeatedly for positive stimuli (in the visual modality, see, e.g., Fitzgerald, Angstadt, Jelsone, Nathan, & Phan, 2006; for a review, see Zald, 2003; in the auditory modality, see e.g., Fecteau, Belin, Joanette, & Armony, 2007; Sander & Scheich, 2001), suggesting that the amygdala is not reserved solely for threat-detection (for a review on the involvement of amygdala in appraisal of general relevance, see Sander, Grafman, & Zalla, 2003). In line with these findings, the assumption of a general relevance principle predicts that basic affective attentional biases are not restricted to negative stimuli, but they are driven by the general relevance of affectively significant cues (in the visual modality, see, e.g., Müller et al., 2015; Wentura et al., 2014). In accordance with the appraisal theories of emotion (e.g., Ellsworth & Scherer, 2003; Scherer, 2001), such attentional biases prioritize highly relevant stimuli concerning the individual's goals and needs regardless of the direction of their valence.

However, intuitively, the adaptive value of rigid affective attentional biases – either in the form of a fixed negativity bias or as a bias for relevant information in general – can be questioned as displaying a certain attentional bias can be adaptive given certain environmental or personal factors but less adaptive or even dysfunctional under other circumstances. This qualm will lead us to the assumption of adaptively flexible affective attentional biases.

ADAPTIVELY FLEXIBLE BIASES: MOTIVATIONAL COUNTER-REGULATION

The assumption of flexible attentional biases that selectively prioritize positive *or* negative information based on the requirements of a motivational context appears advantageous as they can provide higher compatibility with the actual goal pursuit and self-regulation. Flexible adjustment of attentional biases to valent information regarding the current motivational focus is conceivable in two antagonistic forms: (1) Attentional biases are tuned in line with our motivational orientations, meaning that our attentional system prioritizes negative information when prevention or avoidance motivational focus is applied (e.g., trying to prevent a dangerous event), and positive information when promotion or approach motivational focus is applied (e.g., trying to obtain a reward; correspondence principle, see, e.g., N. K. Smith et al., 2006; Strack & Deutsch, 2004). The correspondence principle, on the one hand, appears to be intuitively advantageous as it implies that attentional biases support higher-level motivational orientations. On the other hand, however, it would also act against an adaptive flexibility as promoting a mutual reinforcement between the attentional biases and the current motivational orientation, and thus, in an extreme case, it would result in becoming locked-up in an affective-motivational state (cf. self-maintaining nature of emotional disorders; e.g., Mathews & MacLeod, 2005; see also N. K. Smith et al., 2006). For that reason, Rothermund, Voss, and Wentura (2008) suggested a (2) counter-regulation principle in affective attentional biases, that is, in order to promote a homoeostatic regulation of affective-motivational states, attentional biases to valent stimuli operate incongruently to the current motivational-emotional orientations (i.e., it predicts attentional bias towards positive stimuli in negative motivational context, and towards negative stimuli in positive motivational context). Such highly flexible affective attentional biases would prevent escalation or perseveration of affective-motivational states. The assumption of a counter-regulation in attentional biases towards valent visual stimuli has received numerous empirical support as regarding to the affective-motivational contexts promoted by previously experienced positive or negative events (Rothermund, 2003; e.g., Rothermund, Gast, & Wentura, 2011; Schwager & Rothermund, 2014), current emotional states (Schwager & Rothermund, 2013a), and anticipating positive or negative future outcomes (Rothermund et al., 2008; Schwager & Rothermund, 2013b; Wentura, Voss, & Rothermund, 2009; see also Rothermund, Wentura, & Bak, 2001).

In the present work, we will contrast the above outlined assumptions (i.e., bias towards a specific valence category or general relevance principle in affective attentional biases) systematically first time in the auditory modality, and, additionally, we will investigate whether the affective-motivational context of anticipating positive or negative future outcomes can moderate basic affective attentional biases in the auditory domain. Given this backdrop, we will delineate our three specific research routes that will be presented in the three main sections of the thesis.

The Focus of the Present Work

In brief, the first research route of the present work investigates the boundaries for evaluation of complex, naturally occurring emotional sounds concerning the available time and the intentionality of the evaluation process. The second research route targets the question whether task-irrelevant and to-be-ignored tones with affective valence can elicit preferential attention already at an early, perceptual level of sound encoding. Additionally, we will investigate whether this effect can occur outside of the focus of voluntary attention. The third research route will investigate the temporal limits of auditory attention in a so-called "attentional blink" paradigm (see e.g., Shapiro, Raymond, & Arnell, 1997, in the visual modality), in order to explore whether tones with affective valence are more resistant to the temporal constraints of auditory attention relative to neutral tones.

Furthermore, while in the first research route we will examine automatic (in the sense as fast and unintentional) evaluations of natural, complex emotional sounds with a priori valence, the second and third research routes will take a complementary approach: We will induce affective valence experimentally during a learning phase by assigning positive, negative, and neutral meaning to different tone-frequencies of simple, sinusoidal tones in a balanced design. This method ensures strict control over arbitrary physical differences, and, moreover, it also provides control over the time of availability of the valance-related information, as the evaluative meaning carried by different tone-frequencies will be available already at about sound onset. Both of these characteristics of the latter approach are advantageous for a well-controlled investigation of preferential affective attentional effects that possibly take effect rapidly and already at an early level of sound encoding. Moreover, in the second and third research routes, we will systematically contrast the above outlined assumptions regarding basic affective attentional biases, namely, whether attentional prioritization in the auditory modality is driven by a specific (especially negative) valence category, or by the general relevance of valenced stimuli. In the second research route, additionally, we will investigate whether positive and negative affective-motivational context can modulate these basic affective attentional biases. In the following, we will unfold the most important research questions of these three routes.

FAST AND UNINTENTIONAL EVALUATION OF NATURAL EMOTIONAL SOUNDS? (EXPERIMENTS 1-2)

As emphasized above, an important distinguishing characteristic of auditory stimuli is the fact that sounds carry temporally distributed information. Consequently, while the evaluative content of emotional pictures – the most common type of stimulus in visual affective studies – is available instantly at stimulus onset, one can assume that the evaluative content of natural sounds becomes available only after a considerable exposure time due to temporal unfolding. Thus, on the one hand, one can argue that the complex temporal structure of sounds would necessitate gradual extraction of evaluative information over time. On the other hand, fast extraction of affective information from sounds seems to be crucial to detect and react to possibly beneficial or dangerous events rapidly. Hence, the claim of a slow affective processing in the auditory modality, which in turn would lead to relatively slow detection of significant auditory cues, seems ecologically invalid. Given this apparent contradiction, the first aim of this thesis work is to examine the boundary of sound exposure duration that is needed to evaluate complex, naturally occurring emotional sounds.

Furthermore, we will investigate this question while taking into account the intentionality of the evaluation process: While in Experiment 1, we ask participants to evaluate sounds explicitly by rating the affective valence of brief segments of natural emotional sounds, in Experiment 2, an indirect method will be used to investigate unintentional sound evaluation. While there is converging evidence that valence of visual stimuli (e.g., emotional pictures, valent words) can be processed rapidly and unintentionally (e.g., De Houwer & Eelen, 1998; Öhman et al., 2001; Wentura et al., 2014), we have relatively little knowledge about the automaticity of sound evaluation. Moreover, speeded, reaction time (RT)-based paradigms for assessing relative automatisms of evaluations are available almost exclusively in the visual domain. A possible reason for this discrepancy can be again the time-bound character of auditory processing: As brief stimulus presentation times and/or fast reactions to the presented stimuli are often critical in experimental paradigms investigating automatic evaluations, complex natural emotional sounds have been employed only carefully (see, e.g., L. D. Scherer & Larsen, 2011). In Experiment 2, we will introduce an auditory variant of the affective Simon task (for the original visual version, see De Houwer & Eelen, 1998), in which participants are presented with natural emotional sounds and give intrinsically valent responses to a valence-neutral perceptual stimulus feature that is orthogonally varied to valence. If in this paradigm, in which valence is completely task-irrelevant and, moreover, participants are explicitly instructed to ignore it, affective valence of the sounds influences speeded behavioral responses to a task-relevant feature, we can conclude that the presented sounds were evaluated relatively automatically in the terms of fast and unintentional evaluations.

PREFERENTIAL ATTENTION TO VALENCED TONES AT PERCEPTUAL LEVEL? (EXPERIMENTS 3-5)

A further question targeted by the present work is whether valenced (i.e., negative and/or positive) auditory stimuli can be prioritized at early, perceptual stages of sound encoding by involuntary attentional enhancement. Given the crucial importance of detecting cues that are relevant for our survival and well-being, it is suggested – dominantly based on visual research – that affective information can facilitate stimulus processing in a similar way as “cold” (like task-relevant or perceptual saliency-based) forms of attention via automatic “natural” attentional selection (Easterbrook, 1959; Pourtois et al., 2013; Schupp, Flaisch, Stockburger, & Junghöfer, 2006; Vuilleumier & Huang, 2009), and thereby enhancing cortical sensory processing (e.g., Junghöfer et al., 2006; Keil et al., 2003; Olofsson, Nordin, Sequeira, & Polich, 2008; Pourtois et al., 2013; Schupp et al., 2003; Vuilleumier & Huang, 2009; Vuilleumier, 2005). A rapid activation of subcortical regions, in particular, the amygdala, and the connected sensory cortices is often supposed to account for these early affective effects (see, e.g., Furl, Henson, Friston, & Calder, 2013; Frühholz et al., 2015; Morris et al., 1998; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004; for reviews, see Phelps, 2006; Phelps & LeDoux, 2005; Pourtois et al., 2013; Vuilleumier, 2005; Vuilleumier & Huang, 2009). Moreover, there is recent evidence from the auditory modality indicating that magnetic auditory evoked fields can be enhanced by affectively significant auditory stimuli (Bröckelmann et al., 2011, 2013), suggesting that attentional benefit for evaluative sounds can occur rapidly after stimulus onset; however, these responses were recorded during passive listening, allowing little to conclude about the automaticity of the effect. In the present work, we will investigate whether rapid involuntary attention can enhance sensory encoding of valenced tones relative to neutral ones. We aim to investigate the automaticity of the expected attentional enhancement in the terms as it is fast, unintentional, and it can occur in the absence of voluntary attention to these tones.

Based on dominantly visual studies, recent theoretical models on emotional attention embed preferential attention for affective stimuli into a comprehensive view of different attentional mechanisms (for reviews, see Okon-Singer et al., 2013; Pourtois et al., 2013; Vuilleumier & Huang, 2009). On the one hand, affectively significant stimuli supposed to receive attentional enhancement in a reflexive way even in the absence of other, more voluntary controlled attentional processes. Moreover, under many conditions, several attentional processes (e.g., voluntarily directed attention or attentional capture based on affective and perceptual saliency) operate in parallel with or in addition to each other (for visual studies, see, e.g., Keil, Moratti, Sabatinelli, Bradley, & Lang, 2005; Schupp et al., 2007; Vuilleumier, Armony, Driver, & Dolan, 2001; for reviews, see, Pourtois et al., 2013; Vuilleumier & Huang, 2009; Vuilleumier, 2005; for additive effects of spatial attention and emotional prosody in voice-specific areas, see Grandjean et al., 2005). On the other hand, several visual studies have found abolished or strongly reduced preferential effects for affective stimuli when participants' attention was directed to a demanding concurrent task, indicating that the readiness for prima facie automatic preferential processing of affective stimuli strongly depends on contextual factors like the focus of voluntary attention and task demands (see, e.g., Eimer, Holmes, & McGlone, 2003; Holmes, Vuilleumier, & Eimer, 2003, Wiens & Syrjänen, 2013, suggesting that early differential response to emotional pictures depends on voluntary attention; and

Pessoa, McKenna, Gutierrez, & Ungerleider, 2002; Pessoa, Padmala, & Morland, 2005, for demonstrations that the preferential amygdala response to emotional visual stimuli can be abolished by a demanding competing task; but see also Pourtois, Spinelli, Seeck, & Vuilleumier, 2010, for a demonstration that only the late part of the amygdala effect is affected by voluntary attention; for reviews, see Eimer & Holmes, 2007; Pessoa, 2005; Okon-Singer et al., 2013).

In the present work, Experiment 3 will investigate ERP-correlates of preferential attention to valenced tones when they are entirely task-irrelevant, and participants are engaged in a concurrent demanding task. Additionally, in order to investigate whether prima facie automatic attentional effects for valenced relative to neutral auditory stimuli can occur strictly in the absence of voluntary attention, Experiment 4 employs a stricter control over the direction of voluntary attention by presenting continuous concurrent task-relevant stimulation. Furthermore, in line with the assumption of adaptively flexible affective biases in attentional processes, in Experiment 5 we will investigate whether the affective-motivational context of anticipating negative or positive future outcomes moderates affective attentional biases in the auditory domain.

PREFERENTIAL SELECTION OF VALENCED TONES FROM RAPID SOUND SEQUENCES? (EXPERIMENTS 6-7)

Given the basic importance of the temporal aspect in the auditory modality, we argue that efficient selection of affective information from temporally distributed auditory patterns can be comparably crucial for detecting significant cues in our environment as the extensively studied preferential selection of visual affective information in space (for facilitated detection of emotional pictures in spatial arrays see, e.g., Öhman et al., 2001). Consequently, the question arises as to whether preferential attention to affectively significant tones can provide a detection benefit for these tones in rapid sound sequences, even under the circumstances when voluntary attention is limited in time.

The performance deficit termed as attentional blink is a well-established result in the visual modality during multiple target detection in rapid stimulus sequences: Detection performance for the second target is typically impaired after a correctly detected first target as a function of their temporal distance (for reviews, see Dux & Marois, 2009; Martens & Wyble, 2010; Shapiro et al., 1997). Numerous theories based on visual research interpreted this outcome as for example a first target-induced depletion of capacity-limited cognitive resources (e.g., Chun & Potter, 1995; Jolicœur & Dell'Acqua, 1999), and failure or over-application of attentional control (e.g., Lollo, Kawahara, Ghorashi, & Enns, 2005; Taatgen, Juvina, Schipper, Borst, & Martens, 2009; Olivers & Meeter, 2008). However, of most relevance to the present work, affectively significant visual targets are often relatively preserved from this dual-target deficit compared with neutral targets, indicating their special attentional status as they appear to be more resistant to the temporal constraints of visual attention (e.g., A. K. Anderson, 2005; A. K. Anderson & Phelps, 2001; Keil & Ihssen, 2004; Oca et al., 2012; Schwabe et al., 2011).

A transient performance decrement has been observed for a second target following a successfully detected first target in the auditory modality as well (e.g., Duncan, Martens, & Ward, 1997; Horváth & Burgyán, 2011; Shen & Mondor, 2006; Vachon, Tremblay, Hughes, & Jones, 2010). This outcome is often interpreted as an analogue to the visual attentional blink, leading to the notion of the auditory attentional blink. The main target of the third research route is to explore possible affective modulation of the attentional blink deficit first time in the auditory modality. As the auditory attentional blink is much less established compared with its visual counterpart, and, moreover, the visual and auditory attentional blink appears to have some notable differences, first we will investigate the basic phenomenon by introducing a dual-target detection task in a rapid auditory serial presentation paradigm (Experiment 6). Importantly, thereafter, we will introduce an affective variant of this paradigm with simple tones endowed with affective valence presented as second targets (Experiment 7). Given the assumed importance of selection of significant cues in time from the transient auditory input, we expect a pronounced resistance of valenced tones to a dual-target detection deficit in the auditory modality, due to rapid attentional facilitation to these tones.

Fast and Unintentional Evaluation of Emotional Sounds[1]

While converging evidence suggest that affectively significant visual stimuli are evaluated rapidly and unintentionally (e.g., De Houwer & Eelen, 1998; Öhman et al., 2001; Öhman & Mineka, 2001; Wentura et al., 2014), sound evaluation is relatively sparsely understood concerning these aspects of automaticity. Moreover, RT-based paradigms for assessing automatisms of evaluation are hardly available outside of the visual modality. We argue that this relative neglect can partly arise from the time-bound character of the auditory domain, as the assumption of gradually unfolding evaluative information of natural emotional sounds raises an issue for these para-

[1] This chapter is derived in part from an article of Folyi and Wentura published in Cognition and Emotion (2015), copyright by Taylor & Francis.
Available online at: http://www.tandfonline.com/10.1080/02699931.2015.1110514

digms, in which brief stimulus presentation times and/or fast reactions to the presented stimuli are often critical. Accordingly, brief (i.e., lasting approximately 1 s or less) natural emotional sounds have been employed only with caution in experimental paradigms investigating automatisms of evaluations.

We can illustrate this issue with an exception to the dominantly visual RT-based paradigms for assessing evaluations: Cross-modal evaluative priming effects have been demonstrated with evaluative auditory primes and evaluative visual targets (Carroll & Young, 2005; Goerlich et al., 2012; Marin, Gingras, & Bhattacharya, 2012; L. D. Scherer & Larsen, 2011; Schirmer, Kotz, & Friederici, 2002; Sollberger, Rebe, & Eckstein, 2003; Steinbeis & Koelsch, 2011; for an entirely auditory version employing evaluative speech, see Degner, 2011). Evaluative priming effect refers to the phenomenon that the time needed to evaluate a target stimulus is considerably shorter when a preceding briefly presented prime stimulus has the same affective valence (i.e., the prime and target are affectively congruent) compared to when it has a different valence (i.e., the prime and target are affectively incongruent; for a review, see Klauer & Musch, 2003). Evaluative priming in the visual domain is almost exclusively focused on brief prime durations and brief stimulus onset asynchronies (SOA) of prime and target (i.e., ≤ 300 ms) because longer SOAs are thought to be influenceable by strategic behavior and evaluation of the primes is thought to decay quickly. This is consistent with the finding that evaluative priming effects typically decrease with increasing SOA (Hermans, De Houwer, & Eelen, 2001; Klauer, Roßnagel, & Musch, 1997). Therefore, presenting brief primes that still reliably convey evaluative information is crucial. Thus, because of the supposedly time-bound character of natural emotional sounds, musical primes were often employed in auditory-visual priming studies, as consonant and dissonant chords are assumed to transmit evaluative meaning with short exposure duration (Marin et al., 2012; Sollberger et al., 2003; Steinbeis & Koelsch, 2011). As an alternative solution, Scherer and Larsen

(2011) presented natural emotional sound segments as primes in their cross-modal evaluative priming study. Even though their primes had a considerably long duration (1 s), the authors introduced each sound in its full multi-second length before the priming task. They argued that the snippets would have only “reminded” participants of the previously introduced sounds.

In the present section, we investigate the interplay of the modality-specific factor of gradually unfolding sound information with affective evaluations. The main goal targeted in the present section is to explore the boundaries for evaluation of natural emotional sounds concerning the available time and the intentionality of the evaluation process. More specifically, our aim in the present section is twofold: First, to examine the boundary of the sound exposure that is needed for reliable explicit evaluative judgments on natural emotional sounds (Experiment 1); second, to investigate the question of whether natural emotional sounds can be relatively automatically – in the sense of rapidly and unintentionally – evaluated in an RT-based paradigm in which speeded responses are required to the presented sounds (Experiment 2). Thus, in contrast to the above-described approach presenting musical chords, in our first two studies we employ complex, natural emotional sounds.

Experiment 1: Rapid Explicit Evaluation of Emotional Sounds

In the present chapter, we examine the boundary of sound exposure that is needed for reliable explicit evaluative judgments, thus under the condition in which explicit instructions are given for evaluation. Additionally, we investigate the question whether evaluations that are based on brief exposure time can be mediated by semantic identification of the sound source or content, as opposed to the view that early evaluations emerge from mere combinations of non-symbolic, low-level sound features, just as loudness or spectral entropy contour (see, e.g., the approach of Weninger et al., 2013, for identifying low-level acoustic features that contribute to sound evaluations).

In Experiment 1, we presented brief (200, 400, and 600 ms long) segments of natural emotional sounds sampled from the International Affective Digitized Sounds battery (IADS; Bradley & Lang, 2007). The IADS battery includes language-independent natural emotional sounds across a wide range of semantic categories, like environmental sounds (e.g., bird singing, office noise, jackhammer), sounds produced by humans (e.g., walking, laughing, vomiting), and sounds covering scenes (e.g., party, attack, car wreck). In a first sample of participants, we collected valence ratings on the brief segments of emotional sounds, and we tested whether the ratings of these segments mirror the valence ratings of the corresponding full-duration sounds (with approximately six seconds duration) based on the normative sample reported by Bradley and Lang (2007) and an own native German sample (see below). With a second sample of participants, we investigated whether semantic identification can occur based on these brief segments of emotional sounds (which would encompass the possibility that early evaluations can be based on semantic processing). We used two measures of sound identification: (1) a rather coarse-grained identification of the sound source requiring the participants to differentiate whether a sound was produced by an animate or an inanimate agent; and (2) a more fine-grained identification of the sound regarding its content and source. Additionally, the second sample provided valence ratings on the full-length emotional sounds.

METHODS

Participants

Sample 1 and 2 each had 30 participants (undergraduate students from Saarland University) who participated for monetary compensation (Sample 1: 22 females, aged 18–32 years, *Mdn* = 23 years; Sample 2: 19 females, aged 19–31 years, *Mdn* = 24.5 years). The sample size was determined by considerations about the reliability of mean ratings (see below).

Materials

We selected 39 *positive* (e.g., applause, slot machine, bird singing), 39 *negative* (e.g., vomiting, attack, car wreck), and 39 *neutral* (e.g., office noise, walking, yawn) natural sounds from the IADS battery (Bradley & Lang, 2007). Our selection aimed to (a) incorporate a wide range of different natural sounds concerning their source and content; and to (b) maximize the differences between the positive, negative, and neutral stimulus pools on normative valence ratings, thereby creating stimulus pools with non-overlapping rating ranges. A further selection criterion (c) aimed to minimize silent periods in the 0-600 ms excerpts of the sounds. Mean normative valence ratings of the full-length stimuli on a 9-point scale ranging from very unpleasant (1) to very pleasant (9) were *M* = 7.02 (*SD* = 0.43; in the range of 6.31–7.90) for positive, *M* = 4.75 (*SD* = 0.40; in the range of 4.01–5.35) for neutral, and *M* = 2.41 scores (*SD* = 0.48; in the range of 1.57–3.08) for negative sounds, respectively. We coded the sounds as produced by animate or inanimate agents (21, 16, and 24 animate and 18, 23, and 15 inanimate sounds in positive, neutral and negative conditions, respectively). From each sound, we created three new sound files by extracting the 0-200 ms, 0-400 ms, and 0-600 ms segments. Sounds were organized into three stimulus sets. Each set contained the 39 positive, 39 negative, and 39 neutral sounds, with one-third of each valence category selected in the 200, 400, and 600 ms version, respectively. Thereby each set contained one version of all available sounds. Each participant received one of the three stimulus sets in a balanced design, and across the final sample, we thus collected ten rating scores for each version of each sound. Previous studies conducted in our lab suggest that an aggregate of ten ratings secures high reliability of the aggregate measure.

Procedure

Each participant received 117 trials, featuring the sounds of one of the stimulus sets. Auditory stimuli were presented binaurally via headphones (HD-212 Pro, Sennheiser, Wedemark, Germany) in a comfortable loudness of approximately 70 dB(A)[2]. Trials were presented in an individually randomized order.

For Sample 1, each trial started with the presentation of a rating screen featuring a nine-point scale ranging from very unpleasant (-4) to very pleasant (+4), with zero as the neutral point. After 500 ms, a sound segment was presented. Participants were asked to rate the pleasantness of each sound by clicking on one of the nine scale points. The next trial started immediately after the response was registered.

For Sample 2, each trial started with the presentation of a fixation cross without auditory stimuli. After 500 ms, a sound segment was played. Thereafter, participants were asked to accomplish two tasks. First, participants had to categorize the presented sound according to whether it was produced by (an) animate or inanimate agent(s) as the direct source of the sound by clicking to the corresponding category label. For instance, a person or an animal was considered as an animate agent, while musical instruments, tools, or natural phenomena (e.g., thunder) were considered as inanimate agents. Second, participants had to identify the presented sound by describ-

[2] The sound pressure level was adjusted to the perception characteristics of human hearing concerning different tone-frequencies by applying a commonly used filter (A-weighting; B. C. J. Moore, 2012).

ing it in their own words. Participants were asked to type a one or two words long answer that ideally refers to both the sound source and the "nature" or content of the sound (e.g., "woman screams"). Participants were also encouraged to cover a complex situation by using only one word if it was apposite (e.g., "party"). The next trial started immediately after pressing the Enter key. Additionally, participants of Sample 2 were asked to perform a valence rating task on the *full-length* stimuli at the end of the session. The procedure was identical to the procedure of Sample 1, but importantly, on each trial the emotional sounds were now played in their full-length version (6 s).

Design

We applied a 3 × 3 mixed factorial design on the valence ratings and on the two measures of sound identification with the a priori valence category (positive vs. neutral vs. negative) as the grouping factor and the duration of the sound segment (200 ms vs. 400 ms vs. 600 ms) as the repeated measures factor.

RESULTS

A significance level of $\alpha = .05$ (two-tailed) was adopted for all inferential statistics. All analyses are based on items as the units of analyses with values aggregated across participants.

Valence Ratings

Valence ratings were transformed to a 1-9 scale to stay in line with the normative ratings provided by Bradley and Lang (2007). Results of the valence ratings are presented in *Table 1*. First of all, it can be seen in the upper part of the table that the full-length rating provided by a German sample closely resembles the norm rating provided by Bradley and Lang (2007). Thus, there seem to be no important cultural differences between the two samples. Intraclass-correlations indicated that interrater-agreement was high for all ratings; however, it was considerably lower for the 200 ms rating than for the 400 and 600 ms ratings.

Table 1. Mean valence ratings, intraclass-correlations (ICC), and correlations with the norm rating (r_n) and with the full-length rating based on a native German sample (r_G) for the three duration conditions of Experiment 1. Mean valence ratings are also provided for the full-length stimuli based on the normative (Full-Length$_n$) and the native German sample (Full-Length$_G$). Valence ratings range from very unpleasant (1) to very pleasant (9); *SD* in parentheses.

	Valence Category					
	Negative	**Neutral**	**Positive**	ICC[a]	r_n	r_G
Full-Length$_n$	2.41 (0.48)	4.75 (0.40)	7.02 (0.43)			
Full-Length$_G$	2.36 (0.57)	4.49 (0.98)	6.39 (0.96)	.97	.93	
200 ms	3.73 (1.05)	4.59 (0.81)	5.32 (1.01)	.86	.58	.60
400 ms	3.05 (1.07)	4.56 (0.89)	5.99 (1.21)	.92	.77	.77
600 ms	3.03 (1.16)	4.71 (1.06)	6.32 (1.25)	.92	.78	.81

[a] Average intraclass-correlation for random raters (ICC[1, 10]) according to Shrout and Fleiss (1979).

The pattern of means of the three valence conditions clearly reflects a differentiation into positive, neutral, and negative evaluations, and an increasing difference between the means of positive and negative ratings with longer sound durations (see *Table 1*). A 3 (valence: positive vs. neutral vs. negative) × 3 (duration: 200 ms vs. 400 ms vs. 600 ms) MANOVA for repeated measures with duration as a within-items factor and valence as a grouping factor on the valence ratings yielded a main effect of valence, $F(2,114) = 73.28$, $p < .001$, $\eta_p^2 = .562$, that was moderated by the sound duration, $F(4,228) = 16.30$, $p < .001$, $\eta_p^2 = .222$. There was no significant main effect of duration, $F(2,113) = 2.12$, $p = .125$, $\eta_p^2 = .036$. To test valence differentiation in the different duration conditions, separate ANOVAs were performed for each duration condition with valence (positive vs. neutral vs. negative) as grouping factor. These analyses showed significant valence effects for all three durations, $F(2,114) = 26.60$, $p < .001$, $\eta_p^2 = .318$, for the 200 ms duration; $F(2,114) = 74.77$, $p < .001$, $\eta_p^2 = .567$ for the 400 ms duration; and $F(2,114) = 78.87$, $p < .001$, $\eta_p^2 = .580$ for the 600 ms duration. To understand the interaction pattern that emerged in the analysis on the full design, we tested the increase in valence differentiation for the two duration transitions (i.e., the transition from 200 to 400 ms and the transition from 400 to 600 ms): The first 3 (valence: positive vs. neutral vs. negative) × 2 (duration: 200 ms vs. 400 ms) planned interaction contrast was significant, $F(2,114) = 34.17$, $p < .001$, $\eta_p^2 = .375$, thereby signaling a gain in differentiation by using 400 ms excerpts compared with 200 ms snippets. The second 3 (valence: positive vs. neutral vs. negative) × 2 (duration: 400 ms vs. 600 ms) planned interaction contrast did not show significant differences, $F(2,114) = 1.70$, $p = .187$, $\eta_p^2 = .029$, thereby indicating that gain in differentiation by using 600 ms excerpts compared to those of 400 ms length is modest. The difference between the 200 ms condition on the one hand and the 400 ms and 600 ms conditions on the other hand can be seen additionally in the correlation coefficients of the ratings for the brief segments with the full-length ratings (see *Table 1*).[3]

Additionally, we carried out analyses focusing on duration effects within the a priori valence categories. We found a significant duration effect both within the positive and negative valence category, $Fs(2,37) > 21.27$, $p < .001$, $\eta_p^2 > .534$ ($F < 1$, *n.s*,

[3] All correlations are associated with $p < .001$. However, due to the multimodal distribution of the norm ratings, inferential statistics might be biased. Thus, the correlations should be dominantly taken as a descriptive index of the correspondence between brief segment ratings and the full-length ratings.

within the neutral category). Within the two valenced categories, we found significant gain in differentiation by using 400 ms excerpts compared with 200 ms excerpts, $Fs(1,38) > 22.40$, $p < .001$, $\eta_p^2 > .370$. However, gain in differentiation by using 600 ms excerpts compared to those of 400 ms length was modest, $F(1,38) = 4.06$, $p = .051$, $\eta_p^2 = .096$ within the positive category; and $F < 1$, *n.s.*, within the negative category.

There are two more sources of evidence to evaluate the validity of the ratings. First, since our selection was category-focused (i.e., we a priori selected positive, neutral, and negative sounds such that the norm-rating distributions of the three samples were non-overlapping), we attempted to predict category membership on the basis of the 200, 400, and 600 ms ratings, respectively, using multinomial logistic regression. Corresponding to the results reported above, even the 200 ms rating significantly improved prediction in comparison to random assignment, $\chi^2(2) = 45.22$, $p < .001$ ($\chi^2[2] \geq 92.72$, $p < .001$ for 400 and 600 ms). However, while classification accuracy (ACC) was 71.8% for the 600 ms rating, with only 1.7% severe misclassifications (i.e., classification of a positive sound as negative and vice versa), predictions based on the 200 ms rating were considerably weaker: Classification ACC was 59.0%, with 7.7% severe misclassifications (for 400 ms, classification ACC was 74.4%, with 6.0% severe misclassifications).

Second, standard deviations of mean norm ratings of the IADS are available. These can be considered as an index of the relative ambiguity of evaluation. Thus, a new valid rating should be sensitive to this ambiguity; thereby it should be more predictive of the original norm ratings for less ambiguous sounds and less predictive for more ambiguous sounds. In statistical terms, we can assume the interaction term of a new valid rating and the ambiguity index to be significant when predicting the norm rating. This holds true for both the 400 ms rating, $B = -0.12$, $t(116) = 2.07$, $p = .040$, for the product term, and the 600 ms rating, $B = -0.15$, $t(116) = -2.73$, $p = .007$, but not for the 200 ms condition, $B = -0.10$, $t(116) = -1.36$, $p = .178$.

Semantic Identification

The results of the two measures on the semantic identification of sounds are presented in *Table 2* and below.

Sound source categorization

Participants were able to differentiate whether emotional sound segments were produced by animate or inanimate agents with remarkable precision (see *Table 2*).

Table 2. Mean ACC (%) for sound source categorization and mean ACC for specific sound identification (ranging from fully incorrect [0] to fully correct [4]) in the three duration conditions of Experiment 1; *SD* in parentheses.

	Valence Category		
	Negative	**Neutral**	**Positive**
1. Sound source categorization			
200 ms	84.1 (21.2)	74.9 (31.0)	83.6 (23.7)
400 ms	91.8 (15.5)	77.7 (29.7)	90.8 (20.6)
600 ms	92.8 (13.4)	78.5 (31.2)	90.5 (17.2)
2. Specific sound identification			
200 ms	1.49 (1.19)	0.80 (1.04)	1.99 (1.19)
400 ms	2.17 (1.06)	1.46 (1.24)	2.64 (1.24)
600 ms	2.36 (1.09)	1.70 (1.22)	2.86 (1.09)

A 3 (valence: positive vs. neutral vs. negative) × 3 (duration: 200 ms vs. 400 ms vs. 600 ms) MANOVA[4] for repeated measures on sound source categorization ACC yielded a main effect of duration, $F(2,113) = 6.67$, $p = .002$, $\eta_p^2 = .105$, with a significant difference between the 200 and 400 ms conditions, $F(1,114) = 11.46$, $p = .001$, $\eta_p^2 = .091$, but with no significant difference between the 400 and 600 ms conditions, $F < 1$, *n.s*; and a significant valence effect, $F(2,114) = 4.07$, $p = .020$, $\eta_p^2 = .067$ which emerged due to lower ACC for neutral sounds compared with valent sounds, $F(1,115) = 8.14$, $p = .005$, $\eta_p^2 = .066$ ($F < 1$, *n.s*, for positive versus negative sounds). Duration and valence did not interact significantly, $F < 1$, *n.s.* Note that these MANOVA results have to be taken with some caution because of the skewness of the sound source identification scores.

Specific sound identification

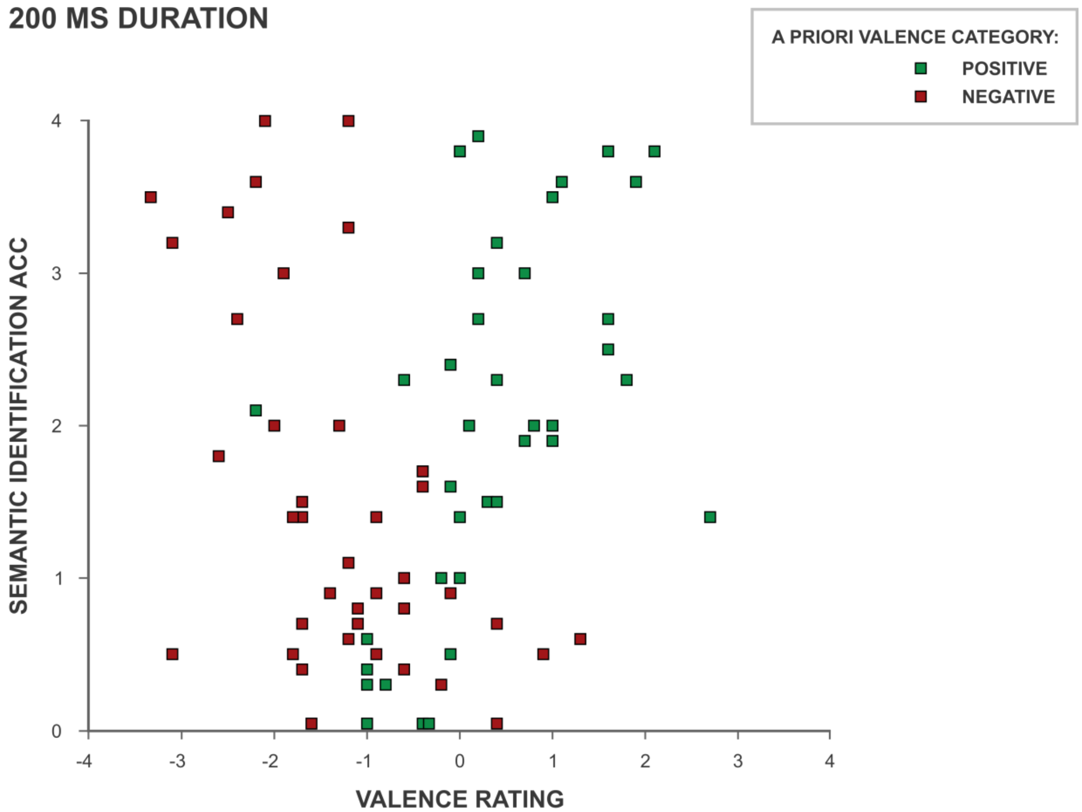
Two native German raters assessed the correctness of the specific sound identifications by scoring the answers as "correct", "partly correct", or "incorrect". The label "partly correct" was applied to the situations when either the sound source or the content of the sound was not or was incorrectly described (e.g., "women" instead of "woman screams"). The interrater-agreement between the two raters was good in all duration conditions as shown by high intraclass-correlations, ICCs > .93. We aggregated the sums of the evaluations of the two raters; thus, the procedure resulted in a five-point accuracy measurement (0-4; ranging from "judged as incorrect by both raters" to "judged as correct by both raters").

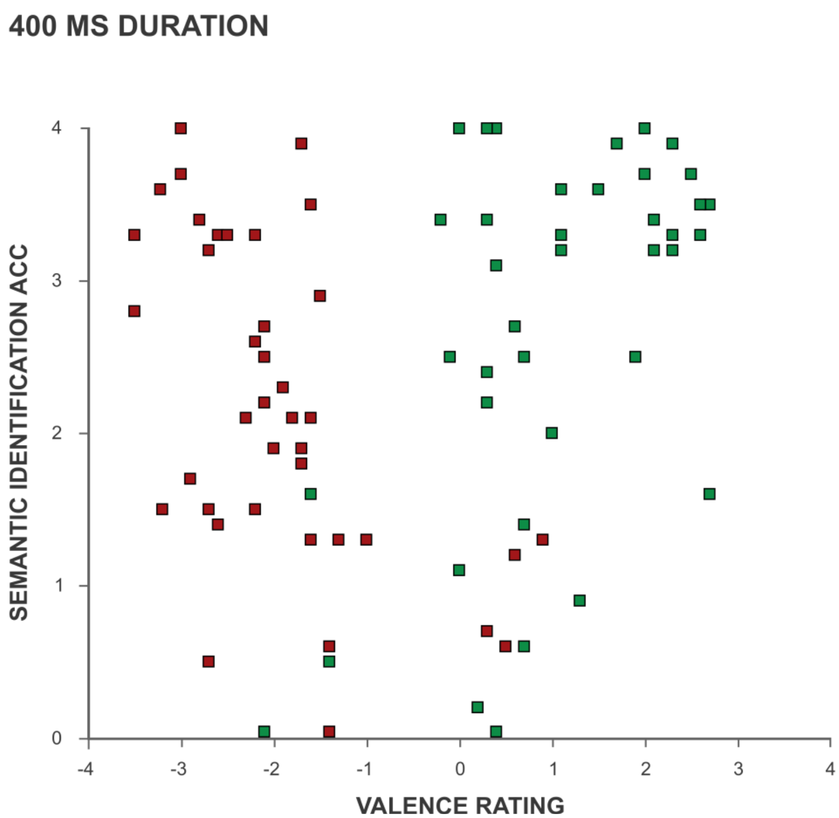
[4] Alternatively, we conducted a 3 (valence) × 2 (animacy category: animate vs. inanimate) × 3 (duration) MANOVA. All effects reported below are essentially the same in this analysis. Additionally, there were significant effects involving animacy: A significant animacy main effect, $F(1,111) = 9.96$, $p = .002$, $\eta_p^2 = .082$; a significant animacy × duration interaction, $F(2,110) = 3.81$, $p = .025$, $\eta_p^2 = .065$; and a significant animacy × valence interaction, $F(2,111) = 4.95$, $p = .009$, $\eta_p^2 = .082$. However, for the sake of succinctness and because these effects are rather uninteresting due to their ambiguity (i.e., they might be an effect of better discriminability of one category relative to the other or they might reflect a response bias), we report only the reduced analysis.

Table 2 presents the results of the specific sound identification ACC. A 3 (valence: positive vs. neutral vs. negative) × 3 (duration: 200 ms vs. 400 ms vs. 600 ms) MANOVA for repeated measures on sound identification ACC yielded a main effect of duration, $F(2,113) = 47.78$, $p < .001$, $\eta_p^2 = .458$, with significant differences between the 200 and 400 ms conditions, $F(1,114) = 65.92$, $p < .001$, $\eta_p^2 = .366$, and between the 400 and 600 ms conditions, $F(1,114) = 7.98$, $p = .006$, $\eta_p^2 = .065$. Additionally, a significant valence effect was found, $F(2,114) = 12.82$, $p < .001$, $\eta_p^2 = .184$, with significantly lower ACC for neutral sounds compared with valent sounds, $F(1,115) = 20.73$, $p < .001$, $\eta_p^2 = .153$; and $F(1,76) = 4.41$, $p = .039$, $\eta_p^2 = .055$ for positive versus negative sounds. Duration and valence did not interact, $F < 1$, *n.s.*

Evaluation and Semantic Identification

To obtain evidence for co-processing of semantic and evaluative features (which would encompass the possibility that semantic processing is a precondition of evaluation), we employed the following logic: If semantic processing would be a necessary precondition of evaluation (or they would occur strictly parallel), non-neutral sounds that are not identifiable for a given duration condition should be rated as neutral; sounds that are clearly identifiable should have received a marked mean valence ratings – either a positive one for positive sounds or a negative one for negative sounds. Finally, (non-neutral) sounds with a medium accuracy score (i.e., sounds that were identified only by some raters) should have received moderate mean valence ratings as a result of some marked ratings (for those who identified) and some neutral ratings (for those who did not identify). Plotting specific identification scores (on the Y-axis) against the ratings (on the X-axis) should therefore yield a parabola-shaped scatterplot (see the actual distribution on *Figure 1*).


200 MS DURATION
A PRIORI VALENCE CATEGORY:
POSITIVE
NEGATIVE
SEMANTIC IDENTIFICATION ACC
VALENCE RATING
4
3
2
1
0
-4
-3
-2
-1
0
1
2
3
4


400 MS DURATION
SEMANTIC IDENTIFICATION ACC
VALENCE RATING
4
3
2
1
0
-4
-3
-2
-1
0
1
2
3
4

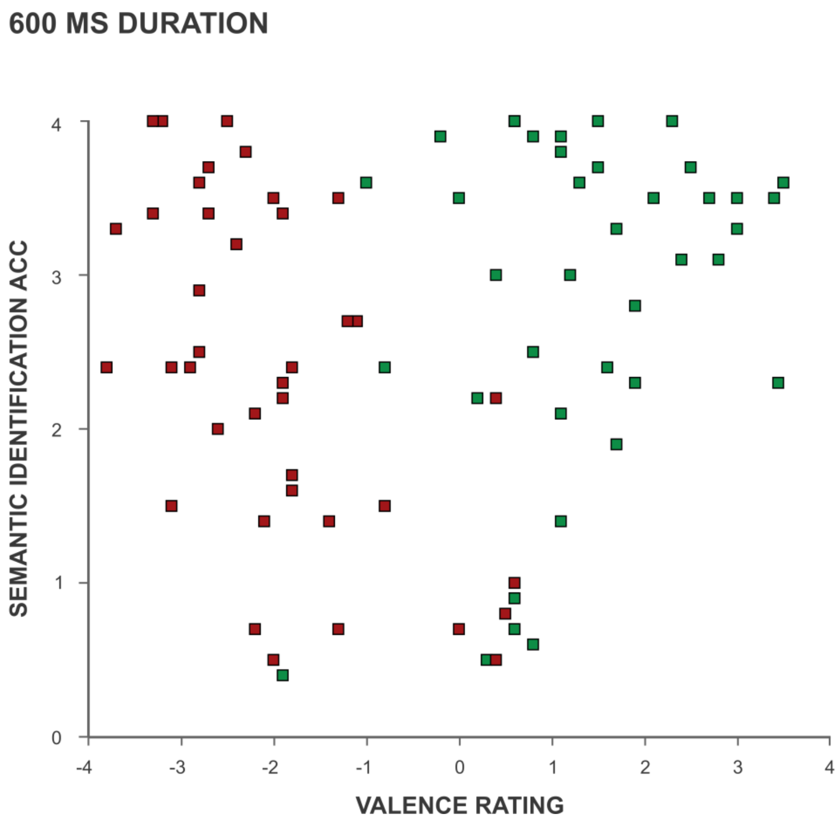


Figure 1. Scatter plots for specific sound identification accuracy versus valence rating in the three duration conditions of Experiment 1.

Specific sound identification accuracy ranges from fully incorrect (0) to fully correct (4); valence rating ranges from very unpleasant (-4) to very pleasant (+4). A priori defined positive valence category is represented by green boxes; a priori defined negative valence category is represented by red boxes.

Moreover, if the original ratings (scaled from -4 to +4) will be used, the parabola should have its vertex at x = 0. Therefore, we regressed the specific identification scores of positive and negative sounds on the quadratic term of the original ratings only (i.e., we left out the first-order term) which forced the regression algorithm to fit a parabola with vertex x=0 to the data. Note that this is a rather strong constraint. The quadratic relationship was significant for all durations, $B = 0.28$, $t(76) = 2.52$, $p = .014$ for 200 ms, $B = 0.25$, $t(76) = 2.28$, $p = .026$ for 400 ms, $B = 0.30$,

$t(76) = 2.70$, $p = .009$ for 600 ms. The same kind of analysis using the source categorization scores instead of the specific identification scores yielded non-significant results. However, this is probably due to the skewness of distributions.

DISCUSSION

Results of Experiment 1 demonstrate clearly that valence can be extracted from very brief (i.e., a few hundred milliseconds long) segments of natural emotional sounds. Even valence ratings for durations as short as 200 ms are still reliable, although they are slightly more ambiguous compared with ratings of the 400 and 600 ms segments. Evidence for these claims was derived from several sources. First, valence ratings of brief sound segments showed a clear differentiation between positive, neutral, and negative valence categories, which were defined a priori according to the norm ratings. Although ratings reflected significant valence differentiation in each duration condition, a significant interaction emerged between duration and valence: ratings showed clearer valence differentiation as exposure duration increased, with the largest increase in differentiation at the 200-400 ms transition. Second, evaluation of 200, 400, and 600 ms segments showed a close relationship with the normative valence ratings of the full duration sounds. While the 600 ms rating (and with some slight limitation the 400 ms rating) seemed to behave almost like a re-rating of the full-length stimuli, the 200 ms rating was more equivocal. The inconsistency of the 200 ms rating was reflected in a lower correlation with the full-length ratings, a lower interrater-agreement, and less sensitivity to the ambiguity in the norm rating compared with the longer duration conditions.

Furthermore, we raised the question whether it is possible to extract complex semantic meaning during a few hundred milliseconds of presentation time, thus, whether early evaluations can be driven by semantic meaning. We found that (1) participants could differentiate sounds produced by animate and inanimate agents with high precision; and (2) participants could identify the specific sounds still reliably, although with less precision. As expected, both the rather coarse-grained and the more specific index of sound identification showed higher precision as exposure duration increased, with the greatest increase in precision at the 200-400 ms transition. The more specific index of sound identification showed a close relationship with the evaluation of the sound fragments in all duration conditions, suggesting that sound identification could occur before or parallel with the early evaluations. Thus, it is possible that the demonstrated fast evaluations are, at least partly, based on semantic processing; however, low-level acoustic features could also contribute to the evaluation of natural sounds (see, e.g., Weninger et al., 2013).

Results of Experiment 1 suggest that valence is evaluated in a similar way when a standard natural emotional sound of several seconds is available or when there is only a short snippet of sound to base the judgment on. While 400 and 600 ms segments were evaluated highly reliably (i.e., they appeared to be comparable to a re-rating of the full-length stimuli), 200 ms segments were evaluated still reliably but relatively more inconsistently compared with the longer durations. This result thus suggests that 200 ms long exposure time is sufficient at least for partial evaluation of natural sounds under the condition of explicit evaluation. Taken together, our findings lend support to the notion that natural emotional sounds can be evaluated rapidly despite their complex temporal structure, at least in an intentional way, and these fast evaluations can be mediated by early semantic identification of the sounds.

Experiment 2: Unintentional Evaluation of Emotional Sounds

Experiment 1 provided evidence that valence information of complex, natural emotional sounds can be extracted rapidly, that is, even if only a few hundred ms long segment is available to base the judgment on. However, rapid evaluations occurred under the circumstances when explicit instructions were given for evaluation. Hence, the question remained open whether natural emotional sounds can be evaluated rapidly *and* unintentionally, thus without (conscious) intention. Based on the results of Experiment 1, in Experiment 2 we introduce an auditory version of the affective Simon task as an indirect RT-based paradigm to assess evaluations of ecologically valid, complex, natural emotional sounds. If in this paradigm, in which valence is irrelevant concerning the main task and participants are instructed explicitly to ignore it, affective valence of the sounds influences speeded behavioral responses to a task-relevant feature, it can be concluded that sound valence was evaluated rapidly and unintentionally, thus relatively automatically (see, e.g., Moors, 2015).

THE (AFFECTIVE) SIMON PARADIGM

The original spatial (non-affective) Simon task was introduced as a powerful irrelevant feature paradigm requiring participants to give spatial responses (e.g., to press right or left key) to a non-spatial stimulus feature (e.g., color of a light flash) while as a task-irrelevant feature, the stimulus presentation was lateralized (e.g., Craft & Simon, 1970). Although the task-relevant feature was not contingent on the spatial location of the stimulus presentation, faster responses were typically observed when the spatial features of the stimulus presentation and response were matching (e.g., left key press for a light flash presented on the left side) compared with a situation when they mismatched (e.g., left key press for a light flash presented on the right side). In a more abstract level, the Simon paradigm can be decomposed into three factors: (1) the task-relevant feature that determines the correctness of the response (e.g., color); (2) the orthogonally varied task-irrelevant (and often to-be-ignored) stimulus feature (e.g., spatial location of the stimulus presentation); and (3) the response feature (e.g., spatial location of the response key) that is meaningfully related to the task-irrelevant stimulus feature (e.g., they both incorporate spatial information) but not related meaningfully to the task-relevant stimulus feature (here, color; see also De Houwer & Eelen, 1998). A further important characteristic of the paradigm is that speeded responses are required to the task-relevant feature. The question of interest is whether the irrelevant feature influences the speed (and/or accuracy) of behavioral responses in this paradigm. If so, one can conclude that (a) the irrelevant feature was processed; (b) it was processed relatively fast as speeded responses were required; (c) it was processed presumably unintentionally as processing of the task-relevant feature was not beneficial to the task performance and, moreover, participants are often explicitly instructed to ignore it.

The Simon paradigm has contributed not only to the understanding of spatial cognition, but its flexibility opened the road for introducing affective variants in which the irrelevant feature and response are affectively related instead of a spatial relation (see, e.g., De Houwer & Eelen, 1998). In the affective Simon task (AST), participants are presented with valenced stimuli (such as positive and negative words) and have to give intrinsically valent responses (e.g., saying “good” or “bad”) as fast as possible to a valence-neutral stimulus feature (e.g., the word’s color) that is orthogonally varied to valence. Thus, importantly, valence information is completely irrelevant regarding the main task, and not predictive for the task-relevant feature. Moreover, participants are instructed to give speeded responses only to the task-relevant manipulation (e.g., react to the word’s color), while they are often asked explicitly to ignore every other feature aside from the task-relevant modulation. The affective Simon (AS) effect refers to the phenomenon that RTs are typically shorter (and responses are more accurate) for congruent (stimulus valence and response valence match) pairings compared with incongruent (stimulus valence and response valence mismatch) trials, thus, behavioral reactions are influenced by the to-be-ignored affective meaning of the stimulus. The characteristics of the paradigm support the interpretation that the AS effect emerges as a result of automatic processing of stimulus valence in the sense of involuntary and fast evaluation.

The (affective) Simon paradigm provides great flexibility along several aspects: AS effects have been demonstrated by employing a wide variety of stimuli (e.g., written words, schematic faces, simple stimuli associated with valence), task-relevant stimulus features (e.g., grammatical category, color), and responses (e.g., by uttering valence category labels or affectively connoted words, or moving a manikin on the screen towards or away from the stimulus; De Houwer & Eelen, 1998; Moors & De Houwer, 2001; Voß, Rothermund, & Wentura, 2003). This flexibility makes the paradigm an ideal candidate to investigate automatic valence evaluations in the auditory modality.

THE AUDITORY AFFECTIVE SIMON PARADIGM

In the auditory version of the AST, we define the relevant factors of the paradigm in the following way: (1) We employ a purely perceptual stimulus feature as the relevant valence-neutral dimension that determines the correct response: Participants are required to classify the direction of an illusory movement of the sound source from a central position toward the right or left side. This task can be performed successfully without intentional processing of the affective meaning of the sounds. We present natural sounds as stimuli (2) with well-defined a priori valence as a task-irrelevant feature; and (3) we employ valence-related category labels as response categories:

Participants are required to utter “good” or “bad” depending on the direction of the illusory movement. We analyze RTs and error rates (ERs) as a function of stimulus and response valence congruency.

Based on the results of Experiment 1, we can assume that valence information can be extracted successfully after a few hundred milliseconds of sound exposure. However, in contrast to Experiment 1 (i.e., presenting brief snippets of sounds that were explicitly evaluated), in Experiment 2, we apply another approach to investigate rapid sound evaluations: We present full-length auditory stimuli that require fast responses to a task-relevant feature (illusory movement direction). From the onset time of the task-relevant feature we can coarsely estimate the time of valence exposure. Taken into consideration the relative ambiguity of the 200 ms ratings in Experiment 1, we employ two parallel versions of Experiment 2: In Experiment 2a, we make the task-relevant feature (i.e., onset of virtual movement) available at 500 ms post sound onset (i.e., 500 ms feature start onset asynchrony; FSOA). This means that participants are exposed to the valence-relevant content slightly earlier than to the task-relevant information. In Experiment 2b, we use a synchronous version, that is, the task-relevant virtual movement starts at the onset of the sound (0 ms FSOA). If sounds are evaluated in this paradigm (when valence of the sounds is a task-irrelevant and to-be-ignored feature), we can expect shorter (and/or more accurate) responses for congruent (sound valence and response valence match) pairings compared with incongruent (sound valence and response valence mismatch) pairings.

METHODS

Participants

In Experiment 2a, 57 students from Saarland University (39 females; aged 18-36 years, *Mdn* = 25 years; 4 left-handers) participated for monetary compensation. The data of four further participants were discarded because of extreme ERs (≥ 16.7%; i.e., far out values according to Tukey, 1977). In Experiment 2b, 52 students from Saarland University (30 females; aged 19-33 years, *Mdn* = 23 years; 6 left-handers) participated for monetary compensation. Given a sample size of $N = 57$ in Experiment 2a (52 in Experiment 2b) and an α-value of .05 (two-tailed), effects size of $d_z = 0.38$ ($d_z = 0.40$ in Experiment 2b) can be detected with a probability of $1 - \beta = .80$ (calculated with the aid of G*Power 3 software; Faul, Erdfelder, Lang, & Buchner, 2007).

Materials

20 *positive*, 20 *negative*, and 20 *neutral* sounds from the IADS battery were presented via headphones **(**HD-600, Sennheiser, Wedemark, Germany**)** with a maximal loudness of approximately 70 dB(A). Mean normative valence ratings – on a nine-point scale ranging from most unpleasant (1) to most pleasant (9) – were $M = 6.94$ ($SD = 0.51$) for positive, $M = 4.62$ ($SD = 0.52$) for neutral, and $M = 2.48$ ($SD = 0.54$) for negative sounds, respectively (Bradley & Lang, 2007). Additionally, four positive, four negative and four neutral IADS sounds were used in practice trials. To invoke the virtual sound movement, amplitude of the sounds was modulated in the following way: Starting at 500 ms post onset (Experiment 2a) or starting at sound onset (Experiment 2b), intensity in one auditory signal channel of the stereo sound was reduced linearly over a 1000 ms interval by a total of 75%. We created two versions of each sound, one with an illusory movement from a central position toward the right side of the perceiver ("moving to the right" sounds) and one with an illusory movement to the left ("moving to the left" sounds; see, e.g., Rosenblum, Carello, & Pastore, 1987, for the role of amplitude change in locating moving sound sources).

Procedure

All participants were tested individually. Before the experiment, instructions emphasized that participants should attend only to the illusory movement of the sounds and ignore any other stimulus features. The experiment started with 12 practice trials. During the practice phase, participants received visual accuracy feedback after every trial. The experimental phase comprised 60 experimental trials, with 20 trials featuring positive, 20 trials featuring negative, and 20 trials featuring neutral sounds in an individually randomized order. Half of the sounds were presented in "moving to the left" and half of the sounds in "moving to the right" version, respectively, in a random order. Assignment of specific sounds to right and left moving versions was counterbalanced across participants.

An experimental trial started with the presentation of a fixation cross without auditory stimuli. (The fixation cross remained on the screen until the end of the trial.) After 1000 ms, a positive, negative, or neutral sound was played. Participants' task was to categorize the direction of the virtual movement by uttering "good" or "bad" ("gut" and "böse" in German, respectively) as quickly and accurately as possible. The assignment of the correct response (saying "good" or "bad") to the illusory movement direction (right or left) was counterbalanced between participants. While a voice key apparatus recorded RT (i.e., the onset of the utterance), the response category was registered online by the experimenter, who was sitting in front of a second screen next to the participant but separated by a partition wall. That is, the experimenter pressed one key for response "good" and one key for response "bad"; if the voice key was triggered accidentally (e.g., by misutterances or by noises like coughing), a third key was used. After the vocal response was detected by the voice key, the auditory stimulus was terminated.

Design

We employed a 2 (sound valence: positive vs. negative) × 2 (response valence: positive vs. negative) repeated measures design, which reduces to a simple one-factorial congruency (congruent vs. incongruent) design. Neutral sounds were added to obtain a baseline measure against which to assess the effects of congruency (i.e., to obtain rough estimates of "costs" and "benefits").

RESULTS

A significance level of $\alpha = .05$ (two-tailed) was adopted for all analyses. RTs were calculated from the onset of the illusory movement (i.e., 500 ms post sound onset for Experiment 2a, at sound onset for Experiment 2b). RT analyses were restricted to trials with correct responses and error-free response recording (3.0% of the trials for both experiments were excluded because of incorrect or erroneous responses or non-reaction of the voice key). As an a priori criterion, RTs below 300 ms and above 2000 ms were discarded from further analyses (2.4% and 2.9% of the trials in Experiment 2a and 2b, respectively). *Table 3* shows the mean RTs and ERs for the congruent and incongruent conditions, and for neutral sounds in Experiment 2a and 2b.

For Experiment 2a, the RT difference between congruent and incongruent trials was significant, $t(56) = 2.50$, $p = .015$, $d_z = 0.33$. The effect seems to be due mainly to the costs associated with the incongruent pairings: The mean RT for neutral sounds was almost identical to the RT for congruent trials, $|t| < 1$, *n.s.*, for the neutral vs. congruent difference; but the difference between incongruent and neutral conditions was significant, $t(56) = 2.65$, $p = .011$, $d_z = 0.35$. Similar analyses on the ERs did not show any significant differences, all $|t|$s < 1, *n.s.*

For Experiment 2b, the incongruent-congruent RT difference was numerically in the expected direction but fell short of significance, $t(51) = 0.66$, $p = .512$, $d_z = 0.09$. Neutral RTs were numerically faster than congruent RTs, but this difference was not significant, $t(51) = -1.12$, $p = .267$, $d_z = -0.16$. The difference between incongruent and neutral conditions was significant, $t(51) = 2.16$, $p = .035$, $d_z = 0.30$. Similar analyses on the ERs did not show any significant differences, all *t*s < 1, *n.s.*

Table 3. Mean RTs (in ms) and mean ERs (%) as a function of stimulus and response valence congruency in Experiment 2; *SD* in parentheses.

	Experiment 2a (500 ms FSOA)[a]		**Experiment 2b** (0 ms FSOA)[a]	
	RT	**ER**	**RT**	**ER**
Neutral	1020 (229)	2.8 (4.4)	1009 (241)	2.6 (4.1)
Congruent	1017 (231)	3.0 (4.0)	1021 (219)	3.0 (3.9)
Incongruent	1040 (228)	3.3 (4.5)	1027 (219)	3.5 (5.4)
AS Effect[a]	24 [9]	0.3 [0.7]	6 [10]	0.5 [0.8]

[a] FSOA = Feature Start Onset Asynchrony

[b] incongruent-minus-congruent difference; standard errors in brackets

Discussion

In Experiment 2, we used an auditory version of the AST in two variations: In one version, the task-relevant change started after half a second of exposure to the emotional sound (Experiment 2a); in the other version, it started at sound onset (Experiment 2b). We found a significant AS effect on the RTs in Experiment 2a, indicating slower reactions when stimulus and response valence were incongruent compared with the situation when they were congruent. We have to emphasize that for successful task performance participants were not required to process the stimulus valence, as it was entirely task-irrelevant and not predictive of the task-relevant feature. Additionally, participants were explicitly instructed to ignore every characteristic of the sounds other than the task-relevant movement direction. Taken together, results of Experiment 2a support the interpretation that natural emotional sounds can be evaluated automatically, in the sense of fast and unintentional evaluation (see, e.g., Bargh, 1992; Moors, 2015).

In Experiment 2b, a similar RT pattern emerged as in Experiment 2a but fell short of statistical significance. A significant AS effect was thus found only in Experiment 2a, where exposure to the evaluative information started before the task-relevant manipulation, and not in Experiment 2b, where the onsets of the evaluative and the task-relevant information were synchronous. Thus, it appears – at least in the present paradigm – that a head start is needed for the valence information to influence behavioral responses. However, the absence of a significant AS effect in Experiment 2b was largely due to the relatively long RTs in the congruent condition (i.e., numerically longer than RTs in the neutral condition); the costs associated with the incongruent condition (relative to the neutral condition) were significant and corresponded roughly to those found in Experiment 2a. There are at least three possible explanations for this pattern of results: First, one might speculate that the intentional processing of the task-relevant feature attenuated processing of other stimulus features, including valence, when presentation onset was synchronous. Second, taking into account the results of Experiment 1, the valence information provided in the first fraction of a second may be rather ambiguous. If so, some of the congruent trials may have in fact been processed as if they were incongruent. Given that the auditory AS effect (as found in Experiment 2a) seems to arise mainly from the costs associated with incongruency, this would result in a mean RT for the congruent condition that is (at least numerically) higher than the mean RT of the neutral condition. Third, if we assume that factors that can influence the occurrence of sound evaluation are cumulative, in Experiment 1, the brief duration of exposure may have been compensated by increased intentional processing of sound valence that in turn could govern increased attentional resources. However, in Experiment 2, the lack of intentionality may have necessitated longer exposure time for sound evaluation to occur (see the argumentation of Moors, 2015). Hence, while evaluation of natural sounds emerged rapidly (i.e., at least partially already after 200 ms long exposure) when participants were explicitly instructed for evaluation, in an indirect RT-based paradigm, sound evaluation occurred unintentionally but also possibly somewhat slower. Nonetheless, a relatively short exposure time of 500 ms before the task-relevant manipulation was already sufficient for the task-irrelevant sound valence to influence behavioral responses to a task-relevant feature.

Intermediate Summary: Fast and Unintentional Evaluation of Emotional Sounds

The results of Experiment 1 and 2 support the assumption that valence information can be extracted rapidly and even in an implicit fashion from natural emotional sounds. First, explicit valence ratings revealed that valence of natural emotional sounds can be evaluated validly even if only the first few hundred milliseconds of the sounds are presented. Valence ratings on the 400 and 600 ms long segments showed a clear-cut pattern: They firmly reflected the a priori sound valence and showed a strong relationship with the valence ratings of the full-length sounds. Although ratings of natural sound segments with 200 ms duration also reflected valence reliably, they were slightly more ambiguous than the 400 and 600 ms ratings, thus suggesting that 200 ms long exposure may have allowed only partial evaluations. Despite this relative ambiguity of the 200 ms ratings, results of Experiment 1 indicate that natural sounds can convey their affective meaning already after very brief exposure time. Second, we found evidence that this early evaluation, at least partly, can be driven by rapid semantic identification of the sounds. Third, Experiment 2 demonstrated that valence of natural sounds can be evaluated implicitly in a speeded, RT-based paradigm. In the auditory version of the AST, participants responded slower if the valence of the response and the valence of the sound mismatched. We can conclude that sound valence was processed automatically in the sense of involuntariness, as the valence information was completely irrelevant regarding the main task, and the task-relevant feature was purely perceptual (i.e., did not require "deep" processing for successful task performance), and it was not contingent on the stimulus valence. Moreover, participants were asked explicitly to ignore every other feature aside from the task-relevant modulation. However, a significant AS effect emerged only when the task-relevant feature lagged behind the onset of the sound by half a second. This pattern of results might be a consequence of (a) some degree of attenuation for the task-irrelevant feature during task-relevant processing; (b) some degree of ambiguity in the valence information carried by the available brief fraction of sounds; or (c) that the lack of intentionality was compensated by longer evaluation (in line with the view of additive effects between different aspects of automaticity, Moors, 2015).

CONCLUSIONS

In sum, the results of Experiment 1 and Experiment 2 give support to the view that naturally occurring emotional sounds (e.g., environmental sounds, vocalizations, scenes) can be evaluated rapidly and even without (conscious) intention. These results emerged even despite the apparent drawback of temporally extended information conveyed by natural sounds, thereby demonstrating fast extraction of evaluative information from complex emotional sounds. The present results indicate notably rapid intentional and unintentional affective evaluation in the auditory modality.

The Next Steps: Affective Attentional Biases at an Early Level of Sound Encoding?

At this point, we can conclude that affective information of sounds can be extracted rapidly and unintentionally, in the sense that evaluative content of sounds can influence behavioral reactions of the perceiver without her (conscious) intention. To take it one step further, in the following section we will examine how affective connotation of a sound influences the processing of the sound itself before overt behavioral responses occur. The main questions of the following section can be stated as follows: Can affectively significant sounds receive preferential attention relatively automatically already at a perceptual stage of sound encoding? If so, is there an attentional bias for affectively significant sounds in general relative to neutral ones or a specific valence category is prioritized? Can these affective attentional biases operate flexibly depending on the actual affective-motivational context? In the following section, we will investigate the above outlined questions by employing ERP methodology and by adapting methods of basic auditory investigations to the field of affective research.

Preferential Attention to Valenced Tones at an Early Stage of Auditory Processing[5]

A well-established result in the field of visual affective research is that preferential processing of emotional stimuli supports rapid and efficient detection of affectively significant cues (for reviews, see, e.g., Schupp et al., 2006; Pourtois et al., 2013). Converging evidence suggests that this benefit can stem not only from the higher-order, post-perceptual stages of stimulus processing, but even from the stages of

[5] This chapter is derived in part (including Experiment 3) from the article of Folyi, Liesefeld, and Wentura published in Biological Psychology (2016), copyright by Elsevier. Available online at: http://www.sciencedirect.com/science/article/pii/S0301051115300934. Note that some analysis parameters were changed during the review process, thereby some of the here reported results are numerically different than the results reported in the article of Folyi, Liesefeld, and Wentura (2016). The general patterns of results are equivalent between the two reports. Furthermore, Experiment 4 is presented in a draft of Folyi and Wentura (manuscript in preparation).

sensory encoding (e.g., Junghöfer et al., 2006; Keil et al., 2003; Olofsson et al., 2008; Schupp et al., 2003; Stolarova, Keil, & Moratti, 2006; Vuilleumier & Huang, 2009; Vuilleumier, 2005), presumably by a rapid attentional enhancement (Keil et al., 2001; Pourtois et al., 2013; Vuilleumier, 2005; Vuilleumier & Huang, 2009; Yiend, 2010). A rapid activation of subcortical regions, in particular, the amygdala is supposed to mediate early affective effects (Phelps & LeDoux, 2005; Pourtois et al., 2013; Vuilleumier & Huang, 2009). In the present chapter, we raise the following questions: Are valenced auditory stimuli processed preferentially compared with neutral ones in a relatively automatic way? If so, at what stage(s) of the auditory processing chain does this preferential processing occur? Can sensory encoding of valenced sounds be facilitated by rapid involuntary attention?

Targeting these questions ties in with important characteristics of affective and basic cognitive research. In affective studies, the stimuli that are typically used are perceptually complex and carry strong, well-defined intrinsic valence (e.g., human emotional expressions, scenes). While such rather "coarse-grained" materials possess the advantage of ecological validity, in turn, they suffer from a lack of tight control over the physical stimulus attributes. This approach is particularly susceptible to stimulus confounds when early, perceptual stages of affective processing are investigated. By contrast, basic cognitive research into perception and attention typically employs perceptually simple, physically well-controlled stimulus materials. In the following studies, we strove for a meaningful synthesis of these two approaches: On the one hand, we aim to exert strict control over physical stimulus attributes by assigning positive and negative valence to simple tones in a learning phase (for a similar approach in the visual domain, see Müller et al., 2015; Wentura et al., 2014; for investigating differences of simple versus complex affective stimuli, see Bradley, Hamby, Löw, & Lang, 2007). On the other hand, we aim to induce valence in an ecologically valid way. To this end, as a valence induction, we introduce a game-like situation, in which different tone-frequencies signal the danger of losing and the chance of gaining money, respectively. Note, that we use direct, tangible resources (i.e., money that could be obtained immediately during the experiment), in contrast to paradigms employing more indirect punishments and rewards, such as presenting only the representations of aversive or desirable objects (e.g., viewing the picture of a desirable food that cannot be obtained during the experiment).

In the visual domain, several studies have applied the approach of associating initially neutral stimuli with evaluative content during a learning phase. However, these studies have mostly associated threat with simple stimuli via conditioning, without any comparison with positive valence (e.g., Batty, Cave, & Pauli, 2005; Hintze, Junghöfer, & Bruchmann, 2014; Notebaert, Crombez, Van Damme, De Houwer, & Theeuwes, 2011; S. D. Smith, Most, Newsome, & Zald, 2006; for exceptions, see Müller et al., 2015; Wentura et al., 2014). In the following studies, importantly, both negative and positive valence is attached to a priori neutral stimuli. Given this background, we address the question arising from a lingering debate in the field of evaluative picture processing, that is, whether attentional prioritization of valenced stimuli is driven (a) by the evaluative content in general (regardless of the direction of valence), or (b) by a specific valence category. The former assumption (a) is supported by the general relevance bias hypothesis, namely, that valenced stimuli in general trigger attentional resources based on their higher goal-relevance compared with neutral stimuli (e.g., Brosch et al., 2008; Müller et al., 2015; Wentura et al., 2014). We contrasted this assumption with the assumption of (b) a bias for a specific valence category that is supported by the negativity bias hypothesis, claiming that our attentional system is tuned to prioritize negative information over positive information, allowing for rapid detection of negative – or more specifically, threatening – stimuli (e.g., Öhman et al., 2001; Öhman & Mineka, 2001). To the best of our knowledge, the present work is the first to contrast these two assumptions systematically in the auditory domain.

Experiments 1-2 revealed a notably fast evaluation of diverse natural emotional sounds despite their complex temporal structure. Although natural emotional sounds can be evaluated remarkably rapidly, the evaluative connotation of these spectrotemporally rich, naturally valenced sounds is typically not available immediately at sound onset (cf. the evaluative content of emotional pictures that is available instantly at stimulus onset). In the following experiments, we employ a complementary approach to that used in Experiments 1-2: Thus, we associate affective valence to different tone-frequencies of simple sinusoidal tones (i.e., tones consisting of a single sinusoidal wave). We have chosen tone-frequency as the critical stimulus feature to convey valence information because of its high importance in auditory object formation (Bregman, Liao, & Levitan, 1990; Griffiths & Warren, 2004), and as it is processed quickly, definitely within 100 ms following tone onset (e.g., Näätänen & Winkler, 1999; Roberts, Ferrari, Stufflebeam, & Poeppel, 2000). Hence, we can assume that attaching valence to tone-frequency allows for even earlier extraction of valence information compared with that found in Experiments 1-2, in which evaluative content of natural emotional sounds was defined by a combination of several stimulus features and complex semantic meaning. Thus, employing tone-frequency as the valence-relevant feature provides not only strict control over arbitrary physical differences but also the advantage of an immediate signal value.

After the valence induction, we employ a selective listening task as the critical test phase. In the selective listening paradigm, concurrent streams of auditory information are presented (see, e.g., Bidet-Caulet et al., 2007; Grandjean et al., 2005; Ross et al., 2010; Woldorff & Hillyard, 1991). These auditory streams are typically easily distinguishable by low-level acoustic cues (e.g., presented to the participants' left and right ear with pitch segregation between the ears). Participants are often instructed to attend the auditory stimulation presented to one ear (*task-relevant channel*) and to perform a monitoring (e.g., detect a target tone) or a “shadowing” task (e.g., repeat the heard speech as quickly as possible) while ignoring the stimulation presented to the other ear (*task-irrelevant channel*). It is assumed in this paradigm that participants devote their voluntary attention to the task-relevant channel while the task-irrelevant channel remains unattended (in the terms of voluntary attention).

Classical investigations on the neurophysiological basis of auditory selective attention considered the selection of the task-relevant stimulation as compared with the supposedly unattended task-irrelevant stimulation. However, it has been demonstrated that not only voluntary, “instructed” attentional selection of the task-relevant stimulation can take place in the selective listening paradigm, but the task-irrelevant information can also “capture” attention based on perceptual salience (Ross et al., 2010; Schröger, 1996), indicating discriminative analysis of the stimuli presented in the task-irrelevant channel. Moreover, task-irrelevant stimuli can be processed in an unintentional way beyond basic physical features: Personally significant or affectively salient semantic content of the irrelevant speech stimuli could be reported by (some of the) participants or have been demonstrated to influence task-relevant responses implicitly (e.g., Conway, Cowan, & Bunting, 2001; Degner, 2011; Nielsen & Sarason, 1981; Wood & Cowan, 1995; for the classical demonstration of the “own-name effect”, see Moray, 1959).

In our selective listening paradigm, participants are required to perform a demanding perceptual detection task in the task-relevant channel on a series of noise bursts (Experiment 3 and 5) or on a continuous noise including abrupt changes (Experiment 4). Importantly, tones with newly acquired positive, neutral, and negative valence are presented concurrently in the task-irrelevant channel. Participants are explicitly instructed to focus their attention to the task-relevant stimulation and ignore the auditory stimuli presented in the task-irrelevant channel. Thus, we can expect that participants devote their voluntary attention to the task-relevant channel in this paradigm, while the positive, negative, and neutral tones presented in the task-irrelevant channel are (initially) unattended. Of the most interest in the present section, we will investigate possible preferential attentional enhancement by the task-irrelevant and to-be-ignored valenced tones relative to neutral ones.

The excellent temporal resolution of the ERP method allows us to identify the point in time at which evaluative features influence auditory processing. In the auditory domain, the N1 component is most often reported to be modulated by the participant's attentional state in the relatively early time range (i.e., within 100-150 ms following tone onset) during the sensory encoding of the auditory stimulus (e.g., Folyi, Fehér, & Horváth, 2012; Herrmann & Knight, 2001; Woldorff & Hillyard, 1991). The auditory N1 is a negative-going waveform that peaks maximally at frontocentral electrodes approximately 80-120 ms following a tone onset or other transient change. It results from the activity of various neural sources, presumably including primary and secondary auditory cortices and non-modality specific brain regions (Giard et al., 1994; Näätänen & Picton, 1987; Woods, 1995). The N1 is considered as an example of "exogenous" auditory ERP components, as it reacts sensitively to changes in acoustic features and stimulus presentation characteristics (e.g., Barry, Cocker, Anderson, Gordon, & Rennie, 1992; Budd, Barry, Gordon, Rennie, & Michie, 1998; Crottaz-Herbette & Ragot, 2000; Dimitrijevic, Michalewski, Zeng, Pratt, & Starr, 2008; Jacobson, Lombardi, Gibbens, Ahmad, & Newman, 1992; Weise, Schröger, Fehér, Folyi, & Horváth, 2012). In functional terms, it reflects transient detection as it can be elicited by stimulus onsets, offsets, and changes in continuous stimulation (e.g., Dimitrijevic et al., 2008; Näätänen & Winkler, 1999; Weise et al., 2012).

As mentioned above, while the auditory N1 is considered to reflect "obligatory" sensory processing, it also reacts sensitively to the manipulations of attention (for reviews, see Giard et al., 2000; Herrmann & Knight, 2001), making it an ideal candidate for investigating potential early preferential attention for valenced tones. ERP studies on auditory attention typically report an enlarged N1 amplitude for attended compared with identical but unattended tones (e.g., Hillyard et al., 1973; Woldorff & Hillyard, 1991; for a review, see Giard et al., 2000). In line with previous findings in the visual domain (for reviews, see, Olofsson et al., 2008; Schupp et al., 2006) and auditory domain (Bröckelmann et al., 2011, 2013; Kluge et al., 2011), we expect that valenced tones evoke an early attentional enhancement. More specifically, we expect that this early attentional enhancement will be reflected in an increased amplitude of the N1 for valenced (negative and/or positive) as compared with neutral tones. Moreover, as valenced tones are entirely task-irrelevant and to-be-ignored in the present paradigm, we expect that this attentional enhancement occurs relatively automatically.

In Experiment 3, we address the question whether task-irrelevant and to-be-ignored tones with newly acquired valence attract preferential attention already at a perceptual stage of auditory processing. Furthermore, we target the question whether the expected preferential attention is driven by the general relevance of the valenced tones or rather by a specific (especially negative) valence category. Two subsequent experiments further specify (a) the automaticity of the expected attentional enhancement concerning whether it can occur strictly in the absence of voluntary attention (Experiment 4), and (b) whether the motivational context of anticipating positive or negative future outcomes can moderate this attentional enhancement (Experiment 5).

Experiment 3: Rapid Auditory Attention to Valenced Tones

Experiment 3 investigated whether tones with positive and/or negative affective valence are prioritized over neutral ones at an early, perceptual stage of auditory processing. First, during a learning phase, we assigned positive, negative, and neutral meaning to tone-frequencies in a balanced design. In the subsequent test phase, participants were instructed to attend the auditory stimulation presented to one ear (task-relevant channel), while ignoring stimulation presented to the other ear (task-irrelevant channel). On the task-relevant channel, participants performed a perceptual detection task on a series of noise bursts, while positive, negative, and neutral tones were presented concurrently on the task-irrelevant channel. ERPs elicited by these tones were of the most interest in the present experiment. Thus, EEG was recorded for the valenced tones when they were entirely irrelevant to a demanding concurrent task and, moreover, participants were explicitly instructed to ignore them. This task setting supports the interpretation that tones presented in the task-irrelevant channel are likely (initially) unattended. Finally, we tested whether participants learned the associations between tone-frequencies and evaluative meaning.

If positive and/or negative tones receive early facilitation by rapid attention, we can expect an early attentional enhancement for these tones compared with neutral ones: (a) Either for both positive and negative tones, in the case of a general relevance bias; or (b) for a specific valence category (i.e., in line with a negativity bias at an early level of sound encoding). Accordingly, we expected that this early attentional effect will be reflected in a differential enhancement of the N1 amplitude.

METHODS

Overview of the Experimental Design

Figure 2 gives an overview of the main phases of Experiment 3. The experiment consisted of three main parts: (A) a *valence induction phase*, (B) a *test phase*, and (C) a *manipulation check phase*. First, in the valence induction phase (*Figure 2A*), half of the tones were presented with constant intensity and half of the tones with increasing intensity. The participants' task was to detect the tones with increasing intensity. Depending on their performance, participants could win or lose money in a game-like situation. Three tone-frequencies were associated with negative (i.e., danger of losing money in the case of insufficient performance), positive (i.e., chance to win money in the case of sufficient performance), and neutral (i.e., no substantial loss or gain) connotation, respectively. A fourth tone-frequency served as a no-go signal (see *Procedure*). The assignment of tone-frequencies to valence conditions was counterbalanced between participants. In a subsequent test phase (*Figure 2B*), a selective listening task was administered, that is, participants were instructed to attend the auditory stimulation presented to one ear and accomplish a perceptual detection task on a series of noise bursts (task-relevant channel), while ignoring stimulation presented to the other ear (task-irrelevant channel). On the task-irrelevant channel, negative, positive, and neutral tones of the previous valence induction phase were presented, and ERPs elicited by these tones were analyzed after the experiment. Finally, in the manipulation check phase (*Figure 2C*), we tested whether tones indeed acquired positive and negative valence during the valence induction. Therefore, we applied the auditory version of the AST (see Experiment 2) as an indirect method to assess stimulus valence. We expected shorter RTs (and/or lower ERs) for congruent (stimulus valence and response valence match) compared with incongruent (stimulus valence and response valence mismatch) trials. At the end of the experiment, participants rated the valence of the tones explicitly.

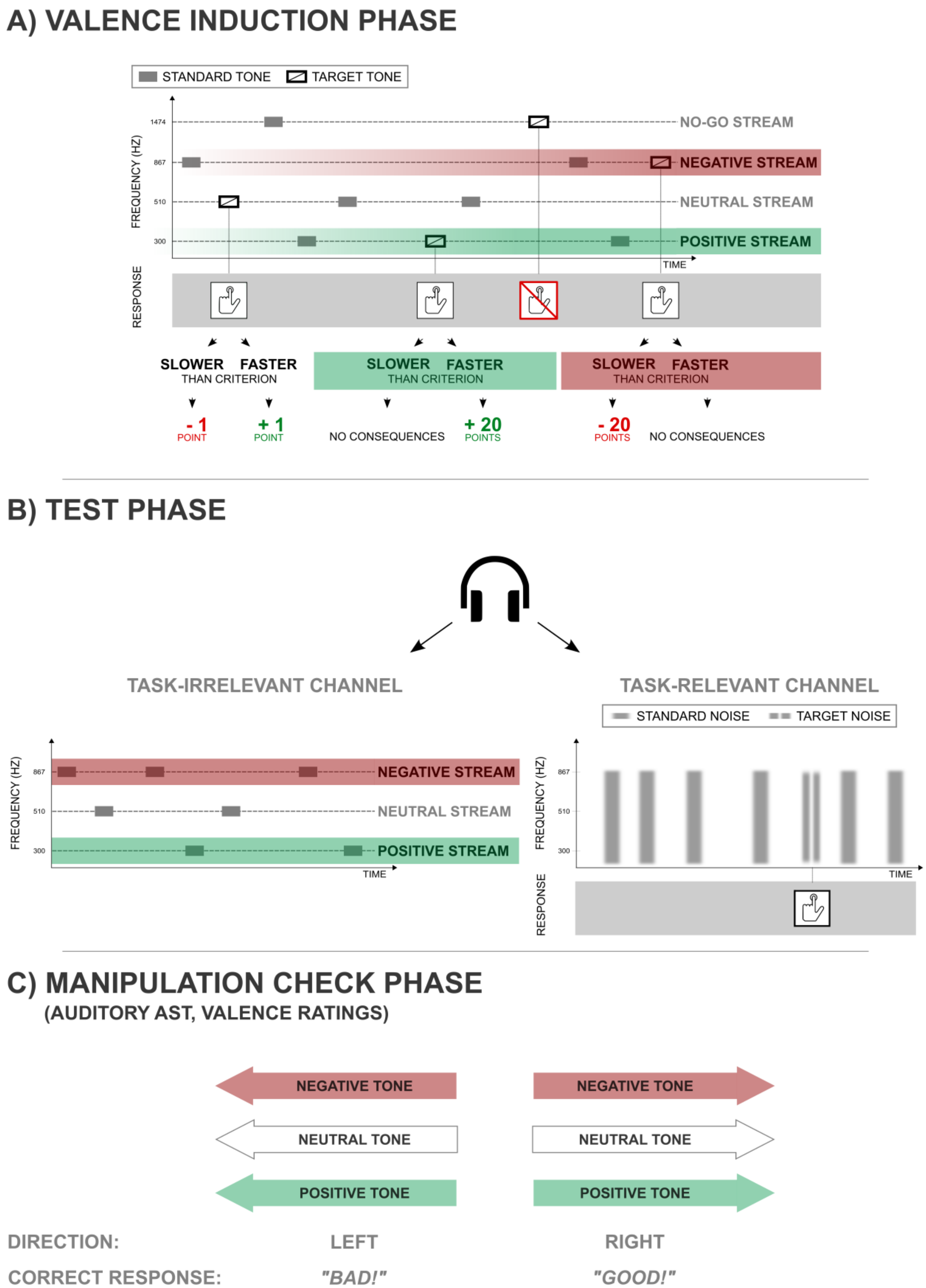

A) VALENCE INDUCTION PHASE
STANDARD TONE
TARGET TONE
FREQUENCY (HZ)
1474
867
510
300
NO-GO STREAM
NEGATIVE STREAM
NEUTRAL STREAM
POSITIVE STREAM
TIME
RESPONSE
SLOWER FASTER
THAN CRITERION
SLOWER FASTER
THAN CRITERION
SLOWER FASTER
THAN CRITERION
- 1
POINT
+ 1
POINT
NO CONSEQUENCES
+ 20
POINTS
- 20
POINTS
NO CONSEQUENCES
B) TEST PHASE
TASK-IRRELEVANT CHANNEL
TASK-RELEVANT CHANNEL
STANDARD NOISE
TARGET NOISE
FREQUENCY (HZ)
867
510
300
NEGATIVE STREAM
NEUTRAL STREAM
POSITIVE STREAM
TIME
RESPONSE
C) MANIPULATION CHECK PHASE
(AUDITORY AST, VALENCE RATINGS)
NEGATIVE TONE
NEUTRAL TONE
POSITIVE TONE
NEGATIVE TONE
NEUTRAL TONE
POSITIVE TONE
DIRECTION:
LEFT
RIGHT
CORRECT RESPONSE:
"BAD!"
"GOOD!"

Figure 2. Schematic illustration of the design of Experiment 3.

Note that only one instance of the counterbalancing scheme is depicted. (A) In the valence induction phase, valence was induced experimentally in a game paradigm by assigning positive (chance to gain money), negative (danger to lose money), and neutral (without substantial gain or loss) meaning to tone-frequencies. (B) In the test phase, positive, neutral and negative tones were presented in a task-irrelevant channel while a perceptual detection task was administered in the task-relevant channel. Participants were instructed to direct their attention to the task-relevant channel, while ignoring the task-irrelevant channel. (C) In the manipulation check phase, the effectiveness of the valence induction was tested in an auditory AST and by collecting explicit valence ratings.

Participants

Twenty-four students from Saarland University (13 females; aged 17–28 years, *Mdn* = 22 years; two left-handers) participated in the experiment for monetary compensation. Given a sample size of $N = 24$ and an α-value of .05 (two-tailed), effects of size $d_z = 0.60$ can be detected with a probability of $1 - \beta = .80$ for the most relevant comparisons that are representing our central hypotheses (i.e., negative vs. positive and valenced vs. neutral comparison; calculated with the aid of G*Power 3 software, Faul et al., 2007). In our experiments employing ERP methodology, attaining adequate test power has been aimed with a consideration of keeping the relatively costly data collection within a reasonable range (we will elaborate on this issue in the *General Discussion*). All participants were native German speakers, had normal or corrected-to-normal vision, and self-reported normal hearing. Before the experiment, all participants gave written informed consent. On average, participants received €34 (compensation was partly dependent on performance; see *Procedure* for details). Six additional individuals were excluded before data analyses because of technical problems resulting in incomplete data or because they failed to understand instructions. Data from three further participants were discarded during data analysis: two due to extensive EEG-artifacts, and one due to the absence of an observable N1 waveform in the averaged ERPs[6]. The assignment of tone-frequencies to valence conditions was counterbalanced between participants according to a Latin square scheme (Winer, 1962) resulting in four counterbalancing groups (*Table A1* in *Appendix B* depicts the exact counterbalancing scheme). In the final sample, two counterbalancing groups had six, one seven and one five participants. In the manipulation check phase, twenty-two students participated as two participants had to be excluded because of technical failure of the voice key apparatus.

Materials

Sinusoidal tones with four different tone-frequencies (300 Hz, 510 Hz, 867 Hz, and 1473.9 Hz) were presented via headphones (HD-600, Sennheiser, Wedemark, Germany). Frequency differences between tones were calculated by the following formula: $\text{frequency}_n = 1.7 \times \text{frequency}_{n-1}$; e.g., 510 Hz = 1.7 × 300 Hz. The resulting tones are perceived as roughly equidistant in pitch, but there was no harmonic or musical relation between them. The maximal intensity level was 45 dB sensation level (SL, above individual hearing level referred to a 1000 Hz sinusoidal tone). All auditory stimuli faded in and out with 10 ms linear rise and fall times.

Valence induction phase

In the valence induction phase, tone duration was 1400 ms, and all four tones were presented in two versions. Standard tones were presented with constant intensity (0.5 × maximal intensity; ~ 39 dB SL). Target tones started with the same constant intensity, but 1000 ms post onset their intensity rose to the maximal level (i.e., 45 dB SL) with a linear rise time of 390 ms.

[6] Including this data set to the final sample does not change the pattern of results reported below.

Test phase

In the selective listening task of the test phase, tones presented in the task-irrelevant channel were identical to the standard tones of the valence induction phase, except for the following changes: Tone duration was reduced to 800 ms and intensity was 0.4 × maximal intensity (~37 dB SL). In the task-relevant channel, white noise bursts were presented with an overall duration of 500 ms and intensity level of approximately 32 dB SL. We used two versions of the noise bursts: While standard noise burst was continuous, target noise burst was interrupted by a 4 ms silent period ("gap")[7] starting at 200 ms post onset.

Manipulation check phase

In the auditory AST, the amplitude of the tones was modulated in an identical way as in Experiment 2a: From 500 ms after stimulus onset, the intensity in the left or right auditory signal channel of the stereo sound was reduced linearly over a 1000 ms interval by a total of 75%. Thereby we created two versions of each tone: with an illusory movement from a central position toward the right ("moving to the right" tones) or left ("moving to the left" tones) side of the perceiver, respectively. Note, that following the design of Experiment 2a, exposure to the evaluative information started 500 ms before the task-relevant manipulation, thereby expectedly securing more marked AS effect compared with a synchronous version (see the comparison of Experiment 2a and 2b). Overall tone duration was 6000 ms, but tones were terminated when a response was recorded. Additionally, 30 natural sounds from the IADS battery (Bradley & Lang, 2007) were presented as filler trials. We selected ten positive (e.g., baby laugh), ten negative (e.g., attack), and ten neutral (e.g., office noise) sounds and created a "moving to the right" and "moving to the left" versions of them,

---

[7] Note that detecting the target noise, thus, the noise bursts with a 4 ms long silent interval can be considered as a demanding perceptual task as the threshold for detecting silent intervals in broadband noises is typically about 2-3ms (B. C. J. Moore, 2012).

in the same way as described above. Averaged normative valence ratings on a 9-point scale ranging from most unpleasant (1) to most pleasant (9) were: M = 7.20 (SD = 0.45) for positive, M = 2.62 (SD = 0.53) for negative, and M = 4.62 (SD = 0.49) for neutral sounds, respectively.

Procedure

Participants initially received €10 as a payment for the first 1.5 hours of the experiment but were obliged to risk the money as the stakes in a subsequent "game". Participants received further €4.50 for each additional started half-hour. The experimental session lasted about 3.5–4 hours, including individual breaks and the time for electrode application and removal.

Participants were comfortably seated in a sound-attenuated room. Hearing thresholds were individually determined by using a continuous, 1000 Hz sinusoidal tone at the beginning of the experiment. This level (0 dB SL) was used as a reference during the experiment. Before the actual experiment started, participants had to accomplish two discrimination tasks: The first required participants to discriminate the four tone-frequencies, the second required them to learn the associations between tone-frequencies and the gain/loss odds. Before the second discrimination task, the possible outcomes related to each tone were shown to the participants. Depending on their outcome odds, three tones were referred in the instructions as "danger", "chance", and "neutral" tones (there was a fourth tone in the valence induction task termed as "no-go" tone, see *Valence induction phase*). The second discrimination task tested for the associations between the four tone-frequencies and the four meanings during the game (i.e., "danger", "chance", "neutral", "no-go"). The discrimination tasks terminated when the accuracy level exceeded 95% in the first task, and 90% in the second task.

The valence induction phase (A) consisted of one practice and three experimental blocks of the valence induction task. The test phase (B) comprised ten blocks of the selective listening task, with an additional block of the valence induction task interspersed after every two blocks of selective listening (i.e., the two phases together consisted of seven valence induction blocks and ten selective listening blocks). Blocks were separated by short, participant-terminated breaks. Additionally, in the manipulation check phase (C), an auditory AST was administered. Finally, explicit valence ratings were collected.

Valence induction phase

In the valence induction phase (*Figure 2A*), participants could win or lose game scores depending on their performance. Every valence induction block started with a score of zero. Participants were informed that they would immediately win €1 if the final total score of the block was zero or above. If the final score was below zero, they immediately lost €1. Thus, there was a chance of winning or losing up to €7 throughout the experiment. At the end of each valence induction block, a feedback indicated the final total score and the actual outcome of the block (thus, plus or minus €1).

A valence induction block comprised 56 auditory trials presented in random order. Each trial started with the presentation of a black fixation cross with a randomized duration of 500-1000 ms. Thereafter, one of the tones was played binaurally. Half the trials featured standard tones (i.e., tones with constant intensity) and half the trials featured target tones (i.e., tones with increasing intensity). Participants were instructed to respond only to the target tones by pressing the space bar as quickly and accurately as possible. A fast and correct response was considered a success trial; a slow or incorrect response (i.e., a missed target or a false alarm) was considered a failure trial. The response-speed criterion for being successful on a given trial was defined by the moving median of the preceding six trials (see, e.g., Rothermund et al., 2008)[8]. Thereby we aimed to ensure that participants experience success and failure trials in a relatively balanced amount. During the valence induction task, the tone-frequency determined the consequences of success or failure. One tone-frequency was associated with a gain of 20 points in case of a success, but no negative consequences in case of a failure (positive tone). Another tone-frequency was associated with a loss of 20 points in case of a failure, but no positive consequences in case of a success (negative tone). A third tone-frequency was associated with a negligible gain or loss of one point in case of either success or failure (neutral tone). Note that the possible outcomes related to each tone were presented explicitly to the participants at the beginning of the experiment (see above). A fourth tone-frequency served as a no-go signal, requiring participants to withholding their response even when the tone was presented in its target form (i.e., with increasing loudness). Responding to the no-go tone resulted in a loss of 20 points. The no-go tone was introduced to make the tone-frequency a task-relevant feature, thereby supporting the learning of the tone-frequency–valence associations. The no-go tone was presented only in the valence induction phase. Visual feedback was given after each target trial and in case of a false alarm, with the feedback indicating (1) the type of the tone that was presented (e.g., "danger" target tone), (2) the consequences of the recent response (e.g., "-20 points"; on the break-even trials – for example in case of successful performance on a negative trial – the feedback stated that "There are no consequences"), and (3) the current total score[9]. In a valence induction block, 14 positive, 14 negative, 14 neu-

---

[8] The actual response-speed criterion (median') was dependent on a participant's current total game score. If the current total game score exceeded 50 points: median' = 1.02 × median, if it was below -50 points: median' = 0.98 × median, if it was between -50 and 50 points: median' = median; thereby leading to current scores that tend to be around the critical value of zero.

[9] 20-point gains and losses additionally elicited feedback sounds (a fanfare-like trumpet sound in case of a gain and a guitar sound with decreasing pitch in case of a loss; both sounds were provided by the

tral and 14 no-go tones were presented (half of them in their standard, and half of them in their target version, respectively). In each valence induction block, seven additional visual detection trials were introduced to ensure that participants keep their visual attention on the screen and thus encode the visually presented feedback. A visual detection trial started with a 500 ms presentation of a black fixation cross, which then turned red; and participants had to press the space bar as quickly as possible when the color changed. If the participant did not respond within 2000 ms of the color change, error feedback was presented visually.

Test phase

In the test phase (*Figure 2B*), participants were instructed to direct their attention to one ear (task-relevant channel), while ignoring the sequence of tones presented to the other ear (task-irrelevant channel). In the task-relevant channel, 18 target noise bursts and 54 standard noise bursts were presented monaurally in a randomized order in each block, with a random interstimulus interval (ISI) of 2500-3500 ms. Participants had to respond to the target noise bursts containing a silent interval as quickly and accurately as possible by pressing the space bar. After each block, visual feedback about the mean hit rate (HR) was presented in order to motivate participants to maintain a high level of performance. In the task-irrelevant channel, positive, neutral, and negative tones of the valence induction phase were presented monaurally in their the standard version with a randomized ISI of 1000-1333 ms. The two series of auditory stimuli presented to the two ears were allowed to overlap at random temporal positions. Task channel and ear assignment was counterbalanced between participants.

FreeSound Project, http://www.freesound.org). Note that these additional sounds were presented only in the valence induction phase.

Manipulation check phase

In the manipulation check phase (*Figure 2C*), the effectiveness of the valence induction manipulation was tested in an auditory AST. The task comprised 12 practice and 30 experimental trials. Positive, negative, and neutral tones were presented ten times each in random order (five times in "moving to the left" and five times "moving to the right" version). The sequence was randomly interspersed by filler trials. That is, ten positive, ten negative, and ten neutral natural sounds were presented once during the task (with half of them in "moving to the left" and half of them in "moving to the right" version, respectively). Participants had to categorize the direction of the "movement" by saying "good" for the "moving to the right" sounds and "bad" for "moving to the left" sounds as quickly and accurately as possible. The assignment of response (saying "good" or "bad") to illusory movement direction (right or left) was not counterbalanced between participants, as Experiment 2 did not show differences between counterbalancing groups.

A trial started with a 1000 ms long visual presentation of a fixation cross without auditory stimuli. (The fixation cross remained on the screen until the end of the trial.) After 1000 ms, a positive, negative, or neutral tone (experimental trials) or a natural sound (filler trials) was played. In the first 500 ms of the stimulus presentation, the intensity was equal in the two auditory signal channels of the stereo stimulus. From 500 ms after stimulus onset, the intensity in one auditory signal channel declined over a 1000 ms interval, thereby creating a movement illusion toward the right or the left side of the perceiver. While a voice key apparatus recorded the RT, the experimenter registered the response category online. After the vocal response was detected by the voice key, the auditory stimulus was terminated. Finally, error feedback was presented visually in the case of an incorrect response.

At the end of the experiment, participants were asked to rate the valence of the auditory stimuli presented in the AST task on a 9-point scale ranging from most unpleasant (1) to most pleasant (9). During the valence rating, auditory stimuli were presented with constant intensity.

EEG Recording and Analysis

The EEG was recorded only during the test phase from 64 scalp locations (following the international 10–10 system, and including left and right mastoids). The common reference electrode was placed on the tip of the nose. The continuous EEG was amplified from DC to 100 Hz at a 500-Hz sampling rate. On-line 70-Hz low-pass filtering was applied, and the signal was band-pass filtered offline (0.5–30 Hz). Horizontal eye movements were monitored with a bipolar setup, with electrodes placed laterally to the outer canthi of both eyes; vertical eye movements were monitored with electrodes placed above and below the right eye.

ERPs elicited by the tones of the task-irrelevant channel were calculated during offline analysis. To this end, the continuous EEG was segmented into 1000 ms long epochs, each including a 200 ms long baseline preceding tone onset. Epochs containing task-relevant sounds and epochs corresponding to the first three trials of each block were discarded (rejection criteria: onset of a task-relevant standard noise burst within a -800 ms to +800 ms time-window relative to the onset of the task-irrelevant tone, or onset of a task-relevant target noise burst within a -1000 ms to +800 ms time-window relative to the onset of the task-irrelevant tone). We applied these rejection criteria, on the one hand, to avoid overlapping of neural responses elicited by the task-irrelevant and the task-relevant stimuli, and on the other hand, to maximize trial numbers and thereby increasing signal-to-noise ratio; and to discard the trials with expectedly enhanced N1 amplitude after the silent breaks between stimulus blocks (see, e.g., Näätänen & Picton, 1987), and thereby reducing error variance. Additionally, epochs contaminated with severe artifacts were rejected (rejection criteria: signal range exceeding 200 µV, or voltage step exceeding 50 µV/ms, or a voltage difference exceeding 150 µV on any channel). On average, ERPs were based on 165, 163, and 159 trials per participant in the positive, neutral, and negative valence conditions, respectively. Epochs were baseline corrected using the 200-ms pre-stimulus interval and averaged separately for the different valence conditions. We formed a region of interest (ROI) from frontocentral electrode sites (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) according to the auditory N1-literature (for a review, see Näätänen & Picton, 1987), and the frontocentral scalp distribution of our grand average N1. N1 amplitude was measured at the frontocentral ROI as mean voltage in a 20-ms time window centered at the latency of the group-average N1 peak (114 ms, experimental conditions collapsed; for comparable method see, e.g., Gilmore, Clementz, & Berg, 2009; Horváth, Maess, Baess, & Tóth, 2012; Jacobsen, Schröger, Horenkamp, & Winkler, 2003).

Design and Statistical Analysis

We applied a 3 × 4 mixed design with valence (positive, neutral, negative) as a within-participants factor and tone-frequency-to-valence assignment (four counterbalancing groups) as a between-participants factor on the N1 amplitudes and the behavioral measures of the valence induction phase. We introduced counterbalancing group as a between-participants factor into the main analysis following the suggestion of Pollatsek and Well (1995) to use the correct error term and to account for the slightly different sample sizes of the final counterbalancing groups (see *Appendix A* for more details concerning this analysis). As effects involving the group factor are not theoretically relevant, we do not report them in order to reduce the complexity of the presentation of results.

We used the multivariate approach to repeated measures analysis, which means that the tripartite factor of valence was transformed into a vector of two orthogonal contrast variables (see, e.g., O'Brien & Kaiser, 1985; Petrova & Wentura, 2012; M. Rohr, Degner, & Wentura, 2012). We a priori chose the contrasts in a way that they represent our specific hypotheses: That is, (1) for the first contrast, amplitude values were averaged across positive and negative stimuli and contrasted with the neutral stimuli. This contrast represents the hypothesis that valenced tones (in general) produce larger attentional effects compared with neutral tones as predicted by a general relevance bias hypothesis. (2) The second contrast was the contrast between the two valenced conditions (i.e., between positive and negative tones), representing the hypothesis of larger attentional effects for negative compared with positive tones in line with the negativity bias hypothesis.

AST data was analyzed in line with Experiment 2: Thus, we employed a 2 (stimulus valence: positive vs. negative) × 2 (response valence: positive vs. negative) repeated measures design, which reduces to a simple one-factorial congruency (congruent vs. incongruent) design.[10] For the valence ratings, 3 × 4 mixed design with valence (positive, neutral, negative) as a within-participants factor and tone-frequency-to-valence assignment (four counterbalancing groups) as a between-participants factor was applied as a standard analysis.

[10] Note that congruency was a within-items factor for the experimental tones in the present design (i.e., positive and negative experimental tones were presented both in congruent and incongruent pairings with the response valence). Thus, the AS effect (i.e., incongruent-minus-congruent difference) accounts for the main effect of tone differences.

Results

A significance level of $\alpha = .05$ (two-tailed) was adopted for all analyses.

Behavioral Performance

RTs below 100 ms were discarded when calculating averaged RTs. In the valence induction phase, RTs were calculated from the beginning of the loudness increase in the ongoing tone. Behavioral measures of the valence induction phase were analyzed in order to ensure that participants complied with the task instructions, and thereby they expectedly learned the valence-to-tone-frequency associations (see *Table 4*).

Table 4. Behavioral results of the valence induction phase in Experiment 3: Mean reaction times (RTs), accuracy rates (ACCs), false alarm rates (FARs), and hit rates (HRs) in each valence condition, *SD* in parentheses.

Valence	**RT**	**ACC**	**FAR**	**HR**
Positive	323 (39)	88.4 (13.5)	17.3 (26.3)	94.0 (7.6)
Neutral	326 (31)	93.7 (10.5)	10.9 (20.8)	98.2 (3.5)
Negative	332 (38)	92.0 (9.5)	7.2 (12.0)	91.2 (10.1)

A 3 × 4 MANOVA for repeated measures with valence as within-participants factor and counterbalancing group as between-participants factor of the RTs did not show any valence differences, $F(2,19) = 0.98$, $p = .394$, $\eta_p^2 = .093$. The average ACC across participants was adequately high ($M = 91.3\%$, $SD = 10.6\%$), and showed a valence main effect, $F(2,19) = 6.98$, $p = .005$, $\eta_p^2 = .423$. The a priori contrast of positive vs. negative valence was significant, $F(1,20) = 6.11$, $p = .023$, $\eta_p^2 = .234$; and valenced vs. neutral conditions also differ significantly, $F(1,20) = 11.81$, $p = .003$, $\eta_p^2 =$

.371. This valence difference in ACC is more understandable in the light of the false alarms: the relatively low mean ACC in the positive condition emerged due to high FARs, while the lowest mean FAR emerged in the negative condition. Note that applying a liberal response criterion pays off only in the positive condition, as false alarms had no negative consequence only in this condition, while they had extreme negative consequence in the negative condition. FARs did not show significant valence effect, $F(2,19) = 2.55$, $p = .104$, $\eta_p^2 = .212$; however, as expected, the contrast of the positive vs. negative conditions was significant, $F(1,20) = 4.89$, $p = .039$, $\eta_p^2 = .196$, while valenced vs. neutral conditions did not differ significantly, $F(1,20) = 0.39$, $p = .538$, $\eta_p^2 = .019$. However, HRs indicated better performance for neutral tones: Analogue analysis on the HRs showed a significant valence main effect, $F(2,19) = 15.89$, $p < .001$, $\eta_p^2 = .626$; the contrast of the valenced vs. neutral conditions was significant, $F(1,20) = 33.43$, $p < .001$, $\eta_p^2 = .626$, indicating higher HRs in the neutral as compared with the two valenced conditions; while positive vs. negative conditions did not differ significantly, $F(1,20) = 1.28$, $p = .271$, $\eta_p^2 = .060$.

In the selective listening task of the test phase, the average RT across participants was 590 ms (*SD* = 74 ms) with an average ACC of 93.2% (*SD* = 5.1%). In sum, behavioral performance was adequate in the valence induction and selective listening tasks, suggesting that participants were engaged in both tasks.

ERP Results

Of most interest for the present study, we analyzed the amplitude of the auditory N1 elicited by the task-irrelevant positive, negative, and neutral tones during the test phase. Prototypical P1-N1-P2 waveform was clearly observable in the group average ERPs (see *Figure 3*, and for the mean amplitudes for the components of interests, see *Table 5*).

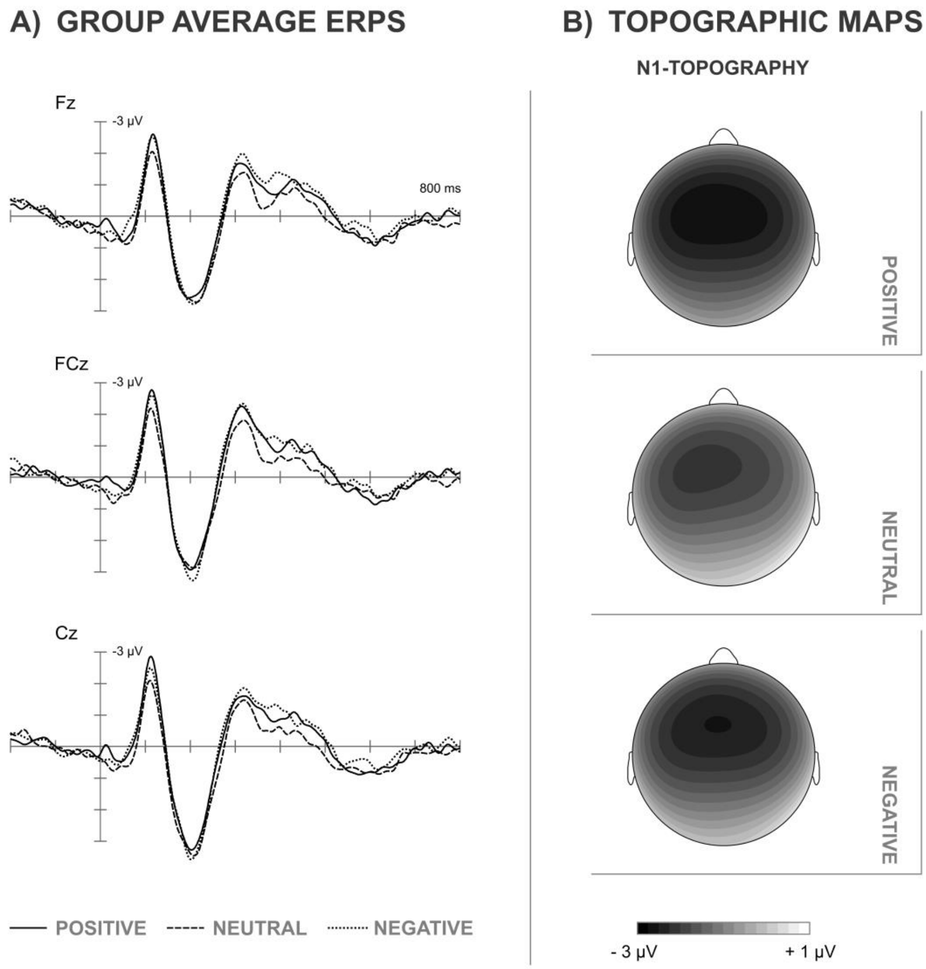


Figure 3. ERP results of Experiment 3.

(A) Group average ERP waveforms to the positive, neutral, and negative tones on the representative Fz, FCz and Cz electrode sites in Experiment 3. The physical onset of the tones is at the crossing of the axes (0 ms). Negative polarity is plotted upwards. (B) Group average topography maps in the N1 time window (104-124 ms) in positive, neutral and negative conditions.

Table 5. ERP results of Experiment 3. Mean amplitudes (in μV) of the P1, N1 and the late negativity (LN; 350 – 500 ms) at the frontocentral ROI in each valence condition, *SD* in parentheses.

Valence Condition	**P1**	**N1**	**LN**
Positive	0.48 (0.72)	-2.42 (1.67)	-0.75 (0.90)
Neutral	0.73 (0.89)	-1.84 (1.59)	-0.34 (1.12)
Negative	0.53 (1.07)	-2.23 (1.58)	-0.91 (0.99)

A 3 × 4 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and counterbalancing group as between-participants factor of the N1 amplitudes showed a tendency for valence main effect, $F(2,19) = 2.66$, $p = .096$, $\eta_p^2 = .219$. The a priori planned contrast of valenced (positive and negative) vs. neutral condition – that tested for the general relevance bias hypothesis – was significant, $F(1,20) = 5.08$, $p = .036$, $\eta_p^2 = .202$, indicating enhanced N1 amplitudes in the valenced conditions compared with the neutral condition (M = -2.32 μV, SD = 1.43 μV in the valenced condition; M = -1.84 μV, SD = 1.59 μV in the neutral condition). The contrast of the two valenced conditions (positive vs. negative) – that tested for the negativity bias hypothesis – did not show any differences on the N1 amplitudes, $F(1,20) = 0.24$, $p = .632$, $\eta_p^2 = .012$.

Although our main hypotheses relate to the N1 component (approximately 100-150 ms following tone onset), we performed further analyses to qualify the apparent differences between valence conditions indicated by *Figure 3A* in the P1 time window (approximately 50 ms following tone onset) and in the later time range (350-500 ms following tone onset). P1 amplitude was measured at the frontocentral ROI as mean voltage in a 20-ms time window centered at the latency of the group-average P1 peak (46 ms, experimental conditions collapsed). A 3 × 4 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and counterbalancing group as between-participants factor, however, showed no valence effect on the P1 amplitudes, $F(2,19) = 0.65$, $p = .531$, $\eta_p^2 = .064$ (with non-significant valenced vs. neutral and positive vs. negative contrasts, Fs < 1.01, *n.s.*).

A further analysis targeted the apparent enhancement of a frontocentrally distributed negativity for valenced compared with the neutral tones in a later time range (approximately 350-500 ms after stimulus onset). A 3 × 4 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and counterbalancing group as between-participants factor on the mean activity measured in the 350-500 ms time range at the frontocentral ROI showed a tendency for a valence main effect, $F(2,19) = 2.76$, $p = .089$, $\eta_p^2 = .225$. The a priori contrast of valenced (positive and negative) vs. neutral condition was significant, $F(1,20) = 5.38$, $p = .031$, $\eta_p^2 = .212$, indicating more negative activity in the valenced conditions compared with the neutral condition (M = -0.83 µV, SD = 0.68 µV in the valenced condition; M = -0.34 µV, SD = 1.12 µV in the neutral condition). The a priori contrast of the two valenced conditions (positive vs. negative) did not show any differences, $F(1,20) = 0.48$, $p = .497$, $\eta_p^2 = .023$.

Manipulation Check

Affective Simon task

RTs were calculated from the beginning of the illusory movement in the ongoing tone. RT analyses were restricted to correct trials (6.3% of the trials was excluded because of erroneous response of the participant or erroneous or non-reaction of the voice key). As an a priori criterion, RTs below 200 ms and above 2000 ms were discarded from further analyses (2.4% of the trials), and RT outliers (1.8% of the trials) were excluded (for each participant an upper and a lower outlier criterion were calculated according to Tukey, 1977, using the distribution of all AST RTs of this person). Mean RTs and ERs for congruent trials (i.e., positive tone/response “good”; negative tone/response “bad”); for incongruent trials (i.e., positive tone/response “bad”; negative tone/response “good”) and for neutral trials, as well as the mean AS effect are presented in *Table 6* for the experimental and filler trials, respectively. As expected, the incongruent-minus-congruent RT-difference showed a significant AS effect for the experimental trials, $t(21) = 2.19$, $p = .040$, $d_z = 0.47$. However, similar analyses on the filler trials (natural sounds) did not yield significant results, $t < 1$, *n.s.* Analogue analyses on the ERs did not show significant AS effect for the experimental trials, $t < 1$, *n.s.*; while on the filler trials participants made more errors in the incongruent compared with the congruent condition, $t(21) = 3.17$, $p = .005$, $d_z = 0.67$.

Table 6. Mean RTs (in ms) and mean ERs (%) as a function of stimulus and response valence congruency for experimental and filler trials in the AST of Experiment 3; *SD* in parentheses.

	Experimental Tones		**Fillers**	
	RT	**ER**	**RT**	**ER**
Neutral	809 (156)	6.0 (13.1)	817 (197)	3.2 (7.8)
Congruent	812 (182)	7.5 (12.5)	828 (174)	3.1 (6.5)
Incongruent	854 (181)	7.1 (8.0)	828 (180)	7.3 (9.5)
AS Effect[a]	42 [19]	-0.4 [2.7]	0 [17]	4.2 [1.3]

[a] incongruent-minus-congruent difference; standard errors in brackets

Valence ratings

The mean valence ratings are presented in *Table 7*. A 3 × 4 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and counterbalancing group as between-participants factor of the ratings did not show valence main effect, $F(2,17) = 0.69$, $p = .515$, $\eta_p^2 = .075$.

Although this standard analysis did not reveal significant differences, further inspection of the explicit valence ratings for the three tones revealed a dominant pattern: Participants oriented themselves on the comparison of stimuli with regard to the most salient feature, that is, pitch (see *Table 7* for the mean ratings). A clear linear pattern of pitch emerged for each participant: The individually lowest pitch was rated as individually most pleasant, the individually highest pitch was rated as individually most unpleasant, and the intermediate pitch received an individually intermediate valence rating, with some exceptions when ratings were equal for the adjacent tones. We used this dominant pattern in the following analysis; thus, we coded tones according to their individual, relative pitch instead of their objective tone-frequency (i.e., lowest, intermediate, and highest tone for a certain participant; note that each participant had to rate only three of the four tones that were assigned with positive, negative, or neutral meaning in a certain counterbalancing group). We have chosen the approach of linear mixed models as an alternative analysis of the valence ratings to assess whether valence has an effect above the conspicuous linear effect of pitch. We used the *lme4* and *lmerTest* packages of R 3.1.3 (Bates, Maechler, Bolker, & Walker, 2014) with significance of predictors assessed using Satterthwaite's approximation for degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2014). First, we ran two random effects model with pitch (-1 = lowest, 0 = medium, +1 = highest) as predictor and rating as dependent variable. Model 1 included random intercepts and random slopes; Model 2 included only random intercepts. Model comparison yielded no significant difference, that is, the removal of the parameter for by-participant random slopes for pitch is justified, $\chi^2(2) = 2.50$, $p = .287$ (see, e.g., Baayen, Davidson, & Bates, 2008). Matching expectations, means (and *SD*s) of the residuals of Model 2 are $M = 0.36$ ($SD = 1.55$) for positive stimuli, $M = -0.29$ ($SD = 1.46$) for negative stimuli, and $M = -.07$ ($SD = 1.59$) for neutral stimuli. Thus, second, we ran two further models on the basis of Model 2: In Model 3 we added valence (-1 = negative, 0 = neutral, +1 = positive) as a predictor including random slopes; in Model 4, we removed the random slopes parameter. Model comparison yielded no significant difference between Models 3 and 4, $\chi^2(2) = 1.35$, $p = .510$, thus, Model 4 can be considered the final model. In this model, pitch as well as valence are significant predictors of the rating, $B = -1.61$, $t(41.39) = -10.33$, $p < .001$ for pitch, and $B = 0.34$, $t(41.16) = 2.20$, $p = .033$ for valence.

Table 7. Mean valence ratings in each valence condition for individually lowest, intermediate, and highest pitch in Experiment 3. Valence ratings range from very unpleasant (1) to very pleasant (9); *SD* in parentheses.

	Pitch			
Valence	**Low**	**Intermediate**	**High**	**All**
Positive	6.00 (2.00)	4.83 (1.47)	3.10 (1.45)	4.64
Neutral	5.40 (1.34)	4.70 (1.77)	2.17 (1.47)	4.09
Negative	5.90 (1.73)	3.17 (1.47)	2.67 (0.52)	3.91
All	5.77	4.23	2.64	

DISCUSSION

The main finding of Experiment 3 is that valenced tones were processed preferentially compared with neutral ones. In particular, we found augmented N1 amplitude for valenced tones compared with neutral ones, thus suggesting enhanced attention for these tones already at a perceptual stage of auditory processing. This result is in line with the view that processing of evaluative stimuli can be facilitated by rapid attention (for a review of dominantly visual studies, see Yiend, 2010). Importantly, similarly to previous findings in the visual domain (e.g., Brosch et al., 2008; Rothermund et al., 2008; Wentura et al., 2014), we did not find a difference between positive and negative valence conditions in the present experiment in the auditory N1-time range. This result is in accordance with a general relevance principle, that is, it is the goal-relevance of the stimulus that possesses attention-grabbing power at the early level of sound encoding rather than a specific valence category.

Besides our main ERP results on the relatively early N1, we found an increased sustained negativity for valenced tones in a later time range (350-500 ms following tone onset). This activity is obviously different from the enhanced positive-going waves for evaluative stimuli starting at about 200-300 ms after stimulus onset (P3 and late positive potentials) that typically occur in several versions of the oddball paradigm with visually presented complex stimulus material. These late positive-going waveforms have been associated with increased attentive processing of the motivationally significant information for subsequent memory storage (for a review, see Olofsson et al., 2008). The late negativity in our results possibly reflects attentional control mechanisms (e.g., Giard et al., 2000), and as in our paradigm valenced tones were entirely task-irrelevant during the test phase, a possible interpretation is that it indexes inhibition for the valenced tones and/or reorientation to the main task (see, e.g., Roye, Jacobsen, & Schröger, 2007, for a similar argumentation).

While the early attentional effect reflected in the enhanced N1 amplitude was comparable for the acquired positive and negative valence, both implicit and explicit behavioral measures differed between valence categories. In the valence induction phase, behavioral results differentiated between valence conditions as reflected in more false alarms for positive compared with negative tones, indicating that participants applied strategies in order to achieve positive monetary outcomes. Although, somewhat surprisingly, HRs were lower for the two valenced conditions compared with the neutral condition, measures of speed and accuracy of reactions indicated an overall high level of performance in each condition of the valence induction task. Altogether, the pattern of behavioral results can be interpreted as an indicator that participants were engaged in the tasks with a constant effort in line with the task instructions. In the manipulation check phase, by using an auditory AST for accessing implicit evaluations, we found evidence that the simple tones had indeed acquired positive and negative affective valence during the previous learning phase. It might appear somewhat surprising that the natural sounds (i.e., the filler trials) did not reveal an AS effect on the RTs. Note, that the manipulation check was administered at the end of a long and demanding experimental session. For precisely that reason we used only a few trials; moreover, each natural sound was presented once and was not balanced for movement direction. However, although RTs for the natural sounds did not show the expected difference, participants made more errors in the incongruent compared with the congruent condition on the filler trials. Additionally, in the independent Experiment 2 using only natural sounds we did find the expected AS effect on RTs. Besides the evidence for acquired implicit valence, we also found support for valence differentiation in explicit self-reports: Although the explicit valence ratings reflected dominantly the most salient feature of the stimuli, that is, pitch (here, the pitch of the three rated tones relative to each other instead of their objective tone-frequency), they also revealed the expected differences in acquired affective valence.

Taken together, the results of Experiment 3 support the view that sensory encoding of valenced stimuli is facilitated by "natural" attention. Several characteristics of the paradigm favor the interpretation of the reported attentional enhancement as automatic in the sense as it occurred involuntarily: EEG was measured when the valenced tones were presented on an entirely task-irrelevant channel, and participants were explicitly asked to direct their attention to the concurrent task-relevant channel to accomplish a demanding perceptual task and ignore the task-irrelevant channel. Moreover, the task-relevant and task-irrelevant channels were easily distinguishable according to low-level cues (spatial separation and low-level acoustic features) that expectedly supports the early selection of the task-relevant stimulation (see, e.g., Woldorff & Hillyard, 1991). Furthermore, participants indeed achieved a high performance on the main task indicating that they were engaged in the task according to the instructions. Moreover, the early temporal locus of the attentional enhancement (i.e., about 100 ms following tone onset) also supports our interpretation. Nevertheless, there is reason to suspect that voluntary attention could contribute to the present results. In Experiment 3, the selective listening task was administered with a relatively low stimulus presentation rate, that, on the positive side, supports the interpretation that our results reflect a selective enhancement of the N1 generator processes rather than activation of functionally distinct "endogenous" ERP components that can temporally overlap with the N1 when sounds are presented with short ISIs (i.e., typically in the range of 200-1000 ms; see Näätänen et al., 1978). However, on the negative side, the relatively long ISIs could allow for voluntary "lapses" of attention between the task-relevant and task-irrelevant channel and thereby contributing to the early attentional effect. In sum, while several aspects of our task settings, the behavioral results of the selective listening task, and the early temporal locus of the valence effect support the interpretation that sensory encoding of valenced tones was facilitated by "natural" involuntary attention (in the visual modality, see, e.g., Schupp et al., 2006), we also leave room for the explanation that voluntary attention could contribute to the present results. Therefore, in a further experiment, we will attempt to shed more light on the nature of the early attentional enhancement for valenced tones concerning its automaticity.

Experiment 4: Increasing the Task-Relevant Attentional Demands

In Experiment 3, we found a differential attentional enhancement for valenced compared with neutral tones at an early level of sound encoding. Although this effect likely reflects, at least partly, involuntary attentional processes, we cannot preclude the possibility that voluntary attention could contribute to the results. Hence, in Experiment 4, we address the question whether the early attentional enhancement for valenced tones can occur strictly outside of the focus of voluntary attention.

In a strict sense of automaticity, attentional enhancement for affective information occurs as an attentional capture, that is presumably mediated by subcortical structures involving the amygdala (Pourtois et al., 2013; Vuilleumier, 2005; Vuilleumier & Huang, 2009), and it is assumed to operate relatively independently of more voluntary controlled attentional processes (see, e.g., Grandjean et al., 2005; Keil et al., 2005; Pourtois et al., 2013). Consequently, if such a "reflexive" preferential attention is elicited by the valenced tones, we can expect an early attentional enhancement for these tones even in the absence of voluntary attention, thus, if participants' voluntary attention is strictly devoted to a concurrent task. Alternatively, preferential attention to affectively significant stimuli may depend – at least to some degree – on voluntary attentional processes as indicated by eliminated preferential processing of emotional facial expressions outside of the focus of voluntary attention (see e.g., Eimer et al., 2003; Eimer & Holmes, 2007, see also

Pessoa, McKenna, Gutierrez, & Ungerleider, 2002; Pessoa, Padmala, & Morland, 2005). In line with the view that prima facie automatic preferential processing of affective stimuli depends on voluntary attention, if valenced tones are presented under the circumstances in which we can assume that voluntary attention is strictly devoted to a concurrent task, we would not expect (substantial) differential attentional effect to the task-irrelevant valenced compared with neutral tones.

For Experiment 3, we cannot preclude the possibility that more voluntary controlled attentional processes could shape the pattern of results: As voluntary attention could switch between the task-relevant and the task-irrelevant channels of the selective listening task, it is possible that (some of the) task-irrelevant tones were in the focus of voluntary attention to some degree. In line with our argumentation, Lachter, Forster, and Ruthruff (2004) provided a line of empirical evidence indicating that the fact that a stimulus was irrelevant to the main task does not necessarily mean that it was also initially unattended as “lapses” (i.e., covert shifts) of attention could have occurred toward it. However, we do not assume a fixed attentional filter that is set in their model at the level of processing basic physical features, but, importantly, we share the concerns of Lachter et al. (2004) as regards (over)interpreting apparent inattention. Accordingly, in Experiment 3, both general characteristics of the selective listening paradigm (i.e., isolated stimuli, non-continuous attentional load in the task-relevant channel), and more specific characteristic of the present design (i.e., relatively slow stimulus delivery rate) may have allowed for “lapses” of attention between the task-relevant and task-irrelevant channel. Therefore, in Experiment 4, we increase the demands for voluntary attentional selection of the task-relevant channel, thereby expectedly decreasing the possibility that participants shift their attention to the task-irrelevant channel.

Although according to the selective listening literature a high rate of stimulus presentation appears to support the attentional selection of the task-relevant channel (e.g., Woldorff & Hillyard, 1991), this approach also increases the possibility that onset and offset related neural responses elicited by the task-irrelevant and task-relevant stimuli overlap in time. To overcome this possible issue, we present a continuous white noise mask in the task-relevant channel, and participants are required to detect infrequent slightly louder target noise bursts in the white noise background (that are perceived as slight, abrupt loudness increments in the ongoing stimulation). Thus, importantly, successful task performance requires constant monitoring of the continuous white noise delivered to the task-relevant ear. Hence, this task setting is expected to prevent "lapses" of attention toward the task-irrelevant channel. If early attentional enhancement for valenced tones emerges under the conditions in which we can assume that voluntary attention is not directed to these tones, we can conclude that it reflects a "pure" involuntary attentional capture. Alternatively, if the differential attentional enhancement for valenced compared with neutral tones demonstrated by Experiment 3 depends on voluntary controlled attentional processes, we can expect that this differential effect will be abolished or substantially reduced by increasing the attentional demands for concurrent task-relevant selection.

METHODS

Stimulus material, sound presentation characteristics, design, and procedure were highly similar to that used in Experiment 3. In the following, we will highlight the changes that were made between the two experiments.

Overview of the Experimental Design

Figure 4 gives an overview of the main phases of Experiment 4. Similarly to Experiment 3, Experiment 4 consisted of three main parts: (A) a valence induction phase, (B) a test phase, and (C) a manipulation check phase. The valence induction phase

(*Figure 4A*), was identical to that in Experiment 3, except for the following changes: In Experiment 4, we did not employ a no-go signal, thus, we presented only positive, negative, and neutral tones in the valence induction phase. We decided to exclude the no-go tone, and thus present only three experimental tones (with the low, middle low, and middle high tone-frequencies of Experiment 3) as we expected that three tone-frequencies can be mapped into clear representations of "low" "middle" and "high" tones, thereby supporting the acquisition of tone-to-valence associations. In order to make tone-frequency a task-relevant feature, we applied an additional tone-discrimination task during the valence induction phase (see *Procedure* for details). In the selective listening task of the test phase (*Figure 4B*), similarly to Experiment 3, participants were instructed to attend the auditory stimulation presented to one ear (task-relevant channel), while ignoring the stimuli presented to the other ear (task-irrelevant channel; thus, positive, negative, and neutral tones of the previous valence induction phase). ERPs elicited by these task-irrelevant positive, negative, and neutral tones were of the main interest. However, importantly, now a continuous white noise was presented in the task-relevant channel throughout the selective listening task. Participants' task was to detect slightly louder target noise bursts in the white noise background. We expected that this characteristic of the design necessitated monitoring the task-relevant channel continuously for successful task-performance. Finally, in the manipulation check phase (*Figure 4C*), we tested whether tones indeed acquired positive and negative valence during the valence induction, following the method of Experiment 3: First, we applied an auditory AST to assess stimulus valence indirectly; second, participants rated the valence of the tones explicitly.

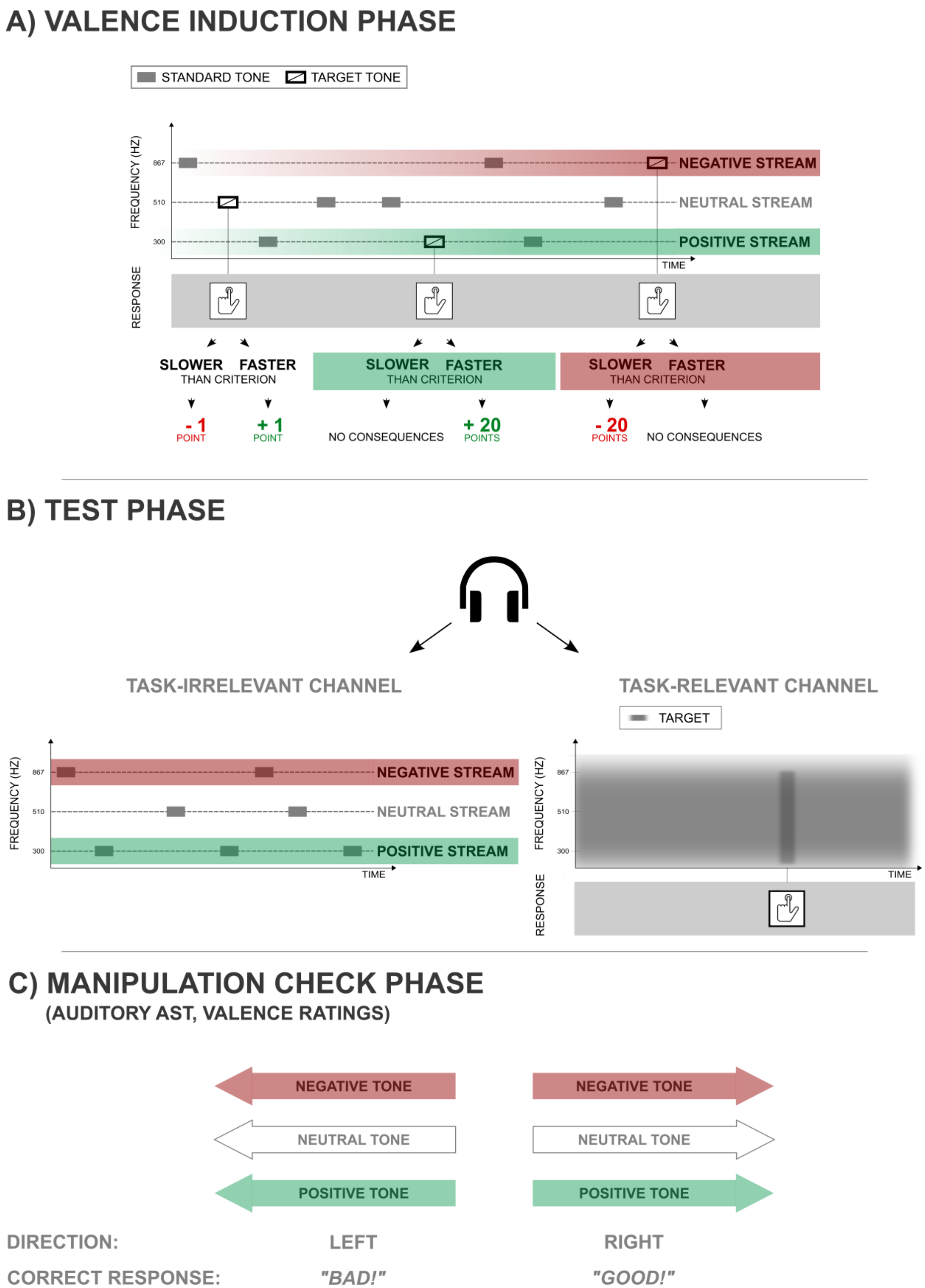

A) VALENCE INDUCTION PHASE
STANDARD TONE
TARGET TONE
FREQUENCY (HZ)
867
510
300
NEGATIVE STREAM
NEUTRAL STREAM
POSITIVE STREAM
TIME
RESPONSE
SLOWER
FASTER
THAN CRITERION
- 1
POINT
+ 1
POINT
SLOWER
FASTER
THAN CRITERION
NO CONSEQUENCES
+ 20
POINTS
SLOWER
FASTER
THAN CRITERION
- 20
POINTS
NO CONSEQUENCES
B) TEST PHASE
TASK-IRRELEVANT CHANNEL
TASK-RELEVANT CHANNEL
TARGET
FREQUENCY (HZ)
867
510
300
NEGATIVE STREAM
NEUTRAL STREAM
POSITIVE STREAM
TIME
RESPONSE
C) MANIPULATION CHECK PHASE
(AUDITORY AST, VALENCE RATINGS)
NEGATIVE TONE
NEGATIVE TONE
NEUTRAL TONE
NEUTRAL TONE
POSITIVE TONE
POSITIVE TONE
DIRECTION:
LEFT
RIGHT
CORRECT RESPONSE:
"BAD!"
"GOOD!"

Figure 4. Schematic illustration of the design of Experiment 4.

Note that only one instance of the counterbalancing scheme is depicted. (A) In the valence induction phase, valence was induced experimentally in a game paradigm. (B) In the test phase, positive, neutral and negative tones were presented in a task-irrelevant channel. In the task-relevant channel, a continuous white noise was presented throughout the selective listening task. Participants' task was to detect slightly louder target noise bursts in the white noise background. For successful task-performance, participants had to monitor the task-relevant channel continuously. (C) In the manipulation check phase, the effectiveness of the valence induction was tested in an auditory AST and by collecting explicit valence ratings.

Participants

Twenty-four students from Saarland University (11 females; aged 18–29 years, *Mdn* = 22 years; three left-handers) participated in the experiment for monetary compensation. The assignment of tone-frequencies to valence conditions was counterbalanced between participants according to a Latin square scheme resulting in three counterbalancing groups (*Table A2* in *Appendix B* depicts the exact counterbalancing scheme). Given a sample size of $N = 24$, and an α-value of .05 (one-tailed), the effect size of $|d_z| = 0.51$ (representing the most relevant valenced-minus-neutral difference on the N1 amplitude in Experiment 3, $M = -0.49$ µV, $SD = 0.96$ µV) can be detected with a probability of $1 - \beta = .78$ (i.e., approximately at the recommended .80 level according to the guidelines of Cohen, 1988; calculated with the aid of G*Power 3 software, Faul et al., 2007)[11].

All participants were native German speakers, had normal or corrected-to-normal vision, and self-reported normal hearing. Before the experiment, all participants gave written informed consent. On average, participants received €34 (compensation was partly dependent on performance). In the final sample, each counterbal-

[11] Note that when increasing sample size, a minimum sample size of $N = 30$ would be necessary to equally complete all cells of our full counterbalancing design.

ancing group had eight participants. In the manipulation check phase, twenty-three students participated as one participant had to be excluded due to technical failure of the voice key apparatus. One further participant was excluded during data analysis of the AST because of generally slow RTs (see below in *Results* section).

Materials

Sinusoidal tones with three different tone-frequencies (300 Hz, 510 Hz, 867 Hz; thus, the low, middle low, and middle high tones of Experiment 3, respectively) were presented in Experiment 4.

Valence induction phase

Materials of the valence induction phase were identical to that in Experiment 3, except for not employing no-go tone in the present experiment.

Test phase

In the test phase, a continuous white noise was presented in the task-relevant channel throughout the selective listening task with an intensity level of approximately 37 dB SL. As targets, additional white noise bursts were presented with a duration of 200 ms and intensity level of approximately of 39 dB SL[12]. Stimuli presented in the task-irrelevant channel were identical to that presented in Experiment 3 (except for not using the highest tone-frequency in any of the counterbalancing groups).

[12] Target intensity was set according to three pilot sessions to make the target noise bursts slightly above the threshold of detectability.

Manipulation check phase

In the auditory AST, amplitudes of the positive, negative, and neutral experimental tones and the filler sounds were modulated in an identical way as in Experiment 2a and Experiment 3. As filler trials, 18 natural sounds from the IADS battery (Bradley & Lang, 2007) were presented (six positive, six negative, and six neutral sounds). Averaged normative valence ratings on a 9-point scale ranging from most unpleasant (1) to most pleasant (9) were: M = 7.16 (SD = 0.37) for positive, M = 2.19 (SD = 0.54) for negative, and M = 4.75 (SD = 0.37) for neutral sounds, respectively. From this pool of natural sounds, one positive, one negative, and one neutral sound were selected for each participant in a way that each sound is presented four times in the planned final sample of twenty-four participants.

Procedure

Valence induction phase

In contrast to Experiment 3, in Experiment 4, we did not employ no-go signal, thus, we presented only positive, negative, and neutral tones (*Figure 4A*). A valence induction block comprised 48 auditory trials presented in a random order. 14 positive, 14 negative, 14 neutral tones were presented (half of them in their standard, and half of them in their target version, respectively). Similarly to Experiment 3, tone-frequency defined the consequences of success and failure on a given trial.[13] To ensure that participants encode tone-frequency as a task-relevant feature, six additional tone-discrimination trials were presented randomly in each valence induction block. A tone-discrimination trial started with the presentation of a black fixation cross with a

[13] To ensure that participants experience success and failure trials in a relatively balanced amount, we again used the moving median of the preceding six trials and the participant's current total game score to define the actual response speed criterion. However, we used the following modified formula in Experiments 4-5: median' = median - 0.2 × current total game score, in order to reflect the participant's game score in a more fine-grained manner.

randomized duration of 500-1000 ms. Thereafter, one of the tones was played binaurally (throughout a valence induction block, each of the three tones was presented once on a tone-discrimination trial in their standard and target form, respectively). After 400 ms following the sound onset, the black fixation cross turned red and remained on the screen during the sound presentation (i.e., 1000 ms long). Thereafter, participants had to choose the "game-meaning" of the presented tone by clicking to the corresponding term on the screen (i.e., "danger", "chance", or "neutral", see also the *Procedure* section of Experiment 3). Error feedback was presented visually if the participant responded incorrectly or did not respond within 1500 ms. Failure on a tone-discrimination trial was associated with a penalty of 10 points. No additional visual detection task was employed in Experiment 4.

Test phase

In the test phase (*Figure 4B*), importantly, a continuous white noise was presented monaurally in the task-relevant channel throughout the selective listening task. Additionally, 15 target white noise bursts were presented monaurally in the task-relevant channel in each test block in random temporal positions with a minimum ISI of 500 ms. Participants had to respond to the target noise bursts (perceived as abrupt loudness increments in the ongoing stimulation) as quickly and accurately as possible by pressing the space bar. Visual feedback was given about the mean HR after each block in order to motivate participants to maintain a high level of performance. Similarly to Experiment 3, positive, neutral, and negative tones were presented monaurally in the to-be-ignored task-irrelevant channel. Task channel and ear assignment was counterbalanced between participants.

Manipulation check phase

In the manipulation check phase of Experiment 4 (*Figure 4C*), not only the experimental trials, but also the filler trials (i.e., natural sounds) were presented 10 times each in random order (five times in “moving to the left” and five times in “moving to the right” version).

EEG Recording and Analysis

EEG recording, filtering, segmentation of the continuous EEG, artifact rejection, baseline correction, choice of ROI, and amplitude calculations were identical to the methods that were applied in Experiment 3. Epochs containing task-relevant target noise bursts were discarded from the analysis (rejection criteria: onset of a task-relevant target noise burst within -1000 ms to +800 ms of the task-irrelevant tone onset). On average, ERPs were based on 248, 247, and 249 trials per participant in the positive, neutral, and negative valence conditions, respectively. Epochs were averaged separately for the each valence condition.

Design and Statistical Analysis

For testing the valence effect, we applied a 3 × 3 mixed design with valence (positive, neutral, negative) as a within-participants factor and tone-frequency-to-valence assignment (three counterbalancing groups) as a between-participants factor on the amplitudes for the ERP components of interest and the behavioral measures of the valence induction phase. Similarly to Experiment 3, we added counterbalancing group as a between-participants factor to use the correct error term (see, Pollatsek & Well, 1995; and Appendix A). In line with Experiment 3, we used the multivariate approach to repeated measures analysis, which means that the tripartite factor of valence was transformed into a vector of two orthogonal contrast variables (see, e.g., O’Brien

& Kaiser, 1985; Petrova & Wentura, 2012; M. Rohr et al., 2012). Similarly to Experiment 3, we applied a priori chosen contrasts: (1) for the first contrast, amplitude values were averaged across positive and negative stimuli and contrasted with the neutral stimuli (i.e., testing for the general relevance bias hypothesis). (2) The second orthogonal contrast was the contrast between positive and negative tones (i.e., testing for the negativity bias hypothesis). AST data and valence ratings were analyzed in line with Experiment 3.

RESULTS

A significance level of $\alpha = .05$ (two-tailed, unless otherwise noted) was adopted for all analyses.

Behavioral Performance

Behavioral performance was adequate on the valence induction and selective listening tasks and reflected strategic differentiation between valence conditions during valence induction, suggesting that participants were engaged in both tasks and followed the task instructions. The behavioral results of the valence induction phase are presented in *Table 8*.

Table 8. Behavioral results of the valence induction phase in Experiment 4: Mean reaction times (RTs), accuracy rates (ACCs), false alarm rates (FARs), and hit rates (HRs) in each valence condition; *SD* in parentheses.

Valence	**RT**	**ACC**	**FAR**	**HR**
Positive	343 (48)	89.3 (12.9)	19.0 (26.3)	97.6 (6.6)
Neutral	348 (33)	94.0 (6.5)	8.6 (10.3)	96.5 (6.0)
Negative	337 (34)	94.5 (5.4)	8.5 (8.2)	97.4 (6.0)

RTs below 100 ms were discarded when calculating averaged RTs. In the valence induction task, RTs were calculated from the beginning of the loudness increase in the ongoing tone. A 3 × 3 MANOVA for repeated measures with valence as within-participants factor and counterbalancing group as between-participants factor of the RTs showed a marginally significant valence main effect, $F(2,20) = 2.97$, $p = .074$, $\eta_p^2 = .229$. The comparisons of valenced vs. neutral conditions and positive vs. negative valence were not significant, $F(1,21) = 2.44$, $p = .128$, $\eta_p^2 = .104$, and, $F(1,21) = 1.27$, $p = .274$, $\eta_p^2 = .057$, respectively.

The average ACC across participants was adequately high ($M = 92.6\%$, $SD = 7.3\%$). Similarly to Experiment 3, lowest mean ACC emerged in the positive condition (see *Table 8*). ACCs showed a marginally significant valence main effect, $F(2,20) = 3.02$, $p = .071$, $\eta_p^2 = .232$. ACC was indeed significantly lower for positive compared with negative tones, $F(1,21) = 6.12$, $p = .022$, $\eta_p^2 = .226$. The valenced vs. neutral comparison was marginally significant, $F(1,21) = 4.02$, $p = .058$, $\eta_p^2 = .161$. Similarly to Experiment 3, in Experiment 4, the relatively low mean ACC in the positive condition emerged also due to high FARs (see *Table 8*; note that in this condition false alarms had no negative consequences). FARs showed a marginally significant valence main effect, $F(2,20) = 3.13$, $p = .066$, $\eta_p^2 = .238$. As expected, participants made significantly more false alarms in the positive compared with negative condition, $F(1,21) = 5.67$, $p = .028$, $\eta_p^2 = .213$. The valenced vs. neutral comparison was also significant, $F(1,21) = 5.73$, $p = .026$, $\eta_p^2 = .214$. Analogue analysis on the HRs did not show any differences, all Fs < 1.05, *n.s.* Furthermore, participants could discriminate the "game-meaning" (i.e., "chance", "neutral", and "danger") of the presented tones with high precision as indexed by high average ACC on the tone-discrimination trials ($M = 93.8\%$, $SD = 5.4\%$).

In the selective listening task of the test phase, the average RT across participants was 367 ms ($SD = 24$ ms) with an average ACC of 78.4% ($SD = 12.9\%$).

ERP Results

Of most interest, we analyzed the amplitude of the auditory N1 elicited by the task-irrelevant positive, negative, and neutral tones during the test phase. Prototypical P1-N1-P2 waveform was clearly observable in the group average ERPs (see *Figure 5*, and for the mean amplitudes for the components of interests, see *Table 9*).

A 3 × 3 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and counterbalancing group as between-participants factor of the N1 amplitudes (measured as mean voltage in a 20-ms window centered at the latency of the group-average peak, i.e., at 144 ms) did not yield significant valence main effect, $F(2,20) = 1.38$, $p = .275$, $\eta_p^2 = .121$. The a priori contrast of valenced (positive and negative) vs. neutral condition showed a tendency for difference, $F(1,21) = 2.23$, $p = .075$ (one-tailed)[14], $\eta_p^2 = .096$; with more negative N1 amplitudes in the valenced conditions compared with the neutral condition (M = -4.33 µV, SD = 1.68 µV in the valenced condition; M = -3.98 µV, SD = 2.00 µV in the neutral condition). The a priori contrast of the two valenced conditions (positive vs. negative) was not significant, $F<1$, *n.s.*

In line with Experiment 3, we performed a 3 × 3 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and counterbalancing group as between-participants factor of the P1 amplitudes measured as mean voltage in a 20-ms window centered at the latency of the group-average peak (62 ms, experimental conditions collapsed). This analysis yielded no valence main effect, $F(2,20) = 0.41$, $p = .668$, $\eta_p^2 = .040$. The a priori contrast of valenced (positive and negative) vs. neutral condition and the a priori comparison of the two valenced conditions (positive vs. negative) did not show any difference, Fs<1, *n.s.*

---

[14] A one-tailed interpretation can be applied given the equivalence of an F-test with df_N=1 (here, our a priori contrast) with a two-tailed t-test and given our specific prediction (i.e., we had the directed hypothesis of enhanced amplitudes to valenced compared with neutral tones for the ERP components of interest that are expected to be augmented by attention; see, Maxwell & Delaney, 1990, p. 144). Furthermore, we had the directed hypothesis of a positive AS effect (see below).

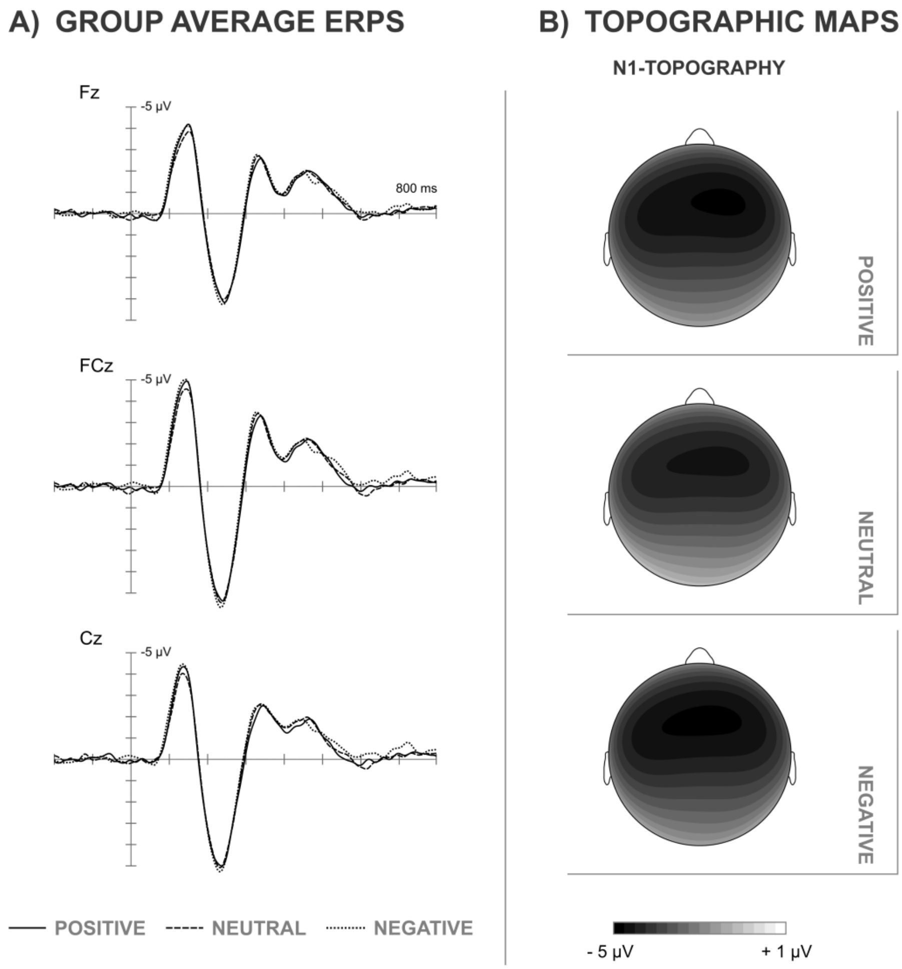


Figure 5. ERP results of Experiment 4.

(A) Group average ERP waveforms to the positive, neutral, and negative tones on the representative Fz, FCz and Cz electrode sites in Experiment 4. The physical onset of the tones is at the crossing of the axes (0 ms). Negative polarity is plotted upwards. (B) Group average topography maps in the N1 time window (134-154 ms) in positive, neutral and negative conditions.

Table 9. ERP results of Experiment 4. Mean P1 and N1 amplitudes, and P1-N1 peak-to-peak amplitudes (in µV) in each valence condition; *SD* in parentheses.

Valence Condition	**P1**	**N1**	**P1-N1**
Positive	0.23 (0.80)	-4.30 (1.87)	-5.72 (2.14)
Neutral	0.06 (0.80)	-3.98 (2.00)	-5.34 (1.69)
Negative	0.24 (0.75)	-4.36 (1.75)	-5.88 (1.96)

In order to better quantify the apparent relative difference between the P1 and N1 peaks as indicated by *Figure 5A*, we applied peak-to-peak amplitude measurement for the P1 and N1 peaks (i.e., using the adjacent peak as a reference, e.g., Handy, 2005). The auditory P1 and N1 are typically elicited together, thereby often termed as P1-N1 complex (Burkard, Eggermont, & Don, 2007); and there is evidence indicating that auditory attention can enhance both components at least in the case of highly focused attention (e.g., Woldorff & Hillyard, 1991). Peak-to-peak amplitudes were derived for each participant as the voltage difference between the P1 and N1 peaks (P1 and N1 peaks were identified in the individual ERPs as the strongest positive/negative local peaks within a 40/60 ms long time window centered around the latency of the group-average P1/N1 peaks). A 3 × 3 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and counterbalancing group as between-participants factor of the P1-N1 peak-to-peak amplitudes showed a marginally significant valence main effect, $F(2,20) = 2.74$, $p = .089$, $\eta_p^2 = .215$. The a priori contrast of valenced (positive and negative) vs. neutral condition was significant, $F(1,21) = 4.43$ $p = .024$ (one-tailed), $\eta_p^2 = .174$, indicating enhanced P1-N1 amplitudes for valenced compared with neutral tones ($M = -5.80$ µV, $SD = 1.86$ µV in the valenced condition; and $M = -5.34$ µV, $SD =$

1.69 μV in the neutral condition). The contrast of the two valenced conditions (positive vs. negative) was not significant, $F<1$, *n.s.*

In line with Experiment 3, we conducted a further analysis in the later time range (350-500 ms post onset) at the frontocentral region. A 3 × 3 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and balancing group as a between-participants factor on the mean activity measured in the 350-500 ms time range at the frontocentral ROI did not show any valence difference, $F(2,20) = 0.02$, $p = .977$, $\eta_p^2 = .002$. (The mean amplitude was $M = -1.65$ μV, $SD = 1.23$ μV in the positive condition; $M = -1.64$ μV, $SD = 1.36$ μV in the negative condition; and $M = -1.68$ μV, $SD = 1.16$ μV in the neutral condition.)

Manipulation Check

Affective Simon task

Analysis of the AST data was in line with that in Experiment 3: RTs were calculated from the beginning of the illusory movement in the ongoing tone. RT analyses were restricted to correct trials (5.6% of the trials was excluded because of erroneous response of the participant or erroneous or non-reaction of the voice key). As an a priori criterion, RTs below 200 ms and above 2000 ms were discarded from further analyses (4.3% of the trials), and RT outliers were excluded (1.1% of the trials; for each participant an upper and a lower outlier criterion were calculated according to Tukey, 1977, using the distribution of all AST RTs of this person). One participant was excluded during data analysis because of an insufficient number of trials remaining after this procedure (73% of the trials were excluded due to generally high RTs, individual mean RT was 2836 ms). Mean RTs and ERs for congruent, incongruent, and neutral trials, as well as the mean AS effect are presented in *Table 10* for the experimental and filler trials, respectively. On the experimental trials, the incongruent-minus-congruent RT difference showed the expected AS effect, $t(21) = 1.77$, $p = .046$ (one-tailed), $d_z = 0.38$. However, similar analysis on the filler trials (natural sounds) did not yield significant results, $|t| < 1$, *n.s.* Analogue analyses on the ERs did not show significant AS effects for the experimental trials, $t(21) = 1.16$, $p = .129$ (one-tailed), $d_z = 0.25$. Similarly, there was no significant AS effect on the ERs for the filler trials, $t(21) = 1.30$, $p = .104$ (one-tailed), $d_z = 0.28$.

Table 10. Mean RTs (in ms) and mean ERs (%) as a function of stimulus and response valence congruency for experimental and filler trials in the AST of Experiment 4; *SD* in parentheses.

	Experimental Tones		**Fillers**	
	RT	**ER**	**RT**	**ER**
Neutral	917 (316)	6.4 (10.5)	814 (233)	6.4 (11.8)
Congruent	880 (273)	3.6 (6.6)	802 (240)	4.6 (11.4)
Incongruent	912 (281)	5.5 (8.6)	799 (251)	7.3 (10.3)
AS Effect[a]	32 [18]	1.8 [1.6]	-3 [19]	2.7 [2.1]

[a] incongruent-minus-congruent difference; standard errors in brackets

Valence ratings

A 3 × 3 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and counterbalancing group as between-participants factor of the valence ratings did not show valence main effect, $F(2,19) = 0.01$, $p = .994$, $\eta_p^2 = .001$ (for the mean ratings, see *Table 11*).

Table 11. Mean valence ratings in each valence condition for low, intermediate, and high pitch in Experiment 4. Valence ratings range from very unpleasant (1) to very pleasant (9); *SD* in parentheses.

	Pitch			
Valence	**Low**	**Intermediate**	**High**	**All**
Positive	4.50 (1.20)	5.38 (1.69)	2.43 (1.51)	4.10
Neutral	5.62 (1.06)	3.86 (1.22)	2.75 (1.17)	4.08
Negative	4.29 (1.80)	3.88 (1.46)	4.00 (2.00)	4.05
All	4.80	4.37	3.06	

In line with the analysis of the valence ratings of Experiment 3, we applied the approach of linear mixed models to assess whether valence has an effect above the linear effect of pitch that was conspicuous in the present data as well. We used the *lme4* and *lmerTest* packages of R 3.1.3 (Bates et al., 2014) with significance of predictors assessed using Satterthwaite's approximation for degrees of freedom (Kuznetsova et al., 2014). First, we ran two random effects model with pitch (-1 = lowest, 0 = medium, +1 = highest) as predictor and rating as dependent variable. Model 1 included random intercepts and random slopes; Model 2 included only random intercepts. Model comparison yielded no significant difference, that is, the removal of the parameter for by-participant random slopes for pitch is justified, $\chi^2(2) = 0.19$, $p = .911$ (see, e.g., Baayen et al., 2008). Second, we ran two further models on the basis of Model 2: In Model 3 we added valence (-1 = negative, 0 = neutral, +1 = positive) as a predictor including random slopes; in Model 4, we removed the random slopes parameter. Model comparison yielded no significant difference between Models 3 and 4, $\chi^2(2) = 1.09$, $p = .581$, thus, Model 4 can be considered our final model.

However, in this model, only pitch was significant predictor of the rating, B = -0.87, $t(44)$ = -4.61, p < .001 for pitch; while B = 0.03, $t(44)$ = 0.15, p = .885 for valence.

DISCUSSION

In Experiment 4, we increased the demands for voluntary attentional selection of the task-relevant auditory stimuli in order to reduce "lapses" of attention to the task-irrelevant positive, negative, and neutral tones. To reach this aim, we modified the selective listening task in the following way: In the task-relevant channel, a continuous white noise was presented, and participants' task was to detect infrequent slightly louder target noise bursts in the white noise background. In consequence, successful task performance required constant monitoring of the continuous white noise. Positive, negative, and neutral tones were presented in a task-irrelevant channel, and we can assume that they were not in the focus of voluntary attention in the present task design.

In line with our effort to increase task demands, we indeed observed substantially lower ACC on the selective listening task compared with Experiment 3 (93.2% in Experiment 3 vs. 78.4% in Experiment 4). Moreover, despite the perceptually challenging task, mean RT was considerably shorter in the present experiment (590 ms in Experiment 3 vs. 367 ms in Experiment 4) with markedly lower standard deviation (74 ms in Experiment 3 vs. 24 ms in Experiment 4). This pattern is consistent with the interpretation that in the present experiment participants focused their attention constantly to the continuous stimulation delivered to the task-relevant ear rather than switching their attention between channels. In Experiment 3, however, participants' attentional focus might have been on the task-irrelevant channel on some of the trials, and prolonged RTs reflect the time cost of an attentional shift to the task-relevant channel when a target was presented.

Regarding the ERP results, under the present task conditions, we found only marginally significant evidence in our standard analysis on the N1 amplitude for a differential attentional enhancement of valenced compared with neutral tones. However, there is a further source of evidence that points toward an early valence-related attentional enhancement: We found an indication that the amplitude of the P1-N1 complex was enhanced for valenced compared with neutral tones, thereby indicating a moderate but even earlier attentional effect compared with what was found in Experiment 3. Although the N1 is often considered to be the earliest component of the auditory ERP that shows reliable sensitivity to manipulations of participants' attention (Herrmann & Knight, 2001), there are evidences that enhancement of auditory ERP components occurring before the N1 – including the P1 – can index very early attentional processes (e.g., Fritz et al., 2007; Woldorff & Hillyard, 1991). Thus, we can assume that P1 and N1 components reflect attentional processes operating at an early stage of stimulus encoding; hence, they can be enhanced together in the case of an attentional effect with an early temporal locus. Consequently, both the time course of the valence effect and the present task characteristics support the interpretation that the differential enhancement of early auditory components to valenced compared with neutral tones reflects involuntary attention to valenced tones.

As mentioned above, although we found significant valence effect on the P1-N1 complex, indication for a differential attentional effect on the N1 amplitude in our standard analysis was weaker in Experiment 4 compared with Experiment 3: The most relevant valenced-minus-neutral difference on the N1 amplitudes was associated with the effect size of $|d_z| = 0.30$ ($M = -0.35$ µV, $SD = 1.15$ µV) in Experiment 4, compared with the effect size of $|d_z| = 0.51$ ($M = -0.49$ µV, $SD = 0.96$ µV) in Experiment 3. Furthermore, we also did not find a differential effect for the valenced compared with neutral tones on the later sustained negativity that possibly indexes attentional control processes.

In line with this pattern, one can claim that we did not only reduce voluntary attention switches between the channels in the present paradigm but, by presenting a continuous white noise in the task-relevant channel, we might have considerably limited available attentional resources more generally, that in turn could contribute to a weakening of the valence effect in the standard analysis. In line with this argumentation, a remarkable line of research suggests that the addition of perceptual load to the main task prevents involuntary "spilling over" of attention for processing task-irrelevant stimuli (for visual demonstrations, see, e.g., Forster & Lavie, 2007, 2008; Lavie, Hirst, de Fockert, & Viding, 2004; Rees, Frith, & Lavie, 1997; for reviews, see Lavie, 2005, 2010; Lavie, Beck, & Konstantinou, 2014; for visual load induced "inattentional deafness", see Macdonald & Lavie, 2011; and Raveh & Lavie, 2014, suggesting that the phenomenon is not restricted to visual attention). The perceptual load theory of attention assumes that task sets with low perceptual load leave room for involuntary "spilling over" of attentional resources for processing task-irrelevant distractor stimuli, and, as a result, in these settings successful task-relevant selection requires active cognitive control (i.e., promoting late selection effects). On the contrary, tasks involving high perceptual load fully engage attentional resources and leave no spare capacity for the task-irrelevant stimuli (i.e., promoting early selection effects; see, Lavie, 2005, 2010; Lavie et al., 2014). Moreover, increasing the amount of perceptual stimulation in the task appears to eliminate so well-established results as the prima facie automatic differential amygdala response to task-irrelevant emotional facial expressions compared with neutral faces (Bishop, Jenkins, & Lawrence, 2007; Pessoa et al., 2002; see also Gupta, Hur, & Lavie, 2015). However, contradictory to an explanation that the continuous noise in the task-relevant channel prevented involuntary "spilling over" of attention to the task-irrelevant tones, N1-amplitudes were generally higher in the present experiment compared with Experiment 3, while the perceptual load theory of attention would predict generally weaker sensory responses to the task-irrelevant stimulation with perceptual load (see, e.g., Handy, Soltani, & Mangun, 2001). Overall enhanced sensory responses in the present experiment can reflect more general differences between the test phases of Experiment 3 and Experiment 4, like heightened level of non-selective arousal or alertness related to the higher task demands (amplitude of the auditory N1 is larger at higher level of general alertness even outside of the focus of voluntary attention; for a review see, Näätänen & Picton, 1987). We will return to this issue in the *Intermediate Summary* below.

To sum it up, if we consider only our standard analysis, a substantial weakening of the valence effect in Experiment 4 would support the account that the valence-related attentional enhancement depends to some degree on voluntary attention. However, importantly, a combined analysis of our experiments did not confirm a weakening of the valence effect on the N1 amplitude statistically. We will return to this point in details below in the chapter on a combined analysis of Experiments 3-5.

Similarly to the results of Experiment 3, indications for an attentional enhancement of early auditory components were comparable for positive and negative valence but did not show any differentiation between valence categories. Hence, results of Experiment 4 are also in accordance with a general relevance hypothesis of affective attentional biases, that is, that valenced tones receive preferential attention based on their higher relevance compared with neutral tones rather than based on a specific valence category. Contrary to the early ERP results, both implicit and explicit behavioral measures differentiated between positive and negative valence. In the valence induction phase, similarly to Experiment 3, participants employed the most liberal response criteria for positive tones as reflected in a high rate of false alarms. Moreover, the high ACC achieved on the tone-discrimination trials indicates that participants could identify the valence-related "meaning" of the tones with high precision, thereby suggesting a high level of contingency learning. In the manipulation check phase, we found evidence that the simple tones acquired positive and negative affective valence during the learning phase, at least implicitly, as the auditory AST showed the expected AS effect on the RTs for the valenced tones. However, natural emotional sounds of the filler trials did not reveal an AS effect either on the RTs nor the ERs, although they were presented balanced for movement direction in the present experiment. Nonetheless, we have to keep in mind that the manipulation check phase was administered at the end of a long and even more demanding experimental session compared with Experiment 3. Consequently, similarly to Experiment 3, we presented relatively few AST trials (10 filler trials in each valence condition in the present experiment vs. 20 trials in each valence condition in Experiment 2). Importantly, in the independent Experiment 2, we demonstrated the expected AS effect using natural emotional sounds. The explicit valence ratings did not reveal significant differences concerning the newly acquired valence in the present experiment; however, they showed a highly similar pattern to that in Experiment 3: They reflected dominantly the most salient stimulus feature, that is, pitch.

To sum up, in Experiment 4, we found indication for an early attentional enhancement for task-irrelevant valenced tones as reflected by the augmented P1-N1 complex, and we also found a marginally significant indication for differential valence effect on the N1 amplitude in our standard analysis. Importantly, this early attentional enhancement emerged under the conditions in which strict voluntary selection of the concurrent task-relevant stimulation was promoted. Taken together, we interpret the pattern of results (together with the results emerged from a combined analysis of Experiments 3-5 that will be introduced below) as involuntary attention to affectively significant auditory information that can operate outside of the focus of voluntary attention.

Experiment 5: Moderation by Anticipation of Positive or Negative Future Outcomes?

In Experiments 3 and 4, we targeted the question whether rapid attentional biases to valenced auditory stimuli are driven by their general relevance (thus, both positive and negative tones are prioritized relative to neutral ones given their higher goal-relevance regardless of valence category; in the visual domain, see, e.g., Brosch et al., 2008; Keil & Ihssen, 2004; Müller et al., 2015; Nummenmaa et al., 2009; Wentura et al., 2014); or our auditory attentional system is tuned to selectively prioritize negative information, as detection of dangers possesses arguably higher survival value compared with detection of opportunities (in the visual domain, see, e.g., Carlson et al., 2012; Esteves, Dimberg, & Öhman, 1994; Ito, Larsen, Kyle, & Cacioppo, 1998; Öhman et al., 2001; Öhman, 2005; Öhman et al., 2012; Schupp et al., 2004). The pattern of results that emerged from Experiments 3-4 was clear concerning this question: While explicit and implicit behavioral results differentiated between the newly acquired positive and negative valence of the tones, ERP results suggested that at the early level of sound encoding the general relevance of the valenced tones that elicits preferential attention instead of a specific valence category. A third – although not exclusive – possibility is that affective biases in early auditory attention are not static and undifferentiated for specific valence categories, but they can promote asymmetric facilitation between positive and negative valence according to current affective-motivational demands. The counter-regulation principle (Rothermund et al., 2008) proposes a highly adaptive mechanism of affective attentional biases: In order to promote a homoeostatic regulation of affective-motivational states, attentional biases to valent stimuli operate incongruently to the current motivational-emotional orientations, thereby preventing escalation or perseveration of affective-motivational states. Accordingly, the counter-regulation principle predicts a bias for positive information in negative affective-motivational context, and a bias for negative information in positive affective-motivational context. Moreover, the assumption of a counter-regulation in attentional biases to valent visual stimuli has received empirical support not only regarding previously experienced positive or negative events (Rothermund, 2003; Rothermund et al., 2011; Schwager & Rothermund, 2014), but also when anticipating positive or negative future outcomes (Rothermund et al., 2008; Schwager & Rothermund, 2013b; Wentura et al., 2009; see also Rothermund et al., 2001).

In accordance with the assumption of flexible affective attentional biases, it is possible that in salient positive or negative motivational contexts negative and positive tones can receive differential attentional enhancement already at the early level of sound encoding. However, as motivational demands were rather ambiguous in the test phase of Experiments 3-4, one can argue that a differential attentional weighting mechanism for specific valence categories could not have been manifested. Hence, in Experiment 5, we introduce a further manipulation in order to promote unequivocal motivational focus: In the positive and negative outcome blocks, we promote the motivational focus of anticipating positive and negative future outcomes by the prospect of substantial monetary reward and substantial monetary loss, respectively. If a counter-regulation principle operates on the attentional biases to valent information at the early stage of sound encoding, we can expect enhanced attention reflected in enhanced N1 for positive compared with negative tones when anticipating negative future outcomes, and in turn enhanced attention reflected in enhanced N1 for negative compared with positive tones when anticipating positive future outcomes. However, if positive and negative tones are facilitated by early attention in an undifferentiated way (i.e., in line with a rather fixed general relevance hypothesis), we can expect a similar pattern of results as in Experiments 3-4, that is, enhanced attention to valenced tones compared with neutral ones without moderation by motivational outcome focus.

METHODS

Stimulus material, sound presentation characteristics, design, and procedure were highly similar to that used in Experiment 3. In the following, we will highlight the changes that were made between the two experiments.

Overview of the Experimental Design

Figure 6 gives an overview of the main phases of Experiment 5. The experiment consisted of two main parts: (A) a valence induction phase, and (B) a test phase. Manipulation check phase was not administered in Experiment 5 in order to keep the duration of an experimental session within acceptable range (the duration of the test phase was increased in the present experiment in order to increase the number of trials per experimental condition for the averaged ERPs; see *Procedure*). Importantly, Experiments 3-4 indicated that the applied valence induction method secures the acquisition of valence. The valence induction phase (*Figure 6A*) was identical to that in Experiment 3, except that the no-go signal was associated with the highest tone-frequency in all counterbalancing group to reduce the complexity of the design. The selective listening task of the test phase (*Figure 6B*) was identical to that in Experiment 3, except that we additionally introduced positive and negative outcome focus manipulation in the present experiment.[15] Thus, positive outcome blocks were associated with the chance of substantial monetary reward, while negative outcome blocks were associated with the danger of substantial monetary loss at the end of a selective listening block depending on the participants' performance on the task-relevant channel.

[15] We did not employ a continuous noise in the task-relevant channel to not further increase task difficulty as the complexity of the selective listening task was already increased by the outcome focus modulation.

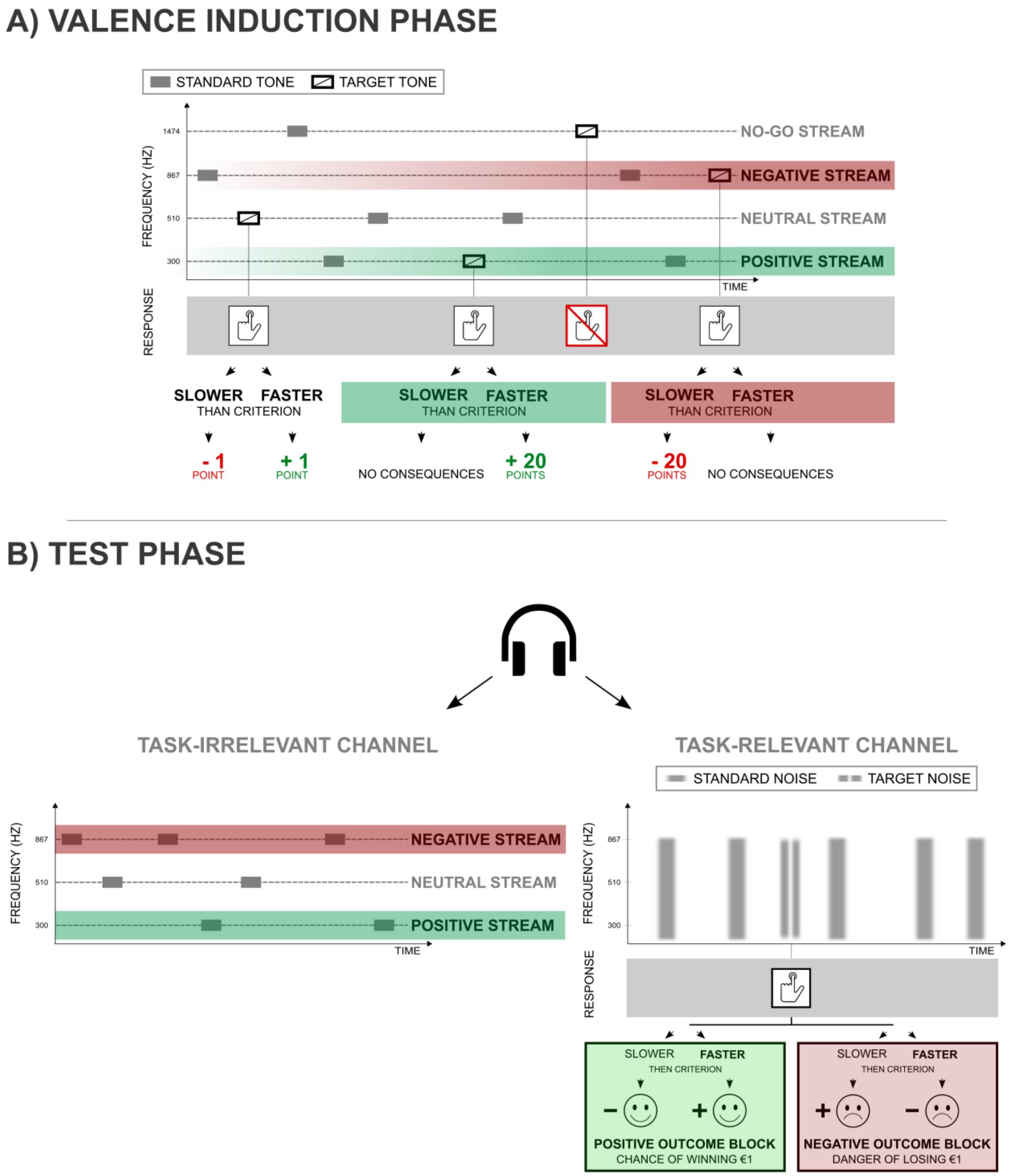


Figure 6. Schematic illustration of the design of Experiment 5.

Note that only one instance of the counterbalancing scheme is depicted. (A) Valence was induced experimentally in a game paradigm in the valence induction phase. (B) In the selective listening task of the test phase, positive, neutral and negative tones were presented in a task-irrelevant channel, while a perceptual detection task was administered in the task-relevant channel. Participants were instructed to direct their attention to the task-relevant channel, while ignoring the task-irrelevant channel. Additionally to the design of Experiment 3, positive and negative outcome focus manipulation was introduced: Positive outcome blocks were associated with the chance of substantial monetary reward, while negative outcome blocks were associated with the danger of substantial monetary loss at the end of the block depending on the task performance.

Participants

Twenty-four students from Saarland University (11 females; aged 18–29 years, *Mdn* = 23 years; two left-handers) participated in the experiment for monetary compensation. The assignment of tone-frequencies to valence conditions was counterbalanced between participants according to a Latin square scheme resulting in three counterbalancing groups (see *Table A2* in *Appendix B* for the exact counterbalancing scheme). Given a sample size of $N = 24$, and an α-value of .05 (one-tailed), the effect size of $|d_z| = 0.51$ (representing the most relevant valenced-minus-neutral difference on the N1 amplitude in Experiment 3; we oriented ourselves on the results of Experiment 3 that had comparable characteristics) can be detected with a probability of $1 - \beta = .78$ (i.e., approximately at the recommended .80 level according to the guidelines of Cohen, 1988; calculated with the aid of G*Power 3 software, Faul et al., 2007).[16] All participants were native German speakers, had normal or corrected-to-normal vision, and self-reported normal hearing. Before the experiment, all participants gave written informed consent. On average, participants received €30.5 (compensation was partly dependent on performance; participants received €12.5 on average during the test phase, see below).

[16] Note that when increasing sample size, a minimum sample size of $N = 36$ would be necessary to equally complete all cells of our full counterbalancing design.

Materials

Similarly to Experiment 3, sinusoidal tones with four different tone-frequencies (300 Hz, 510 Hz, 867, and 1473.9 Hz) were presented in Experiment 5.

Valence induction phase

Materials were identical to that applied in Experiment 3.

Test phase

In the test phase, we employed identical auditory stimuli and sound presentation characteristics as in Experiment 3. Furthermore, to emphasize the motivational character of the negative and positive outcome focus blocks, we presented additional visual feedback: schematic faces depicting smiling and sad emotional expressions. In line with the experimental instructions, we will refer to these schematic faces as "smileys" and "saddies", respectively. The size of these images was 100 x 100 pixels.

Procedure

The procedure was identical to that in Experiment 3, except for the changes outlined below. For the valence induction phase, participants received €11 as an initial payment but were obliged to risk the money as the stakes during the subsequent "game". Additionally, for the test phase, participants received €11 as an initial payment, also with the obligation of risking the money as the stakes during the selective listening task (see below). Furthermore, participants received €4 for each half-hour of the preparation preceding the experiment (i.e., electrode application, hearing level measurement).

The valence induction phase (A) consisted of one practice and three experimental blocks of the valence induction task. The test phase (B) comprised one practice and twelve experimental blocks of the selective listening task, with an additional block of the valence induction task interspersed after every two blocks of selective listening (i.e., the two phases together consisted of eight valence induction blocks and twelve selective listening blocks). We increased the duration of the test phase in the present experiment compared with Experiments 3-4 in order to provide a sufficient number of trials per condition for the averaged ERPs. Outcome focus was varied block-wise in the following way: Half of the selective listening blocks featured the motivational character of positive outcome focus, half of the selective listening blocks featured the motivational character of negative outcome focus. After two consecutive blocks of the same outcome focus, the opposite outcome focus was applied on the subsequent selective listening block. We counterbalanced between participants whether the test phase started with positive or negative outcome focus. Blocks were separated by short, participant-terminated breaks.

Valence induction phase

The valence induction phase of Experiment 5 (*Figure 6*A) was almost identical to that in Experiment 3, except that the highest tone (with tone-frequency of 1473.9 Hz) was associated with no-go meaning in all counterbalancing group in Experiment 5. Similarly to Experiments 3-4, participants had the possibility for winning or losing €1 on each valence induction block. However, as eight valence induction blocks were administered in the present experiment, there was a chance of winning or losing up to €8 in the valence induction task.

Test phase

Stimulus presentation characteristics and participants' task were identical to that applied in Experiment 3, apart from the introduction of positive and negative outcome focus in the present experiment (*Figure 6B*). In the positive and negative outcome blocks, the motivational focus of anticipating positive and negative future outcomes was introduced by the prospect of substantial monetary reward and substantial monetary loss, respectively. Depending on their performance, participants could collect "smileys" in the positive outcome blocks, and "saddies" in the negative outcome blocks. Every selective listening block started with an initial score of twelve "smileys" or "saddies". Participants were informed that they would immediately win €1 if the final score of "smileys" was more than twelve at the end of a positive outcome block (termed as "winning block"). While if the final score of "smileys" was twelve or below, there were no negative consequences. However, participants immediately lost €1 if the final score of "saddies" was more than twelve in a negative outcome block (termed as "losing block"). While if the final number of "saddies" was twelve or below, there were no positive consequences. Thus, there was a chance of winning or losing up to €6 during the selective listening task.

Similarly to the selective listening task of Experiment 3, participants were instructed to respond only to the target noise bursts (i.e., noise bursts including a brief "gap") by pressing the space bar as quickly and accurately as possible. Fast detection of a target was considered a success trial; a slow or incorrect response (i.e., a missed target or a false alarm) was considered a failure trial. The response-speed criterion for being successful on a given trial was defined by the moving median of the preceding five trials (see, e.g., Rothermund et al., 2008).[17] The outcome focus of the block de-

[17] The actual response-speed criterion (median') was dependent on a participant's current score of "smileys" or "saddies" in the following way: In positive outcome focus blocks: median' = median – (current total score of "smileys" - 12)*3, while in the negative outcome focus blocks: median' = median + (current total score of "saddies" - 12)*3, thus, leading to current scores that tend to be around the critical value of 12.

termined the consequences of success or failure: In a positive outcome block, fast detection of a target noise resulted in increasing the number of “smileys” by one, thereby increasing participants’ chance to win monetary reward at the end of the block. In case of a failure, the number of “smileys” was reduced by one. On the contrary, in a negative outcome block, fast detection of a target noise resulted in decreasing the number of “saddies” by one, thereby increasing the possibility to avoid monetary loss at the end of the block. In case of a failure, the number of “saddies” was increased by one, thereby increasing the possibility for monetary loss. Visual feedback was given after each target trial and in case of a false alarm, with the feedback indicating the explanation and the consequences of the recent (non-)response (“slow” / ”fast” / ”missed target” / ”false alarm” → +/- 1 “smiley” / ”saddie”). Additionally, the feedback included the current total in the following form: Possible positions of smileys or saddies were arranged into two 3 × 4 matrices that were separated by a blank line, thus, the critical value of twelve smileys or saddies was highlighted by the visual arrangement. After the feedback, participants could start the next trial by pressing the space bar. At the end of each selective listening block, a further feedback indicated the final total score of smileys or saddies and the actual outcome of the block (thus, plus or minus €1, or no consequences).

EEG Recording and Analysis

EEG recording, filtering, artifact rejection, baseline correction, choice of ROI, and amplitude calculations were identical to the methods that were applied in Experiment 3. As exception, in the present experiment, we segmented the continuous EEG into 800 ms long epochs, each including a 200 ms long baseline preceding tone onset. We decided to extract shorter epochs compared with the previous experiments, as – based on the results of Experiments 3-4 and according to our hypotheses relating to relatively early ERP components – no ERP components of interest were expected to occur later than in the time range of about 500-600 ms following tone onset. With decreased epoch duration we could increase trial numbers thereby increasing the signal-to-noise ratio (i.e., now the automatized rejection of trials containing severe artifacts or with temporal overlap between a task-irrelevant and task-relevant stimulus is based on the most relevant -200 ms to +600 ms time range relative to the task-irrelevant tone onset). In the positive outcome condition, on average ERPs were based on 85, 84, and 79 trials per participant in the positive, neutral, and negative valence conditions, respectively. In the negative outcome condition, on average ERPs were based on 86, 84, and 78 trials per participant in the positive, neutral, and negative valence conditions, respectively. Epochs were averaged separately for each condition. However, after introducing outcome focus manipulation, even with an increased duration of the test phase, we could base the calculation of averaged ERPs for each cell of our design on considerable fewer numbers of trials in the present experiment compared with Experiments 3-4.

Design and Statistical Analysis

We applied a 2 × 3 × 3 mixed design with outcome focus (positive, negative) and valence (positive, neutral, negative) as within-participants factors, and tone-frequency-to-valence assignment (three counterbalancing groups) as a between-participants factor (see Pollatsek & Well, 1995, and *Appendix A*) on the amplitudes for the ERP components of interest. Furthermore, we applied a 3 × 3 mixed design on the behavioral measures of the valence induction phase with valence (positive, neutral, negative) as a within-participants factor, and counterbalancing group as a between-participants factor. We used the multivariate approach to repeated measures analysis, with two orthogonal contrast variables (see, e.g., O'Brien & Kaiser, 1985; Petrova &

Wentura, 2012; M. Rohr et al., 2012). Similarly to Experiments 3-4, we applied a priori chosen contrasts: (1) for the first contrast, amplitude values were averaged across positive and negative stimuli and contrasted with the neutral stimuli (i.e., testing for the general relevance bias hypothesis). (2) The second orthogonal contrast was the contrast between positive and negative tones (i.e., testing for the negativity bias hypothesis).

RESULTS

A significance level of $\alpha = .05$ (two-tailed, unless otherwise noted) was adopted for all analyses.

Behavioral Performance

Behavioral performance was adequate on the valence induction and selective listening tasks, suggesting that participants were engaged in both tasks and followed the task instructions. RTs below 100 ms were discarded when calculating averaged RTs. In the valence induction task, RTs were calculated from the beginning of the loudness increase in the ongoing tone. The behavioral results of the valence induction phase are presented in *Table 12*.

Table 12. Behavioral results of the valence induction phase in Experiment 5: Mean reaction times (RTs), accuracy rates (ACCs), false alarm rates (FARs), and hit rates (HRs) in each valence condition; *SD* in parentheses.

Valence	**RT**	**ACC**	**FAR**	**HR**
Positive	333 (50)	88.5 (12.9)	16.7 (25.1)	93.7 (7.8)
Neutral	335 (46)	91.3 (9.8)	10.3 (19.8)	92.9 (8.4)
Negative	333 (41)	92.0 (6.9)	8.6 (10.4)	92.7 (7.6)

A 3 × 3 MANOVA for repeated measures with valence as within-participants factor and counterbalancing group as between-participants factor of the RTs did not show valence differences, $F < 1$, *n.s.* The average ACC across participants was adequately high ($M = 90.6\%$, $SD = 9.0\%$). Similarly to Experiments 3-4, the numerically lowest mean ACC emerged in the positive condition (see *Table 12*). ACCs did not show significant valence main effect, $F(2,20) = 2.30$, $p = .126$, $\eta_p^2 = .187$. However, as expected, ACC was significantly lower for positive compared with negative tones, $F(1,21) = 4.63$, $p = .044$, $\eta_p^2 = .181$, while valenced vs. neutral conditions did not differ significantly, $F < 1$, *n.s.* Again, comparably to Experiments 3-4, the relatively low mean ACC in the positive condition emerged due to more liberal response criterion as indexed by high FARs (as in this condition false alarms had no negative consequences), while the lowest mean FAR emerged in the negative condition (in which false alarms had the most extreme negative consequences; see *Table 12*). FARs did not show significant valence main effect, $F(2,20) = 2.30$, $p = .126$, $\eta_p^2 = .187$. However, as expected, participants made significantly more false alarms in the positive compared with negative condition, $F(1,21) = 4.77$, $p = .040$, $\eta_p^2 = .185$, while valenced vs. neutral conditions did not differ significantly, $F < 1.24$, *n.s.* Analogue analysis on the HRs did not show any differences, all Fs < 1.11, *n.s.*

Behavioral results on the selective listening task of the test phase are presented in *Table 13*. RTs did not differ statistically between positive and negative outcome focus conditions, $t(23) = -1.16$, $p = .257$, $d_z = -0.24$. The average ACC across participants was adequately high ($M = 96.1\%$, $SD = 3.1\%$). General ACCs and HRs did not differ between positive and negative outcome focus conditions, $t(23) = -1.03$, $p = .313$, $d_z = -0.21$, for ACCs; and $t(23) = -0.57$, $p = .572$, $d_z = -0.12$ for HRs. However, participants made more false alarms on blocks with the prospect of monetary reward compared with blocks with the prospect of monetary loss, $t(23) = 2.54$, $p = .018$, $d_z = 0.52$.

Table 13. Behavioral results of the test phase in Experiment 5: Mean reaction times (RTs), accuracy rates (ACCs), false alarm rates (FARs), and hit rates (HRs) in positive and negative outcome focus conditions; *SD* in parentheses.

Outcome Focus	**RT**	**ACC**	**FAR**	**HR**
Positive	500 (41)	95.7 (3.9)	3.5 (1.5)	95.0 (6.8)
Negative	504 (46)	96.5 (3.4)	2.8 (1.0)	95.8 (6.7)

ERP Results

Prototypical P1-N1-P2 waveform was observable in the group average ERPs elicited by the task-irrelevant positive, negative, and neutral tones during the test phase (see *Figure 7*, and for the N1 amplitudes in each condition, see *Table 14*). A 2 × 3 × 3 MANOVA for repeated measures with outcome focus (positive, negative) and valence (positive, neutral, negative) as within-participants factors and counterbalancing group as a between-participants factor of the N1 amplitudes (measured as mean voltage in a 20-ms window centered at the latency of the group-average N1 peak, at 108 ms) yielded no significant valence main effect, $F(2,20) = 1.56$, $p = .235$, $\eta_p^2 = .135$. The a priori contrast of valenced (positive and negative) vs. neutral condition was significant, $F(1,21) = 3.20$, $p = .044$ (one-tailed), $\eta_p^2 = .132$, indicating enhanced N1 amplitudes for valenced compared with neutral tones (M = -4.99 µV, SD = 3.10 µV in the valenced condition; M = -4.48 µV, SD = 2.64 µV in the neutral condition). The contrast of positive vs. negative conditions was not significant, $F<1$, *n.s.* (M = -4.99 µV, SD = 3.15 µV in the positive condition; M = -4.99 µV, SD = 3.28 µV in the negative condition). Outcome focus did not show significant main effect, $F(1,21) = 1.29$, $p = .269$, $\eta_p^2 = .058$. Of most interest, there was no indication of an interaction between valence and outcome focus, $F(2,20) = 0.11$, $p = .900$, $\eta_p^2 = .011$.

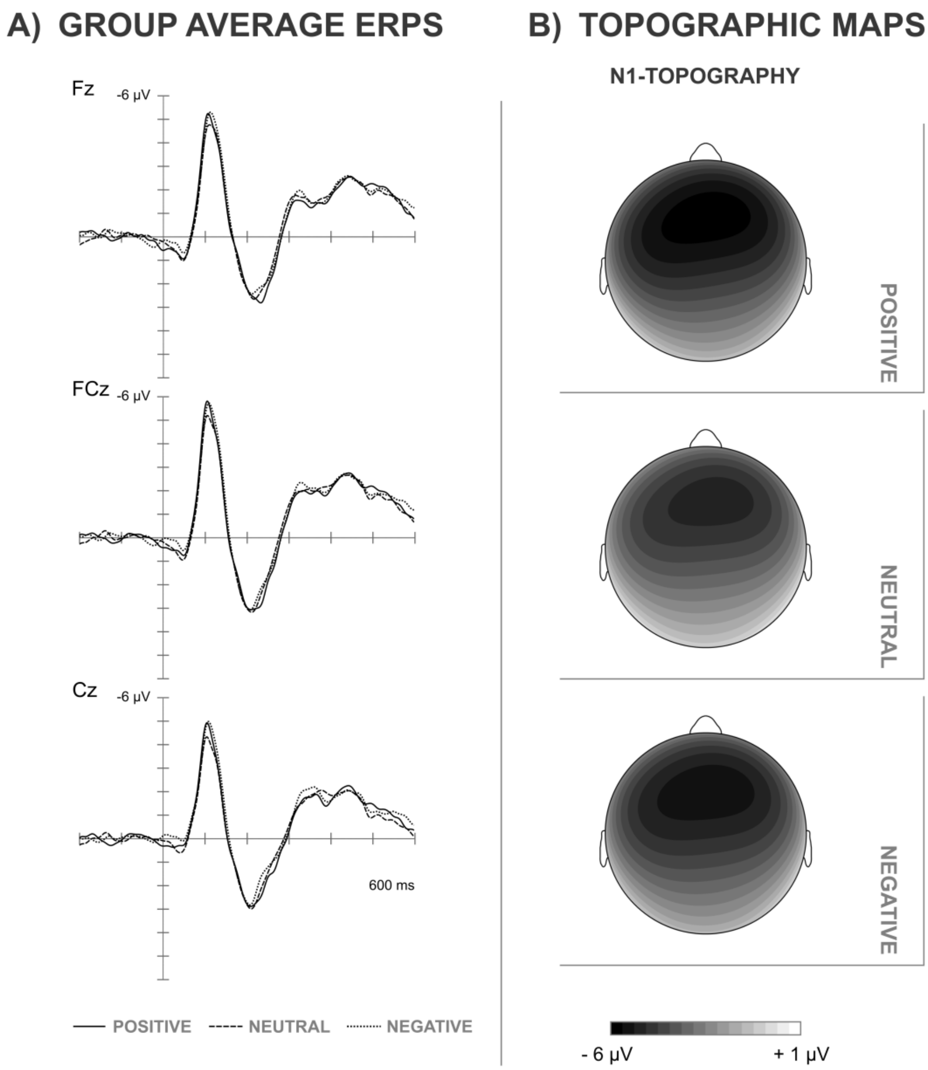


Figure 7. ERP results of Experiment 5.

(A) Group average ERP waveforms to the positive, neutral, and negative tones on the representative Fz, FCz and Cz electrode sites in Experiment 5. We present the ERP results collapsed across the outcome focus conditions, as our results did not show any indication for outcome focus modulation. The physical onset of the tones is at the crossing of the axes (0 ms). Negative polarity is plotted upwards. (B) Group average topography maps in the N1 time window (98-128 ms) in positive, neutral and negative conditions.

Table 14. ERP results of Experiment 5. Mean N1 amplitudes (in µV) in each valence and outcome focus condition; *SD* in parentheses.

	Outcome Focus	
Stimulus Valence	**Positive**	**Negative**
Positive	-4.80 (3.43)	-5.18 (3.05)
Neutral	-4.40 (2.33)	-4.56 (3.15)
Negative	-4.88 (3.54)	-5.10 (3.32)

In line with the standard analysis of Experiments 3-4, we performed a further analysis on the P1 amplitude measured as mean voltage in a 20-ms window centered at the latency of the group-average P1 peak (47 ms, experimental conditions collapsed). A 2 × 3 × 3 MANOVA for repeated measures with outcome focus (positive, negative) and valence (positive, neutral, negative) as within-participants factors and counterbalancing group as a between-participants factor of the P1 amplitudes did not yield significant valence main effect, $F(2,20) = 1.29$, $p = .297$, $\eta_p^2 = .114$. The a priori contrast of valenced (positive and negative) vs. neutral condition and the contrast of the two valenced conditions (positive vs. negative) did not show significant differences, *F*s < 1.78, *n.s.* (mean P1 amplitudes were $M = 0.60$ µV, $SD = 0.93$ µV in the positive condition; $M = 0.39$ µV, $SD = 0.94$ µV in the negative condition; $M = 0.75$ µV, $SD = 0.75$ µV in the neutral condition). Outcome focus did not show significant main effect, $F(1,21) = 2.37$, $p = .139$, $\eta_p^2 = .101$. There was no indication of an interaction between valence and outcome focus on the P1 amplitudes, $F(2,20) = 0.23$, $p = .793$, $\eta_p^2 = .023$.

Additionally, in line with Experiments 3-4, we conducted a further analysis in the later time range (350-500 ms post onset) at the frontocentral ROI. A 2 × 3 × 3

MANOVA for repeated measures with outcome focus (positive, negative) and valence (positive, neutral, negative) as within-participants factors and counterbalancing group as a between-participants factor did not show valence difference on the mean activity in the 350-500 ms time range, $F(2,20) = 0.05$, $p = .995$, $\eta_p^2 < .001$ (mean amplitudes were $M = -1.98$ µV, $SD = 1.80$ µV in the positive condition; $M = -1.99$ µV, $SD = 1.84$ µV in the negative condition; $M = -1.96$ µV, $SD = 1.81$ µV in the neutral condition). Outcome focus did not show significant main effect, $F(1,21) = 0.88$, $p = .770$, $\eta_p^2 = .004$. There was no indication of an interaction between valence and outcome focus, $F(2,20) = 0.61$, $p = .555$, $\eta_p^2 = .057$.

DISCUSSION

In Experiment 5, we introduced positive and negative motivational focus concerning future outcomes by assigning experimental blocks of the selective listening task with the chance of substantial monetary reward in the positive outcome condition, and with the danger of substantial monetary loss in the negative outcome condition. Behavioral results differentiated between positive and negative outcome focus conditions as indicated by the higher rate of false alarms when having the prospect of monetary reward compared with the prospect of monetary loss. This difference in false alarm rates can be interpreted as an indication of basic approach versus avoidance behavioral tendencies in line with the motivational focus (e.g., Duckworth, Bargh, Garcia, & Chaiken, 2002; Lang, 1995); or as an adjustment of task strategies on the selective listening task performance for achieving positive monetary outcomes. Nevertheless, importantly, the early attentional enhancement for valenced tones was clearly not modulated by outcome focus as reflected by a general enhancement of the auditory N1 for affectively significant tones relative to neutral ones regardless of the motivational context. Thus, the general pattern of results in Experiment 5 appeared to be similar to that in Experiment 3: The N1 was increased for valenced compared with neutral tones, and it did not show any differentiation between positive and negative valence.

However, the valence effect appeared to be weaker compared with the results that emerged from Experiment 3: The most relevant valenced-minus-neutral difference on the N1 amplitudes was associated with an effect size of $|d_z| = 0.36$ (M = -0.51 µV, SD = 1.43 µV), compared with the effect size of $|d_z| = 0.51$ for Experiment 3 (M = -0.49 µV, SD = 0.96 µV). Additionally, in line with the results of Experiment 4, we did not find a differential effect for valenced compared with neutral tones on the later sustained negativity. One possible explanation is that increasing the complexity of the design reduced test power for detecting valence differences. Furthermore, in line with the results of Experiment 4, we found generally enhanced sensory responses in the present experiment compared with Experiment 3 (similarly to Experiment 4 in which task-demands were increased compared to Experiment 3). However, we have to keep in mind that by introducing a differentiated positive versus negative motivational outcome focus, we also increased the motivational demands in general on the selective listening task in the present experiment compared with Experiment 3. Thus, the between experiments difference on the N1 might reflect an increase in non-selective arousal or alertness due to the higher motivational demands in the present experiment (see the *Intermediate Summary* below).

Behavioral results of Experiment 5 showed a similar pattern to that found in Experiments 3-4. Behavioral measures of the valence induction phase indicated a differentiation between positive and negative valence: In line with a more liberal response criterion for the positive compared with negative tones, relatively high rate of false alarms emerged in the positive condition, while the lowest rate of false alarms emerged in the negative condition.

In sum, Experiment 5 provided a replication of the early attentional enhancement to valenced compared with neutral tones, with no indication of a counter-regulation principle on the attentional biases to valent information at the early stage of sound encoding. Nevertheless, according to the present results, it is unclear whether there are such (more extreme) affective-motivational contexts and personal factors (e.g., already experienced substantial gains and losses; or affective disorders) that can promote differential attentional enhancement for specific valence categories during sound encoding. Future research can delineate whether affective-motivational factors can exert influence on early affective biases of auditory attention.

Comparable Early Attentional Effect across Experiments?

While we found indications in the single Experiments 4-5 for an early differential attentional enhancement for valenced tones, the attentional effect on the N1 amplitude was admittedly smaller in these experiments as compared with Experiment 3. The question targeted in the present chapter is whether this early valence effect is moderated by the different task settings of the three experiments, which presumably promoted different levels of demand for voluntary attentional selection of the task-relevant channel. In Experiment 4, a strict attentional selection of the task-relevant channel was promoted by continuous task-relevant stimulation, thus, we can conclude that differential attentional enhancement for valenced tones compared with neutral ones occurred involuntarily. Although we used the same perceptual detection task in Experiment 5 as in Experiment 3, by introducing a motivational outcome focus manipulation, we also increased the motivational demands on the selective listening task compared with Experiment 3. Accordingly, we can expect that participants had a generally higher motivation to achieve successful performance on the selective listening task compared with Experiment 3 that in turn could support the intentional attentional selection of the task-relevant channel. In line with this reasoning, participants detected the targets presented in the task-relevant channel even more accurately in Experiment 5 compared with Experiment 3 (93.2% in Experiment 3 vs. 96.1% in Experiment 5). Moreover, mean RT on the selective listening task Experiment 5 fell in between the mean RTs of Experiment 3 and Experiment 4 (590 ms in Experiment 3 vs. 367 ms in Experiment 4 vs. 502 ms in Experiment 5). These differences in the behavioral measures of the selective listening task can indicate that participants kept their attention on the task-relevant channel more constantly in Experiment 5 compared with Experiment 3, however, presumably less strictly compared with Experiment 4.

Hence, it is possible that participants' attentional focus was on the task-irrelevant channel on a proportion of trials in Experiment 3 but not (or on a smaller proportion of trials) in Experiments 4-5. Thus, the question emerges whether the apparent difference between the valence results of Experiment 3 and Experiments 4-5 mean that the prima facie involuntary early attentional enhancement for valenced tones depend (to some degree) on voluntary attention and concomitant task demands? We present a combined analysis of the three experiments concerning the most relevant early attentional effect on the N1 amplitude: That is, N1 amplitude measures obtained in Experiments 3-5 were included in a single analysis with valence (positive, neutral, negative) as within-participants factor and experiment as between participants factor. On the one hand, if an interaction emerges between the attentional enhancement for task-irrelevant valenced tones and the experiment factor (representing the three experiments with different task-relevant demands), we can conclude that the early attentional effect depends (at least to some degree) on the different task-characteristics of these experiments. On the other hand, the absence of such interaction would indicate rather a comparable process across the different task situations.

A 3 × 3 MANOVA for repeated measures with valence (positive, neutral, negative) as within-participants factor and experiment (Experiment 3, 4, 5) as a between-participants factor of the N1 amplitudes yielded a significant valence main effect, $F(2,68) = 5.01$, $p = .009$, $\eta_p^2 = .128$. The a priori planned contrast of valenced (positive and negative) vs. neutral condition was significant, $F(1,69) = 10.16$, $p = .002$, $\eta_p^2 = .128$, indicating enhanced N1 amplitudes in the valenced conditions compared with the neutral condition (M = -3.88 µV, SD = 2.45 µV in the valenced condition; M = -3.43 µV, SD = 2.39 µV in the neutral condition). The a priori contrast of the two valenced conditions (positive vs. negative) did not show any differences, $F(1,69) = 0.06$, $p = .801$, $\eta_p^2 = .001$ (M = -3.90 µV, SD = 2.54 µV in the positive condition; M = -3.86 µV, SD = 2.59 µV in the negative condition). There was a significant experiment main effect, $F(2,69) = 10.60$, $p < .001$, $\eta_p^2 = .235$, with lowest N1-amplitudes in Experiment 3 (M = -2.16 µV, SD = 1.41 µV in Experiment 3; M = -4.21 µV, SD = 1.71 µV in Experiment 4; and M = -4.82 µV, SD = 2.87 µV in Experiment 5). However, importantly, no interaction emerged between valence and experiment, $F(4,138) = 0.15$, $p = .965$, $\eta_p^2 = .004$, indicating comparable valence effect across the three experiments.

Thus, the results of this combined analysis suggest comparable early differential attentional enhancement for task-irrelevant valenced compared with neutral tones across the three experiments, although the single experiments employed different task-relevant attentional and motivational demands. This relative "immunity" to different degrees of concomitant task demands further supports our claim that the early attentional enhancement for valenced tones occurred relatively automatically.

Intermediate Summary: Rapid Preferential Attention to Valenced Tones

The main question addressed in this section was whether valenced (i.e., positive and/or negative) tones can receive enhanced attention relative to neutral ones at an early, perceptual stage of auditory processing. Furthermore, we targeted the question whether valenced tones govern preferential attention in general (i.e., in line with the hypothesis of a general relevance principle in affective attention) or rather a specific valence category is prioritized (specifically negative valence in line with the negativity bias hypothesis of affective attentional biases). Additionally, we investigated whether such basic affective attentional biases in the auditory modality are modulated by the affective-motivational context of anticipating positive or negative future outcomes.

In three ERP-experiments, we strove for a synthesis between the typical approaches of basic auditory attentional and affective research: On the one hand, in order to avoid perceptual confounds, we induced valence experimentally during a learning phase by assigning positive, neutral, and negative valences to tone-frequencies in a balanced design. On the other hand, we aimed to induce valence in an ecologically valid context by creating a task situation involving the danger of losing and the chance of winning actual resources. After the valence induction, EEG was recorded during a selective listening task in which tones with newly acquired affective valence were entirely task-irrelevant. In Experiments 3-4, we additionally tested whether participants indeed learned the associations between tone-frequencies and evaluative meaning using explicit and implicit measures of sound evaluation.

The main result of Experiment 3 was an amplitude enhancement of the auditory N1 for valenced (i.e., positive and negative valence together) compared with neutral tones, indicating enhanced attention to these tones at an early stage of sound encoding (i.e., approximately 100 ms after tone onset). In Experiment 4, we investigated the relative automaticity of this effect concerning whether it can occur independently from voluntary attention. To this end, we increased the demands for voluntary selection of a concurrent task-relevant stimulation. In Experiment 4, we found only a marginally significant valence effect in our standard analysis on the N1 amplitude. However, on the P1-N1 complex, we found indication for an early valence effect. We interpreted the pattern of results emerged from Experiment 4 as an indication for an involuntary attentional enhancement for valenced tones outside of the focus of voluntary attention. This interpretation was also supported by a combined analysis of our experiments. In Experiment 5, we investigated whether this early attentional enhancement is influenced by the motivational context of anticipating positive or negative future outcomes in line with the hypothesis of adaptively flexible affective attentional biases. Experiment 5 provided a replication of the early attentional enhancement on the N1 for valenced compared with neutral tones; however, this effect was not moderated by anticipating positive or negative future outcomes. Thus, the pattern of results concerning affective attentional biases at the level of sound encoding appeared unequivocal: In three experiments, we found only indications for valenced versus neutral differential effect at the early stage of sound encoding, but we found no indication for preferential attention to a specific valence category over another. Consequently, our results suggest that it is the general relevance of the valenced tones that governs early attentional processes rather than the priority of negative valence specifically.

Although the attentional effect for valence tones on the N1 amplitude appeared admittedly weaker in the single Experiment 4 and Experiment 5 compared with Experiment 3, a weakening of the valence effect was not confirmed statistically. A combined analysis of the three experiments showed a clear-cut pattern: We found a clear enhancement of N1 amplitudes for valenced tones across experiments, but, importantly, without any indication of an interaction with the factor of experiment. This pattern of results suggests comparable early attentional facilitation of the task-irrelevant valenced tones across the three experiments with different degrees of task-relevant attentional and motivational demands. Such a relative "immunity" to task-demands further supports our interpretation that the early preferential attentional enhancement for valenced tones occurs relatively automatically.

Although the most important finding of our studies was the differential attentional enhancement for valenced relative to neutral tones that was comparable between experiments, a more general between-experiments difference also appeared: Despite that voluntary attention was likely devoted more strictly to the main task in Experiments 4-5 compared with Experiment 3, the N1 response had generally higher amplitude in these two experiments. A possible explanation is that a change in non-specific arousal or alertness – possibly due to increased demands on the concurrent task – resulted in generally stronger sensory responses (see, e.g., Näätänen & Picton, 1987). Nonetheless, the differential enhancement for valenced relative to neutral tones appeared independently of this general enhancement.

To the best of our knowledge, the present experiments are the first to present evidence for such an early attentional enhancement for valenced stimuli in auditory ERPs, demonstrating a fast interplay between attention and affective factors during sound encoding. ERP studies of valenced stimuli show great variability, especially in an early latency range (within about 100-150 ms after stimulus onset; see, e.g., Olofsson et al., 2008, for a review of visual studies). One possible source of this variability is that early ERPs are strongly influenced by the physical characteristics of the presented stimuli (e.g., Burkard et al., 2007), which makes experimental designs comparing physically complex stimuli susceptible to stimulus confounds. A strength of the present experiments is the strict control over arbitrary physical differences. This was achieved by assigning a priori neutral tones with positive, negative, and neutral meaning through a learning phase. It is important to note that our main finding – an enhancement of the N1 amplitude for valenced compared with neutral stimuli – cannot be explained by the variable physical feature of tone-frequency, as it was counterbalanced between participants.

Although naturally valenced sounds are evaluated rapidly as demonstrated by Experiments 1-2, the evaluative connotation of spectrotemporally rich, natural emotional sounds is typically not available instantly at sound onset due to temporal unfolding of this information. On the contrary, the approach of Experiments 3-5 has the advantage of an immediate signal value: As we employed constant tone-frequency as the critical stimulus feature to convey valence information, evaluative information was available already at about sound onset. Thus, this approach and the approach to explore naturally valenced sounds (in line with Experiments 1-2) complement each other.

Conclusions

In sum, Experiments 3-5 provided evidence for differential attentional enhancement for valenced compared with neutral tones at a perceptual stage of sound encoding. Thus, besides the great number of previous studies investigating evaluative picture processing, the present study found evidence for rapid preferential processing of valenced tones. Our results suggest that the early attentional enhancement for valenced auditory stimuli emerges involuntarily. While explicit and implicit behavioral measures differentiated between positive and negative valence, our ERP results suggest that the general relevance of the valenced tones governs attentional processes at the early level of sound encoding.

The Next Steps: Preferential Selection of Valenced Tones from Temporal Sequences?

Briefly, based on the results of the hitherto conducted experiments, we can conclude that auditory stimuli can be evaluated rapidly and unintentionally, and tones endowed with affective valence can attract preferential attention already at the level of sound encoding in an automatic way (i.e., rapidly, unintentionally, and relatively independently of voluntary attention). In the following section, given the special importance of the temporal dimension in the auditory modality, we will explore whether auditory temporal attention is biased in the favor of affectively significant tones compared with neutral ones in sound sequences. More specifically, we will investigate whether tones with positive and/or negative valence are selected preferentially compared with neutral ones during target detection in a rapid auditory serial presentation paradigm.

Preferential Selection of Valenced Tones from Temporal Sequences[18]

Time is an essential organizing principle in the auditory modality: A key characteristic of our auditory world is its transient and also temporally extended nature. First, natural sounds are dynamic and carry temporally distributed information. Second, in natural environments sounds rarely occur in isolation, but rather as several simultaneous and successive acoustic events. Given the basic importance of the temporal aspect in our auditory perception, efficient selection of affectively significant cues in temporally distributed auditory patterns can be comparably crucial as extracting affective information from visual space (see, e.g., Öhman et al., 2001, for facilitated detection of emotional pictures in spatial arrays). In Experiments 3-5, we found an

[18] This chapter is presented in a draft of Folyi and Wentura (manuscript in preparation).

early attentional enhancement on the auditory N1 for valenced relative to neutral tones, that is, on an ERP component that is functionally related to transient detection mechanisms, such as detection of sound onsets and rapid changes. As a functional link is suggested by visual studies between early attentional enhancement for affectively significant visual stimuli and a benefit of detection performance in visual space (e.g., Pourtois et al., 2013; Vuilleumier & Driver, 2007), the question arises as to whether rapid preferential attention to affectively significant tones – as demonstrated by Experiments 3-5 – can provide a detection benefit for these tones in time. Hence, in the present chapter, we will investigate whether tones with positive and/or negative valence can be selected preferentially from rapid sound sequences compared with neutral tones in a relatively automatic way, leading to a detection benefit for these tones under the conditions in which auditory attention is limited in time.

In the visual modality, a performance deficit termed as attentional blink is a well-established result during multiple target detection in rapid stimulus sequences: Detection performance typically drops for a second target after a correctly detected first target as a function of their temporal distance (for reviews, see Dux & Marois, 2009; Martens & Wyble, 2010; Shapiro et al., 1997). Importantly, affectively significant visual targets are often preserved from the attentional blink, indicating their special attentional status, in the sense that they are more resistant to the temporal constraints of visual attention compared with neutral targets. In the present section, we aim to explore possible affective modulation of the attentional blink deficit first time in the auditory modality. Due to the fact that the auditory attentional blink is much less established compared with its visual counterpart, first, in Experiment 6, we will investigate the basic phenomenon by introducing a dual-target detection task in a rapid auditory serial presentation (RASP) paradigm. Critically, in Experiment 7, we will introduce an affective variant of this paradigm by presenting simple tones endowed with affective valence as second targets. We expect that tones with positive and/or negative valence receive early attentional facilitation, leading to a relative resistance to a dual-target detection deficit. In the following, we will give an overview of the auditory perceptual and attentional factors that can possibly influence detection of significant cues in rapid sound sequences. On this note, we will delineate the main characteristics of the (visual) attentional blink, and highlight possible differences between the visual attentional blink and its auditory counterpart. Given this background, we will describe how the visual attentional blink is moderated by the affective significance of the targets and our hypothesis about the affective modulation of the expected dual-target deficit in the auditory modality.

Auditory Temporal Processing

Without doubt, extensive temporal analysis of the dynamic acoustic environment is essential for most of our auditory abilities (e.g., Shinn, 2003). In a broad sense, the term of auditory temporal processing refers to the ability of the auditory system to represent the complex temporal structure of the acoustic input (see, e.g., Eggermont, 2015). This umbrella term includes such important abilities of the auditory system as temporal ordering (i.e., the ability of defining the order of rapidly presented auditory stimuli, e.g., Fitzgibbons & Gordon-Salant, 1998), temporal discrimination and resolution (i.e., the ability of discriminating durations of briefly presented auditory stimuli or detecting silent intervals in sounds, e.g., Eddins & Green, 1995; Wright, Buonomano, Mahncke, & Merzenich, 1997), temporal integration (i.e., the inverse relationship between sound duration and detection threshold, e.g., Eddins & Green, 1995; Viemeister, 1996); and release from auditory temporal masking (referring to the threshold shift for detecting a sound when another sound is presented non-simultaneously but in close temporal vicinity, e.g., Massaro, 1975; see also Picton, 2013; Shinn, 2003, for reviews). The term encompasses perceptual abilities that have been demonstrated to react sensitively to higher-order influences, such as attentional manipulations and training on higher-level cognitive functions (e.g., Leek, Brown, & Dorman, 1991; D. R. Moore, Halliday, & Amitay, 2009; Strait, Kraus, Parbery-Clark, & Ashley, 2010). Without question, the refined auditory temporal processing abilities are of vital importance when perceiving rapid sound sequences. Hitherto, auditory temporal processing has been in the focus of many investigations mainly due to its relevance for speech perception and language development (e.g., Alvarez, Fuente, Coloma, & Quezada, 2015; Grube, Cooper, Kumar, Kelly, & Griffiths, 2014; Grube, Kumar, Cooper, Turton, & Griffiths, 2012; D. R. Moore et al., 2009; Tallal & Gaab, 2006; Wright, Lombardino, et al., 1997). However, given its basic importance in our "auditory world", the question occurs whether a bias can arise for affectively significant cues in the temporal aspect of auditory processing.

Temporal Deficit of Auditory Attention: Analogue to the Visual Attentional Blink?

A transient performance decrement is often observed for a second target following a successfully detected first target during multiple target detection in rapid sound sequences. This intriguing phenomenon is mostly interpreted as an analogue to the visual attentional blink (AB) leading to the notion of the auditory AB (for auditory and auditory-visual AB studies, see, e.g., Arnell & Jolicœur, 1999; Duncan et al., 1997; Hein, Parr, & Duncan, 2006; Horváth & Burgyán, 2011; Koelewijn, Burg, Bronkhorst, & Theeuwes, 2008; Mondor, 1998; Shen & Alain, 2010; Shen & Mondor, 2006, 2008; Tremblay et al., 2005; Vachon & Tremblay, 2005; Vachon et al., 2010).

In brief, the concept of AB refers to a transient impairment of conscious identification or report of a second target after a successfully detected first target in a rapid sequence of targets and distractors (i.e., task-irrelevant items; for reviews, see Dux & Marois, 2009; Martens & Wyble, 2010; Shapiro et al., 1997). Despite the above-mentioned handful of auditory demonstrations, the notion of AB has been established mainly based on visual research. Accordingly, theoretical explanations of the AB also principally considered vision with some authors even claiming that AB is a strictly visual phenomenon (e.g., Chun & Potter, 2001), while others restricting it to within sensory modalities (Duncan et al., 1997; Hein et al., 2006; Martens, Johnson, Bolle, & Borst, 2008; but see also e.g., Arnell & Jenkins, 2004; and Arnell & Jolicœur, 1999, for demonstrations of cross-modal AB suggesting a central, modality-independent limitation). Initial accounts of the (visual) AB proposed that the transient detection impairment for the second target occurs due to suppression of its perceptual processing following the first target (gating theory, Raymond, Shapiro, & Arnell, 1992), while later interpretations have emphasized that the locus of AB is at post-perceptual level (see, e.g., Dux & Marois, 2009; Martens & Wyble, 2010). This latter assumption has been supported by empirical evidence indicating that even unreported targets receive full perceptual analysis (e.g., Jacoby, Visser, Hart, Cunnington, & Mattingley, 2011; Luck, Vogel, & Shapiro, 1996; Sergent, Baillet, & Dehaene, 2005), while a suppression is typically observed on ERP-correlates of post-perceptual, working memory-related processes (e.g., Dell'Acqua et al., 2015; Rolke, Heil, Streb, & Hennighausen, 2001; Vogel & Luck, 2002; Vogel, Luck, & Shapiro, 1998).

From the numerous accounts of the (visual) AB, we highlight two major lines of interpretations. The earlier resource-limitation accounts understand the AB as a first target-induced depletion of capacity-limited cognitive resources, leading to impaired consolidation of the second target in visual working memory (e.g., Chun & Potter, 1995; Jolicœur & Dell'Acqua, 1999), or a reduced weighting for the second target relative to the first target and distractor representations during retrieval (interference theory, Isaak, Shapiro, & Martin, 1999; Shapiro, Raymond, & Arnell, 1994; Shapiro et al., 1997). An influential interpretation of this kind proposes two stages of processing: The first stage has a high capacity for parallel processing of several stimuli, and it is functionally related to initial stimulus detection and identification. Representations at the first stage are transient and vulnerable, and thus they are susceptible to interference from the following items, but they also form a basis for selection for later processing. The second stage, however, has limited capacity, and its "costly" processes enable conscious report of a certain stimulus. Critically, if the second target cannot benefit from the capacity-limited second stage because the "costly" processing is devoted to a first target, its representation will decay quickly (two-stage model of AB, Chun & Potter, 1995).

In the visual modality, a special temporal pattern of target detection performance is often observed following a correctly detected first target: Paradoxically, performance for the second target is relatively preserved when it follows the first target directly (i.e., at the first temporal lag relative to the first target), while the AB is typically the "deepest" in the interval of about 200–500 ms following the first target (or, according to an item-based interpretation, when the second target is presented at about the second or third temporal lag relative to the first target). The preserved performance for the second target when it is presented directly after the first target is referred as the lag 1 sparing effect, and this outcome is often considered as a hallmark of a "pure" attentional blink without additional costs like task or spatial switching (see, e.g., Chun & Potter, 2001; Visser, Bischof, & Di Lollo, 1999). This intriguing phenomenon has posed a challenge to many theories of the AB. The resource-limitation accounts accommodate the lag 1 sparing effect by assuming that the two targets can be processed together in a single episode when they are presented consecutively (e.g., Chun & Potter, 1995; Jolicœur & Dell'Acqua, 1998). Explanations for this prima facie contradictory outcome are inherent in a second line of theories in which the focus has shifted from a first target-induced resource depletion to attentional control mechanisms, emphasizing the role of disruptions in or over-application of attentional control (e.g., temporary loss of attentional control account, Lollo et al., 2005; threaded cognition model of AB, Taatgen et al., 2009; or the boost and bounce theory of AB, Olivers & Meeter, 2008). These accounts challenged the capacity limitation view, especially by demonstrating that several targets can be detected without severe performance deficit given that they are presented in direct succession (e.g., Olivers, Stigchel, & Hulleman, 2007). Hence, the role of the presence or absence of intervening distractors between the targets has been emphasized instead of the post-first target time course per se (e.g., Lollo et al., 2005; Olivers & Meeter, 2008). According to the boost and bounce theory of visual temporal attention (Olivers & Meeter, 2008), when encountering relevant information, a dynamic attentional system boost the visual input via a transient attentional enhancement; while when encountering irrelevant information, a transient attentional suppression blocks the visual input, in order to provide or gate access to working memory, respectively. The transient attentional response does not take effect immediately, but peaks about 100 ms after having encountered relevant information. During multiple target detection in rapid sequences, the theory proposes that a first target-induced attentional enhancement is the common cause of both the lag 1 sparing and the transient performance impairment for the second target at later lags. Thus, a preserved performance at the first lag occurs as the second target receives from the attentional boost induced by the first target. If a distractor item follows the first target in close temporal vicinity, it also receives an initial attentional enhancement. However, in this case, the initial enhancement will be followed by a strong transient suppression ("bounce") in order to prevent the distractor from entering working memory. This counteracting response results in inhibition of the items following the distractor. Accordingly, performance impairment for the second target occurs due to this strong suppression triggered by the distractor(s) presented after the first target.

Consequently, between-targets distractors have a key role in producing AB via closing the "attentional gate", while the theory does not predict performance deficit for a situation in which several targets are presented consecutively without intervening distractors (thus, the "attentional gate" is open; Olivers & Meeter, 2008).

In sum, the lag 1 sparing can be an important distinguishing characteristic of a "genuine" (visual) AB, especially as further factors that can cause performance decrement for the second target – like perceptual interference between the targets and spatial, stimulus category, or task switching effects – are maximal at lag 1 and decay rapidly thereafter (see, e.g., Chun & Potter, 2001; Visser et al., 1999). Hence, a "genuine" (visual) AB pattern can be described with a U-shaped temporal pattern of performance deficit for the second target with the most pronounced deficit in the AB-time window (about 200-500 ms after the first target, or at about lag 2-3 position relative to the first target), thus, in the situation in which distractor(s) intervene(s) between the targets (and, in line with the interpretation of Olivers & Meeter, 2008, closes the "attentional gate" for the successive items). Moreover, this pattern is characterized by a relatively preserved performance at the first lag (open "attentional gate" at successive target presentation), and a relatively recovered performance at later lags (after about 500 ms, following the transient inhibition, Olivers & Meeter, 2008).

A further remarkable outcome of the AB paradigm is that participants report the presentation order of the two targets reversed on a proportion of trials (obviously mainly when the two targets follow each other directly, thus, at lag 1). These so-called inversion errors are understood by the resource-limitation accounts of the AB similarly to the lag 1 sparing effect, namely that the two targets can be processed together when they are presented consecutively (e.g., Chun & Potter, 1995), merging into a single episodic representation, although that comes at a price of lost temporal order information (e.g., Akyürek et al., 2012). According to Olivers, Hilkenmeier, and Scharlau (2011), the law of prior entry, meaning that attended stimuli can have a benefit in processing speed (for the original assumption, see Titchener, 1908; and for a review, see Spence & Parise, 2010), can provide an alternative explanation for the inversion errors in the AB paradigm. The boost and bounce theory of AB (Olivers & Meeter, 2008) predicts that when the two targets are presented in direct succession, the second target also receives from the attentional boost induced by the first target (i.e., in line with the explanation of the lag 1 sparing outcome). On a proportion of trials, the second target can even "win the race" over the first target, supposedly by a relative benefit in processing speed as a consequence of the transient attentional boost, thereby leading to a benefit in the time of arrival to higher stages of information processing and its prior entry into working memory (Hilkenmeier, Olivers, & Scharlau, 2012; Olivers, et al., 2011). Importantly, empirical evidence indicates that focused attention can speed up perceptual processing also in the auditory modality (Folyi et al., 2012), opening the way to an interpretation of inversion errors based on prior entry in the auditory modality in line with the visual theory of Olivers et al. (2011).

However, there is hardly any proof of preserved performance at the first lag during dual-target detection in the auditory modality. Quite the contrary, the auditory target detection deficit termed as AB is typically most pronounced at lag 1 (e.g., Arnell & Jenkins, 2004; Arnell & Jolicœur, 1999; Duncan et al., 1997; Horváth & Burgyán, 2011; Mondor, 1998; Shen & Alain, 2010; Shen & Mondor, 2006; Vachon et al., 2010; Vachon & Tremblay, 2005; for an exception, see Tremblay et al., 2005, experiment 2). This temporal pattern of performance decrement in the auditory modality can be interpreted as a contribution of short-range and relatively low-level effects to the "genuine" AB that is assumed to reflect limitations or interruptions in task-relevant cognitive processes. These additional effects at short temporal lags can reflect a failure of auditory temporal processing like temporal masking by the preceding salient target, a cost of having to reallocate attention from the frequency region of the first target to the frequency region of the second target (e.g., Tremblay et al., 2005), or auditory distraction (referring to an "exogenous" attentional capture by the first target driven by its perceptual salience rather than its task-relevance, Horváth & Burgyán, 2011). As the auditory AB is much less established compared with its visual counterpart, before introducing affective manipulation (Experiment 7), in Experiment 6 we investigate the temporal dynamics of auditory attention during dual-target detection in rapid sound sequences. In line with the auditory AB literature (and contrary to the typical visual AB-pattern), we expect the most pronounced performance decrement for the second target when it follows a successfully detected first target directly (i.e., at lag 1).

Affective Information Breaks through the Auditory Attentional Blink?

Of particular interest to the present work, in the visual modality, AB has been demonstrated to vary by the affective significance of the presented stimuli, indicating a detection benefit for emotional targets compared with neutral ones. One line of research demonstrated that affectively significant first targets – or even affectively significant task-irrelevant items presented before a target – produce a more severe AB deficit compared with neutral stimuli (Arnell, Killman, & Fijavz, 2007; Engen, Smallwood, & Singer, 2015; Huang, Baddeley, & Young, 2008; Ihssen & Keil, 2009; Most, Chun, Widders, & Zald, 2005; S. D. Smith et al., 2006; Stein, Zwickel, Ritter, Kitzmantel, & Schneider, 2009; for a review on emotion-induced attentional blink, see McHugo, Olatunji, & Zald, 2013), an outcome explained by preferential attentional processing (supposedly a delayed disengagement of attention, Ihssen & Keil, 2009; McHugo et al., 2013; Most, et al., 2005) triggered by the emotional stimulus. A further affective variant of the visual AB paradigm presumably taps into more “reflexive” affective attentional processes (see, e.g., Schwabe et al., 2011): Affectively significant second targets have found to be more resistant to the AB deficit compared with neutral second targets (A. K. Anderson, 2005; A. K. Anderson & Phelps, 2001; Keil & Ihssen, 2004; Keil, Ihssen, & Heim, 2006; Oca et al., 2012; Raymond & O’Brien, 2009; Schwabe et al., 2011; D. Zhang et al., 2014). This outcome can be understood as the affectively significant targets “break through” the AB more readily compared with neutral targets as receiving early preferential processing that in turn can lead to a facilitation at later stages, such as working memory processes (e.g., Keil & Ihssen, 2004; Keil, Ihssen, & Heim, 2006; Schwabe et al., 2011). The “AB-sparing” indicates a special attentional status of affectively significant information: The preferential attentional selection of affectively significant stimuli can be considered automatic in the sense that it occurs relatively independently from the temporal constraints of voluntary attention – regardless of whether these constraints emerge due to a first target-induced cognitive resource depletion or due to attentional control mechanisms.

Given this background, the main goal of the present section can be summarized as follows: We target to explore possible affective modulation of the dual-target detection deficit first time in the auditory modality. To achieve this goal, first we demonstrate a dual-target deficit in a rapid serial auditory presentation (RSAP) paradigm in order to reveal the basic characteristics of this effect with a special focus on its temporal pattern (Experiment 6). Second, in our main experiment (Experiment 7), we aim to explore possible affective modulation of this deficit by the affective valence of the second target. Similarly to the approach of Experiments 3-5, we associate simple tones with affective valence during a learning phase. By employing this method, we not only gain strict control over arbitrary physical differences, but we also have the advantage of immediate signal value, thus, evaluative information carried by different tone-frequencies will be available already at about sound onset.

Experiments 3-5 demonstrated that simple tones associated with affective valence can receive “natural” preferential attention relative to neutral tones already at an early, perceptual level of sound encoding. In line with these results and with the findings of affective “AB-sparing” in the visual modality, we expected that valenced tones can attract early attentional enhancement in a temporal sequence and thereby, they will be relatively spared from dual-target deficits compared with neutral targets. Given the special importance of rapid and effective detection of significant cues in the dynamic acoustic environment, we expect that a similar or even more pronounced preferential selection emerges for affective auditory stimuli in rapid temporal sequences as in the visual modality. In line with previous auditory AB-findings showing the most severe deficit when the targets are presented in direct succession, we expect that the affective “AB-sparing” will be manifested most markedly at lag 1 in the auditory modality. Furthermore, in line with the prior entry hypothesis applied to inversion errors (Olivers et al., 2011), it is possible that affectively significant tones also profit from an attentional speed-up of auditory processing in a rapid sequential presentation paradigm, leading to a benefit in the time of arrival to higher stages of information processing. If so, we can expect that the attentional benefit for valenced second targets is reflected in the higher proportion of inversion errors compared with neutral second targets.

Experiment 6: Dual-Target Detection in a RASP Paradigm

In Experiment 6, we introduced an RSAP paradigm presenting relatively simple auditory stimuli (for using simple sinusoidal tones and complex tones with a few frequency components in AB studies, see e.g., Horváth & Burgyán, 2011; Mondor, 1998; Shen & Mondor, 2006; Vachon & Tremblay, 2005), with a complex tone interrupted by a silent interval as first target (*Target 1*, *T1*), and with simple sinusoidal tone as second target (*Target 2*, *T2*). Targets were embedded into a rapid sequence of task-irrelevant distractors tones. In each sequence, either a single target (T1 or T2) or both targets (*dual-target* or *T1 + T2 trials*) or only distractors were presented. The sequential position of T2 was varied in a way that resulted in a delay of lag 1, lag 3, or lag 6 relative to T1. We employed a simple target detection task (see, e.g., Horváth & Burgyán, 2011; Shen & Mondor, 2006; Vachon & Tremblay, 2005). In our paradigm, participants had to report after every trial whether they have heard only distractors, a single T1 (termed as "dashed" target), a single T2 (termed as "clear" target), or both targets. In the latter case, participants had to report the targets in their correct presentation order.

For the purpose of our study, the most relevant measure is the HR for T2 as a function of whether it was preceded by a T1 or not, and as a function of T1-T2 lag, conditioned on successful T1-performance (i.e., in line with the AB-literature). We expected a lag-dependent performance deficit for T2 on dual-target trials when T1 is detected correctly; with specific hypothesis concerning the temporal pattern of this deficit: Namely, we expected the most marked detection deficit to occur when T1 is presented directly before T2 (i.e., at lag 1) compared with later lags. This pattern is indicated by the majority of auditory AB studies, and we can assume that at least partly it reflects lower-level effects that are maximal at lag 1 (e.g., failure of temporal processing abilities, switching costs, or auditory distraction driven by perceptual saliency). Additionally, we tested whether there is performance deficit for T1 on the trials on which T2 was detected correctly, depending on the temporal distance between the targets. A possible effect of T1-T2 temporal lag on the T1-HRs cannot be considered as a conventional index of AB, but it conveys important information about a possible short-ranged competition or interference between the targets. Such deficit would indicate a trade-off between the two targets presumably by lower-level factors (e.g., failure of temporal processing abilities) than a “genuine” AB. A further analysis targeted the influence of temporal distance between the targets on the inversion errors. For the two latter cases (as self-evident characteristic of these two phenomena), we expected the performance deficit to be maximal when the two targets are presented directly after each other (i.e., at lag 1).

METHODS

Participants

Thirty-five students from Saarland University (29 females; aged 20–33 years, *Mdn* = 25 years) participated in the experiment for monetary compensation. All participants were native German speakers, had normal or corrected-to-normal vision, and self-reported normal hearing. Before the experiment, all participants gave written informed consent. Given a sample size of $N = 35$ and an α-value of .05 (two-tailed), effects size of $d_z = 0.49$ can be detected with a probability of $1 - \beta = .80$ for the most relevant lag 1 vs. later lags comparison representing our main hypothesis (calculated with the aid of G*Power 3 software; Faul et al., 2007).

Materials and Procedure

Sounds were presented binaurally via headphones (K 511, AKG Acoustics, Harman International Industries, Inc.) at a comfortable maximal loudness of approximately 70 dB(A)[19]. All auditory stimuli faded in and out with 5 ms rise and fall times. *Figure 8* gives an overview of the frequency structure of the presented auditory stimuli.

[19] Distractor sounds were presented with maximal intensity (reference amplitude of 1). Amplitudes of the three versions of Target 2 were attenuated relative to the distractors in the following way: (1) low Target 2: 0.8*reference amplitude 1; (2) medium Target 2: 0.5*reference amplitude 1; (3) high Target 2: 0.4*reference amplitude 1. This attenuation aimed to reduce saliency of Target 2 tones relative to the distractors. Attenuations were made according to the data of three pilot participants.

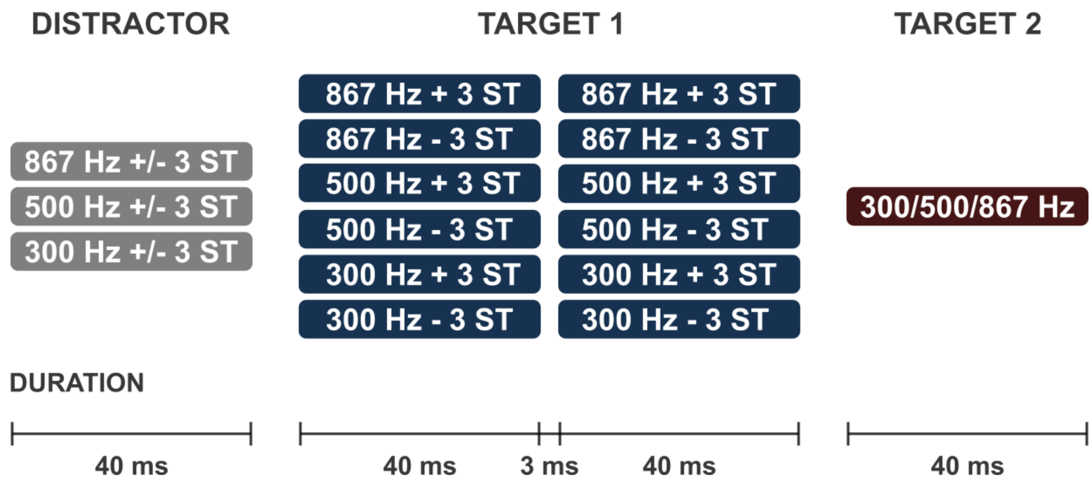


Figure 8. Schematic illustration of the auditory stimuli presented in the RASP paradigm of Experiments 6-7.

Distractors were complex sounds with three sinusoidal components. Distractors were presented with a duration of 40 ms. There were eight types of distractors defined by combinations of different tone-frequencies. *Target 1* was a complex sound with six sinusoidal components. Target 1 was presented with an overall duration of 83 ms that was interrupted by a 3 ms long silent period starting at 40 ms post-onset. *Target 2* was a simple sinusoidal tone. Target 2 was presented with a duration of 40 ms in three different versions: with tone-frequency of 300 Hz or 510 Hz or 867 Hz.

Three types of auditory stimuli were employed: distractor, T1, and T2. As *distractors*, complex sounds were presented with three equally weighted sinusoidal components. Eight types of distractors were created from the combinations of the following frequency components: F_0 = 252.3 Hz or 356.8 Hz; F_1 = 428.9 Hz or 606.5 Hz; F_2 = 729.1 Hz or 1031 Hz. As *T1*, a complex sound was presented with six equally weighted sinusoidal components (frequency components: F_0 = 252.3; F_1 = 356.8; F_2 = 428.9; F_3 = 606.5; F_4 = 729.1; F_5 = 1031 Hz). T1 had an overall duration of 83 ms that was interrupted by a 3 ms long silent period starting at 40 ms post-onset. As *T2*, a simple sinusoidal tone was presented in three different versions: with tone-frequency of 300 Hz or 510 Hz or 867 Hz (in the following referred as low, medium, and high pitch versions of T2, respectively). These three tones are roughly equidistant in pitch, but there are no harmonic or musical relations between them. Duration of the auditory stimuli was 40 ms except T1.

Note, that the tone-frequency components of T1 and the distractors are the tone-frequencies of the three versions of T2 shifted by 3 semitones (ST): 300 Hz +/–3 ST, 510 Hz +/–3 ST, 867 Hz +/–3 ST. We used the same tone-frequencies for T2 as the tone-frequencies of the low, middle low, and middle high tones in Experiments 3-5. In Experiment 6, these different versions of T2 are not of interest, however, in Experiment 7, similarly to the approach of Experiments 3-5, these three tone-frequencies will be associated with positive, neutral, and negative affective meaning.

Four types of experimental trials were presented: (1) *no-target* trials featured only distractors, (2) *only-T1* trials featured a T1 and distractors, (3) *only-T2* trials featured a T2 and distractors, while (4) *T1 + T2 trials* featured both targets and distractors. During the experiment, all the four types of trials were presented in a randomly intermixed order.

Figure 9 gives an overview of the temporal structure of the stimulus sequences. An experimental trial comprised a sequence of 12–37 auditory stimuli with an ISI of 40 ms. A sequence started with the presentation of 5–20 distractors, followed by a middle part that defined the type of the trial, and ended with the presentation of 5–10 distractors. The middle part was defined by three factors: (1) whether it contained a T1 or not; (2) whether it contained a T2 or not; and (3) whether it consisted of two, four or seven stimuli. Combinations of the first two factors defined the four trial types (no-target vs. only-T1 vs. only-T2 vs. T1 + T2 trials). The third factor defined the delay of the last position of the middle part relative to the first position of the middle part (termed as *lag 1*, *lag 3,* and *lag 6* conditions). If a trial contained a T1, it occupied the first position of the middle part. If a trial contained a T2, it occupied the last (thus the second, fourth, or seventh) position of the middle part (i.e., lag 1, lag 3, or lag 6 relative to the first position, respectively). If it was not occupied by a target, first and last positions and all other possible positions of the middle part were filled by distractors. Each of the 4 (trial type: no-target vs. only-T1 vs. only-T2 vs. T1 + T2) × 3 (lag: 1 vs. 3 vs. 6) combinations of trials was presented nine times in a randomized order during an experimental block. One-third of the trials containing T2 was presented with low, one-third with medium, and one-third with high pitch versions of T2 (counterbalanced between trial type and lag conditions). Distractors were presented in an individually randomized order.

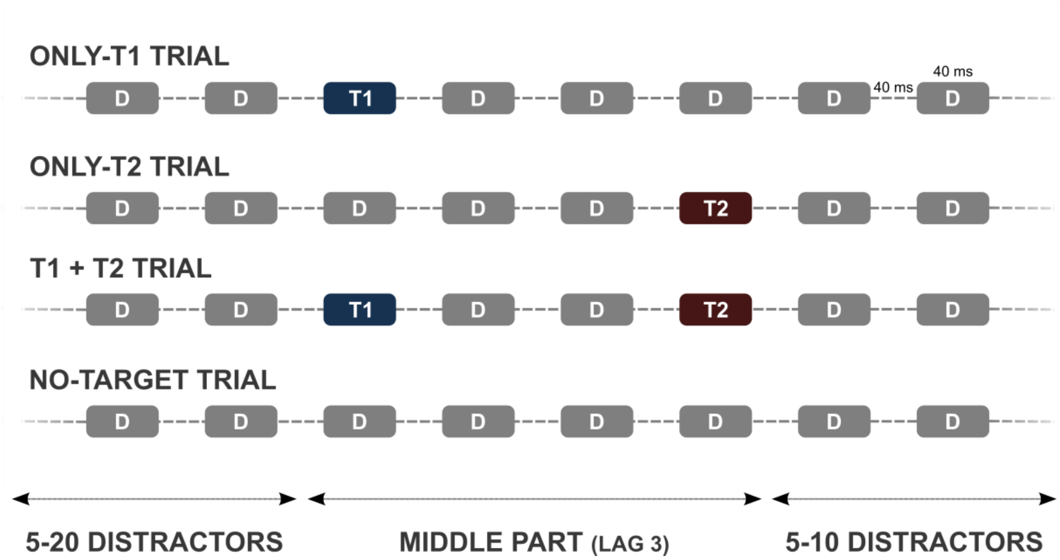


Figure 9. Schematic illustration of the stimulus sequences presented in the RASP paradigm of Experiments 6-7.

Distractors (D) are represented in gray, T1 is represented in blue, and T2 is represented in brown. Each sequence started with the presentation of 5–20 distractors, followed by a *middle part* that defined the type of the trial, and ended with the presentation of 5–10 distractors. The middle part of the sequence was defined by three factors: (1) whether it contained a T1 or not; (2) whether it contained a T2 or not; and (3) whether it consisted of two, four or seven stimuli. Combinations of the first two factors defined the four trial types (*only-T1* vs. *only-T2* vs. *T1 + T2* vs. *no-target* trials). The third factor defined the delay of the last position of the middle part relative to the first position of the middle part (*lag 1*, *lag 3* and *lag 6* conditions). Note that only lag 3 condition is depicted here. Auditory stimuli were presented with an ISI of 40 ms. Stimulus duration was 40 ms except for T1 that had an overall duration of 83 ms.

Each trial started with the presentation of a black fixation cross with a duration of 340 ms without auditory stimuli. (The fixation cross remained on the screen until the end of the trial.) Thereafter, a sound sequence was played binaurally. The participant's task was to report after every trial whether they have heard only distractors, a single T1 (termed as "dashed" target in the instructions), a single T2 (termed as "clear" target in the instructions), or both T1 and T2. In the latter case, participants had to report the targets in the correct presentation order. The task did not include the discrimination of the three versions of T2. Participants had to give a non-speeded answer by pressing the corresponding key on the keyboard that was signed by stickers (in the case of reporting both targets, participants had to choose between the response options of T1 + T2 and T2 + T1 indicating the presentation order of the two targets). No accuracy feedback was provided during the actual experiment. After a response was given, participants could start the subsequent trial by pressing the space bar.

The experiment consisted of a practice and an experimental phase. After familiarizing with the possible sound sequences, participants worked through three practice blocks. In the first practice block, nine no-target and nine only-T1 trials, in the second practice block, nine no-target, and nine only-T2 trials were presented in a random order. In the third practice block, all the four types of trials were presented six times each in a random order. During the practice phase, accuracy feedback was provided after every response. The experimental phase consisted of eight experimental blocks of 36 trials each. An experimental session lasted about 50 minutes, including instructions and individual breaks. Up to five participants were tested parallel in separate cubicles.

Design and Statistical Analysis

Essentially we employed a 2 (trial type: only-T2 vs. T1 + T2) × 3 (lag: 1 vs. 3 vs. 6) repeated measures design on T2-HRs. T2-HR was conditioned on correct T1-performance, thus, only trials on which the presence or absence of T1, and also the order of the two targets were reported correctly were considered in this analysis. Additionally, we applied a repeated measures analysis with lag (1 vs. 3 vs. 6) as within-participants factor on T1-HRs on the dual-target trials.[20] Only trials with correct T2-performance (i.e., the presence of T2, and also the order of the two targets reported correctly) were considered for determining T1-HRs.

We used the multivariate approach to repeated measures analysis with a priori orthogonal contrasts (see also at Experiments 3-5) given our specific hypothesis about the pattern of HRs as a function of the relative temporal position of the two targets: According to the auditory AB literature we expected that the most marked T2-deficit emerges when T1 is followed directly by T2. In line with this hypothesis, our first contrast compared lag 1 condition with longer lag conditions (lag 3 and 6). The second orthogonal contrast was between the two later lags (lag 3 vs. lag 6).

RESULTS

A significance level of $\alpha = .05$ (two-tailed) was adopted for all analyses. Participants showed adequate overall performance in target detection indicating that they complied with the task requirements: Average HR for T1 on only-T1 trials was 86.1% (SD = 18.9%); average HR for T2 on only-T2 trials was 94.0% (SD = 6.7%). The average ACC on dual-target trials was 73.1% (SD = 19.7%).

[20] Note that (contrary to the only-T2 and dual-target trials) lag factor is not meaningful regarding only-T1 trials; thereby only-T1 trials were not included in this analysis.

Performance for T2

First, we examined the detection performance for T2 as a function of whether or not a T1 was presented and as a function of the serial position of T2 (i.e. lag condition). T2-HRs were conditioned on correct T1-performance. This analysis can be considered as the conventional analysis of AB (see, e.g., Shapiro et al., 1997). *Table 15* presents the mean HRs for T2 in each condition (see also *Figure 10* as a visual aid for interpreting the pattern of means). As indicated by *Figure 10*, matching expectations, T2 was associated with diminished performance as a function of the temporal distance between the targets when it was preceded by a successfully detected T1. Moreover, this deficit appeared to be restricted to the first temporal lag, thus to the situation when the two targets were presented consecutively without intervening distractors.

Table 15. Mean HRs (%) for T2 given correct T1-performance in Experiment 6; as a function of the serial position of T2 (lag condition) on only-T2 and dual-target trials; *SD* in parentheses.

	Trial type	
	only-T2	**T1 + T2**
Lag 1	96.0 (6.6)	85.8 (20.8)
Lag 3	96.1 (5.5)	93.7 (9.2)
Lag 6	97.4 (4.6)	94.7 (10.0)

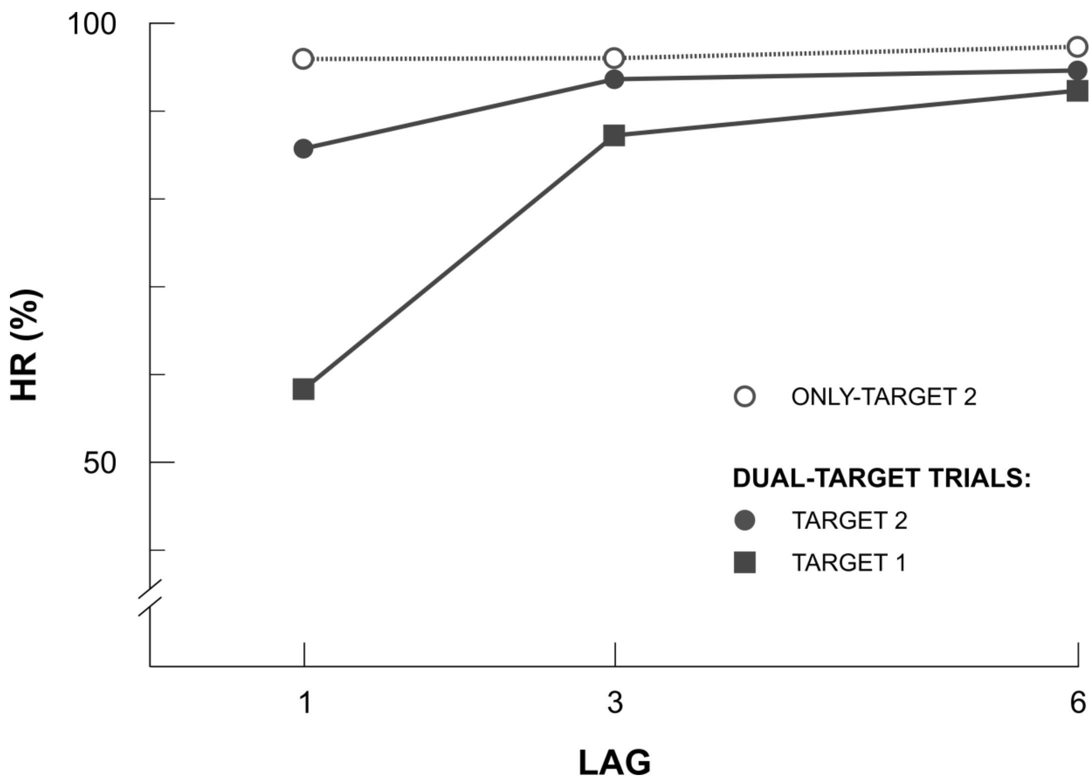


Figure 10. Mean HRs for Target 2 and Target 1 in Experiment 6.

Mean HRs for Target 2 on only-Target 2 and dual-target trials given correct Target 1-performance (represented by open and solid circles, respectively); and mean HRs for Target 1 on dual-target trials given correct Target 2-performance (represented by solid squares), as a function of the serial position of Target 2 (lag condition).

The 2 × 3 MANOVA for repeated measures with trial type (only-T2 vs. T1 + T2) and lag (1 vs. 3 vs. 6) as within-participants factors of T2-HRs showed a trial type main effect, as expected, with better detection performance when only T2 was presented without a preceding target, $F(1,34) = 10.01$, $p = .003$, $\eta_p^2 = .227$, and a lag main effect, $F(2,33) = 6.01$, $p = .006$, $\eta_p^2 = .267$. Importantly, the trial type main effect was significantly moderated by the temporal position of T2, $F(2,33) = 3.62$, $p = .038$, $\eta_p^2 = .180$, as a conventional index of AB. In line with our hypothesis about an auditory AB pattern, the trial type × lag 1 vs. longer lags (lag 3 and 6) interaction contrast was significant, $F(1,34) = 7.35$, $p = .010$, $\eta_p^2 = .178$, indicating reduced

T2-performance when the two targets were presented in direct succession. The trial type × lag 3 vs. lag 6 interaction contrast was not significant, $F(1,34) = 0.40$, $p = .842$, $\eta_p^2 = .001$.

Performance for T1

Figure 10 indicates a further robust pattern of HRs: On dual-target trials, not only T2 was associated with decreased detection performance when it was preceded by a correctly detected T1, but detection performance was also impaired for T1 when it was followed by a successfully detected T2. Moreover, this decrease in performance for T1 was even more pronounced than the performance deficit for T2. Importantly, an effect of T1-T2 temporal lag on the T1-HRs on dual-target trials cannot be considered as a conventional index of AB, but it conveys important information about short-ranged competition or interference between the targets.

In line with the analysis of T2-HRs, we conducted a second analysis on the HRs for T1 as a function of the serial position of T2 on dual-target trials. *Table 16* presents the mean HRs for T1 in each condition (see also *Figure 10* as a visual aid for interpreting the pattern of means). Detection HRs for T1 are reported conditioned on correct T2-detection. A MANOVA for repeated measures with lag (1 vs. 3 vs. 6) as within-participants factor of T1-HRs on dual-target trials showed a lag main effect, $F(2,33) = 25.86$, $p < .001$, $\eta_p^2 = .610$; with a significant lag 1 vs. longer lags (lag 3 and 6) a priori contrast, $F(1,34) = 53.20$, $p < .001$, $\eta_p^2 = .610$, indicating a pronounced degradation of detection performance for T1 when it was followed directly by a second target. The lag 3 vs. lag 6 contrast was also significant, $F(1,34) = 7.13$, $p = .012$, $\eta_p^2 = .173$.

Table 16. Mean HRs (%) for T1 given correct T2-performance in Experiment 6; as a function of the serial position of T2 (lag condition) on dual-target trials; and on only-T1 trials; *SD* in parentheses.

T1 + T2 trials:	
Lag 1	58.4 (31.3)
Lag 3	87.3 (14.5)
Lag 6	92.4 (11.1)
only-T1 trials:	
	90.5 (14.6)

Inversion Errors

The third analysis targeted the inversion errors, that is, the outcome when participants reported the presentation order of the two targets reversed. In the present experiment, 2.3% (*SD* = 3.6%) of the dual-target trials were associated with the report of reversed order. MANOVA for repeated measures with lag (1 vs. 3 vs. 6) as within-participants factor of inversion ER on dual-target trials showed a lag main effect, $F(2,33) = 3.71$, $p = .035$, $\eta_p^2 = .184$ ($M = 3.5\%$, $SD = 5.1\%$, for lag 1; $M = 1.9\%$, $SD = 4.5\%$, for lag 3; $M = 1.6\%$, $SD = 3.0\%$, for lag 6). Expectedly, inversion errors occurred most often when the two targets were presented directly after each other, $F(1,34) = 5.96$, $p = .020$, $\eta_p^2 =.149$, for the lag 1 vs. longer lags (lag 3 and 6) a priori contrast, while the lag 3 vs. lag 6 contrast was not significant, $F(1,34) = 0.43$, $p = .517$, $\eta_p^2 =.012$.

DISCUSSION

In Experiment 6, we investigated the basic characteristics of the dual-target detection deficit in a RASP paradigm. In line with previous AB studies, we found a lag-dependent performance decrement for T2 on dual-target trials. Similarly to the results found in the majority of auditory AB studies, this impairment did not show the U-shaped pattern of a “genuine” visual AB that is predicted by the (visual) theories of AB (see, e.g., the boost and bounce theory of AB that predicts preserved performance for T2 when the two targets are presented consecutively without intervening distractors due to an open “attentional gate”, Olivers & Meeter, 2008). Instead, the T1-related dual-target deficit manifested itself in a performance decrement at the first temporal lag in the present experiment, thus in the situation when the two targets were presented consecutively without intervening distractors. What is more, we found an even more pronounced performance deterioration for T1 when it was followed by a successfully detected T2. This deficit appeared most markedly when the two targets were presented in direct succession (see *Table 16* and *Figure 10*). Furthermore, inversion errors also occurred on a proportion of dual-target trials, meaning that participants reported that they heard T2 before T1, obviously, especially when the two targets followed each other directly.

Taken together, this pattern of results is not consistent with a “genuine” (visual) AB that is predicted by the AB theories outlined above. As expected according to the auditory AB literature, the most marked detection deficit for T2 indeed occurred at lag 1. However, the overall pattern of findings – thus, (1) a T1-related T2-deficit that is almost entirely restricted to lag 1, and (2) an even more pronounced T2-related T1-deficit suggesting mutual interference – does not require assuming a composite effect of a “pure” AB (i.e., as a longer lasting cognitive deficit) and additional short-range effects. The short-lived trade-off between the two targets is rather consistent with an interpretation that only the “additional” short-range deficit took effect.

Furthermore, taken together with the finding of a T2-related deficit on T1-detection and the absence of lag 1 sparing, inversion errors are not likely to reflect target integration in the present experiment.

This pattern of results can comprise disruption in several aspects of auditory temporal abilities like a failure in temporal ordering in the case of inversion errors and temporal masking between the two salient targets in the case of mutually diminished target detection. The auditory temporal or non-simultaneous masking refers to the impairment of auditory detection performance when a further sound is presented in close temporal vicinity. Sounds presented in rapid sequences can impair each other's perception due to two basic forms of temporal masking: During forward masking, a preceding sound interferes with the detection of the target sound; while during backward masking, a subsequent sound interferes with the detection of the previously presented target sound. Importantly, the time course of the short-range interference that was found in the present RASP paradigm is more comparable with the time course of temporal masking as it decays rapidly before about 100-200 ms (e.g., B. C. J. Moore, 2012), while a "genuine" AB-deficit is typically the "deepest" when T2 is presented in a time window of about 200-500 ms following T1.

Notably, the detection deficit for a target sound among rapidly presented and randomly varying sounds is often much more severe than it would be expected based on their overlapping activation patterns on the auditory periphery (the amount of masking due to acoustic overlap is often referred as "energetic masking", e.g., Brungart, Simpson, Ericson, & Scott, 2001). The additional deficit that is not attributable to "energetic masking" is termed as "informational masking", referring to the phenomenon that the detection threshold for a target sound is elevated among other competing sounds, even if the target is clearly above the hearing level. The shift in detection threshold occurs even if the competing sounds do not overlap in tone-frequency, and even if they are presented non-simultaneously in randomly varying, rapid sound sequences (Gutschalk, Micheyl, & Oxenham, 2008; Leek et al., 1991; Tolnai, Dolležal, & Klump, 2015). This poorly understood phenomenon is suggested to reflect an interference between potentially relevant auditory stimuli that are competing for perceptual awareness, with a deficit occurring at several stages of sound processing beyond the auditory periphery, and presumably involving processes such as perceptual grouping and segregation, and auditory object formation and selection (for reviews, see Kidd, Mason, Richards, Gallun, & Durlach, 2008; B. C. J. Moore, 2012; Shinn-Cunningham, 2008; Snyder, Gregg, Weintraub, & Alain, 2012). The interference effect between the two successive targets in the present RASP paradigm can be covered rather by the catchall term of informational masking that is accommodating at least partly perceptually-based and shorter-lived deficits on sound detection in rapid and random sequences than by the notion of a "genuine" AB (i.e., as a depletion of "costly" cognitive processes or a transient attentional suppression triggered by a first target event or the following distractors).

Taken together, we have started with the investigation of the temporal dynamics of auditory attention in a RASP paradigm. Although we found the most marked dual-target performance decrement at lag 1 as indicated by the majority of auditory AB studies, we are cautious to interpret the pattern of results as a deficit of auditory attention that can be considered as an equivalent of a (visual) AB for two reasons: First, the T1-related dual-target deficit on T2-performance was not only maximal at lag 1, but it was almost entirely attributable to lag 1 condition. Second, there was an even more pronounced performance deterioration for T1 when it was followed by T2. Thus, the present results rather reflect at least partly perceptually-based short-range interference between the targets due to temporal (informational) masking than a "genuine" AB.

Given this backdrop, in Experiment 7, we will investigate whether the auditory dual-target deficit is moderated by the affective valence of T2. Notably, informational masking shows sensitivity to a priori knowledge as reflected in training and cuing effects (Oxenham, Fligor, Mason, & Jr, 2003; Richards & Neff, 2004; Strait et al., 2010), and to perceptual saliency-based attentional capture, with a release from informational masking for perceptually salient targets, and a pronounced masking by perceptually salient competing sounds (e.g., Leek et al., 1991). Given the sensitivity of informational masking to attentional factors (e.g., Gutschalk et al., 2008; Richards & Neff, 2004; C. Zhang, Lu, Wu, & Li, 2014), we expect an advantage of valenced relative to neutral targets in the interference between temporally close targets as a consequence of "natural" attention to valenced tones. First of all, in line with the typical results of rapid serial visual presentation (RSVP) paradigms, we expected that this benefit will be reflected in a lag-dependent spared performance for valenced compared with neutral T2s.

Experiment 7: Preferential Selection of Valenced Targets in a RASP Paradigm?

In Experiment 7, we targeted the question whether valenced auditory stimuli can be selected preferentially from temporal sequences. We investigated this question by presenting simple tones with affective valence as the second target in the dual-target RASP paradigm used in Experiment 6. In order to avoid perceptual confounds, we assigned positive, negative, and neutral valences to tone-frequencies in a balanced design following the valence induction method of Experiments 3-5.

In RSVP paradigms, second targets with affective meaning remain relatively spared from AB deficit (A. K. Anderson, 2005; e.g., A. K. Anderson & Phelps, 2001; Ihssen & Keil, 2009; Keil & Ihssen, 2004; Keil et al., 2006; Oca et al., 2012; Raymond & O'Brien, 2009), showing a facilitated detection performance specifically for positive (Oca et al., 2012; Raymond & O'Brien, 2009), or valenced visual stimuli in general (A. K. Anderson, 2005; Ihssen & Keil, 2009; Keil et al., 2006; Keil & Ihssen, 2004; while A. K. Anderson & Phelps, 2001, investigated only negative valence). However, in Experiment 6, we found rather a short-lived competition between the targets in the auditory modality presumably due to temporal masking effects than a “genuine” AB, as indicated by a mutual degradation of T1 and T2 detection performance at lag 1. The question arises whether a preferential attention enhancement for valenced tones can bias the competition between the detection of temporally close targets in favor of a valenced target?

If tones endowed with affective meaning are prioritized in rapid temporal sequences by preferential attention, we can expect a relative detection facilitation and thereby, similarly to the typical results of affective RSVP paradigms, a reduction of the lag-dependent dual-target deficit for these tones. Furthermore, in line with the prior entry hypothesis applied to inversion errors (Olivers et al., 2011), we expect that the attentional benefit for valenced (positive and/or negative) T2s can be reflected in the higher proportion of lag-dependent inversion errors compared with neutral T2s. In line with our general question of whether basic auditory attentional biases to affective valence are characterized by a general relevance bias (as also demonstrated by Experiments 3-5) or a bias to a specific valence category (a negativity bias as indicated by many visual studies employing various paradigms, e.g., Öhman et al., 2001; or a positivity bias as has been demonstrated by visual affective AB studies, Oca et al., 2012; Raymond & O’Brien, 2009) in rapid auditory sequences, our hypotheses about the affective bias in the present paradigm can be stated as follows: (a) a general relevance bias in the mutual interference between the targets; or (b) a bias for a specific valence category in the mutual interference between the targets.

METHODS

Participants

Twenty-nine students from Saarland University (17 females; aged 20–36 years, *Mdn* = 24 years) participated in the experiment for monetary compensation (€15 on average; compensation was partly dependent on performance, see *Materials and Procedur*e for details). All participants were native German speakers, had normal or corrected-to-normal vision, and self-reported normal hearing. Before the experiment, all participants gave written informed consent. The assignment of tone-frequencies to valence conditions was counterbalanced between participants according to a Latin square scheme resulting in three counterbalancing groups (the counterbalancing scheme was identical to that used in Experiment 5, see *Table A2* in *Appendix B*). In the final sample, two counterbalancing groups had ten participants each, and one counterbalancing group had nine participants. Given a sample size of $N = 29$ and an α-value of .05 (two-tailed), effects size of $d_z = 0.54$ can be detected with a probability of $1 - \beta = .80$ for the a priori comparisons that are representing the our central hypotheses (see below, calculated with the aid of G*Power 3 software, Faul et al., 2007).

Materials and Procedure

Up to five participants were tested parallel in separate cubicles. Participants initially received €14 as a starting payment but were obliged to risk the money as the stakes in a subsequent "game". The experimental session lasted about 90 minutes, including instructions and individual breaks. The experiment consisted of (A) a *valence induction phase* that was identical to that applied in Experiment 5, and (B) a *RASP phase* using the same RASP task as in Experiment 6. In order to keep the duration of an experimental session in a convenient range for the participants, we did not employ manipulation check phase in the present experiment. However, note that Experiments 3-4 already showed the validity of the applied valence induction method. The valence induction phase consisted of one practice and three experimental blocks of a valence induction task. During the RASP phase, after three practice blocks (see above at Experiment 6), eight RASP experimental blocks were presented, and every two blocks of RASP task was interspersed by an additional block of the valence induction task (i.e., the two phases together consisted of six valence induction blocks and eight RASP blocks). Blocks were separated by short, participant-terminated breaks. Sounds were presented binaurally via headphones (K 511, AKG Acoustics, Harman International Industries, Inc.) at a comfortable maximal loudness of approximately 70 dB(A).

Valence induction phase

The valence induction phase of the present experiment was identical to the valence induction procedure of Experiment 5. In brief, participants could win or lose game scores depending on their performance, and that resulted in a win or loss of €1 at the end of each valence induction block. Sinusoidal tones with four different tone-frequencies (300 Hz, 510 Hz, 867 Hz, and 1473.9 Hz) were presented in two versions: (1) *standard tones* were presented with constant intensity; (2) *target tones* with increasing intensity. Participants were instructed to respond only to the target tones except for the highest tone that served as a no-go signal in all counterbalancing group. The tone-frequency determined the consequences of successful and unsuccessful performance on a given trial: One tone-frequency was associated with substantial gain in case of a success, but no negative consequences in case of a failure (*positive tone*), another tone-frequency was associated with substantial loss in case of a failure, but no positive consequences in case of a success (*negative tone*), a third tone-frequency was associated with a negligible gain or loss in case of a success and a failure, respectively (*neutral tone*).

RASP phase

Stimulus materials and procedure were identical to that in Experiment 6, except that every two blocks of the RASP task were interspersed by an additional block of the valence induction task. Importantly, note that tone-frequencies of the low, medium and high pitch versions of T2 in the RASP task are identical to the tone-frequencies of the three tones (with positive, negative, and neutral meaning) in the valence induction phase (i.e., 300 Hz or 510 Hz or 867 Hz; the assignment of tone-frequencies to valence conditions was counterbalanced between participants). Similarly to Experiment 6, one-third of the trials containing T2 was presented with low, one-third with medium, and one-third with high pitch versions of T2 (counterbalanced between trial type and lag conditions).

Design and Statistical Analysis

First, we employed a 2 (trial type: only-T2 vs. T1 + T2 trials) × 3 (lag: 1 vs. 3 vs. 6) × 3 (T2-valence: positive vs. negative vs. neutral) × 3 (three counterbalancing groups) mixed design with trial type, lag, and T2-valence as within-participants factors and counterbalancing group as a between-participants factor on T2-HRs. Only trials with correct T1-performance were considered for determining T2-HRs. Additionally, a 3 (lag: 1 vs. 3 vs. 6) × 3 (T2-valence: positive vs. negative vs. neutral) × 3 (three counterbalancing groups) mixed design with lag and T2-valence as within-participants factors and counterbalancing group as a between-participants factor was applied on T1-HRs. For determining T1-HRs, only trials with correct T2-performance were considered. Similarly to the approach of Experiments 3-5, we added counterbalancing group as a between-participants factor to use the correct error term and account for the slightly different sample sizes of the final counterbalancing groups. In the following, we do not report effects involving the group factor as they are not of interest in the present design (see Pollatsek & Well, 1995; and *Appendix A*).

We used the multivariate approach to repeated measures analysis with a priori orthogonal contrasts representing our specific hypotheses. Similarly to the approach applied in Experiments 3-5, we transformed the tripartite factor of valence into a vector of two orthogonal contrast variables (see, e.g., O'Brien & Kaiser, 1985; Petrova & Wentura, 2012; M. Rohr et al., 2012) in a way that represents our specific hypotheses. That is, for the first contrast, T1 and T2-HRs and inversion ERs were averaged across positive and negative T2-valence and contrasted with the neutral condition representing the hypothesis of a general relevance bias. The second contrast was the contrast between positive and negative T2-valence, representing the hypothesis of bias for a specific valence category. Additionally, in line with Experiment 6, we applied a priori orthogonal contrasts of lag condition, given our hypothesis about the most severe dual-target deficit occurring at lag 1 in the auditory modality. Thus, the first orthogonal contrast compared lag 1 condition with later lags (lag 3 and 6) representing our hypothesis about an auditory dual-target deficit. The second orthogonal contrast was between the two later lags (lag 3 vs. lag 6). Combinations of the above outlined a priori contrasts tested our specific hypotheses on T1 and T2-HRs and inversion ERs: (a) a general relevance bias in the auditory dual-target deficit represented by the valenced vs. neutral T2 × lag 1 vs. later lags (lag 3 and 6) interaction contrast; (b) a bias for a specific valence category in the auditory dual-target deficit represented by the positive vs. negative T2 × lag 1 vs. later lags (lag 3 and 6) interaction contrast.[21]

[21] Note that the valenced vs. neutral T2 × lag 3 vs. lag 6 and positive vs. negative T2 × lag 3 vs. lag 6 interaction contrasts are of minor interest in the present context. We will report these contrasts in parenthesis.

Results

A significance level of $\alpha = .05$ (two-tailed) was adopted for all analyses.

Valence Induction Phase

Behavioral measures of the valence induction phase were analyzed to ensure that participants complied with the task instructions, and thus they expectedly learned the associations between valence and tone-frequency. The behavioral performance on the valence induction phase was adequate, suggesting that participants were engaged in the task (see *Table 17*).

Table 17. Results of the valence induction phase in Experiment 7: Mean reaction times (RTs), accuracy rates (ACCs), false alarm rates (FARs), and hit rates (HRs) in each valence condition; *SD* in parentheses.

Valence	**RT**	**ACC**	**FAR**	**HR**
Positive	309 (74)	86.3 (15.6)	23.0 (32.2)	95.5 (5.7)
Neutral	325 (62)	90.1 (11.0)	9.4 (13.9)	89.7 (15.4)
Negative	319 (68)	91.0 (7.5)	9.5 (12.5)	91.5 (8.7)

RTs were calculated from the beginning of the loudness increase in the ongoing tone, and RTs below 100 ms were discarded when calculating averaged RTs. A 3 × 3 MANOVA for repeated measures with valence as within-participants factor and counterbalancing group as between-participants factor of the RTs did not show significant valence differences, $F(2,24) = 2.26$, $p = .126$, $\eta_p^2 = .159$. The a priori contrast of valenced vs. neutral conditions was significant, $F(1,25) = 4.56$, $p = .043$, $\eta_p^2 = .154$, indicating faster reactions to valenced compared with the neutral tones; while positive vs. negative valence did not differ significantly, $F(1,25) = 1.90$, $p = .180$, $\eta_p^2 = .071$.

The average ACC across participants was adequately high ($M = 89.1\%$, $SD = 9.7\%$). Similarly to Experiments 3-5, numerically lowest mean ACC emerged in the positive condition (see *Table 17*). ACCs did not show significant differences across valence conditions, $F(2,25) = 1.90$, $p = .171$, $\eta_p^2 = .132$. However, in line with a more liberal response criteria, ACC was marginally lower for positive compared with negative tones, $F(1,26) = 3.94$, $p = .058$, $\eta_p^2 = .132$, while valenced vs. neutral conditions did not differ significantly, $F(1,26) = 0.74$, $p = .397$, $\eta_p^2 = .028$. Similarly to Experiments 3-5, the relatively low mean ACC emerged due to a high proportion of false alarms in the positive condition. FARs showed a significant valence effect, $F(2,25) = 4.14$, $p = .028$, $\eta_p^2 = .249$. As expected, the a priori contrast of positive vs. negative conditions was significant, $F(1,26) = 8.10$, $p = .009$, $\eta_p^2 = .238$; and valenced vs. neutral conditions also differed significantly, $F(1,26) = 6.25$, $p = .019$, $\eta_p^2 = .194$. Analogue analysis on the HRs showed a tendency for valence differences, $F(2,25) = 3.07$, $p = .064$, $\eta_p^2 = .197$. The a priori contrast of positive vs. negative conditions was significant, $F(1,26) = 6.02$, $p = .021$, $\eta_p^2 = .188$, indicating higher HRs for positive compared with negative tones; while valenced vs. neutral conditions did not differ significantly, $F(1,26) = 2.49$, $p = .127$, $\eta_p^2 = .087$. Taken together, participants' performance in the valence induction phase indicated a high level of task compliance and application of strategies for achieving positive monetary outcomes.

RASP Phase

Participants showed adequate overall performance in target detection suggesting that they were engaged in the task: Average HR for T1 on only-T1 trials was 79.8% ($SD =$

22.9%); average HR for T2 on only-T2 trials was 89.9% (*SD* = 14.7%). The average accuracy on dual-target trials was 67.1% (*SD* = 24.2%).

Performance for T2

First we examined the detection performance for T2 as a function of its serial position and affective valence, and whether or not it was preceded by a T1. T2-HRs were conditioned on correct T1-performance. *Table 18* presents the mean HRs for T2 on only-T2 trials and on dual-target trials in each condition (see also *Figure 11* as a visual aid for interpreting the pattern of means).

Table 18. Mean HRs (%) for T2 given correct T1-performance in Experiment 7; as a function of the serial position (lag condition) and affective valence of T2 on only-T2 and dual-target trials; *SD* in parentheses.

	T2-valence		
	Positive	**Neutral**	**Negative**
only-T2 trials:			
Lag 1	93.7 (14.9)	97.0 (7.2)	98.3 (5.5)
Lag 3	93.8 (14.9)	96.1 (10.8)	97.7 (6.2)
Lag 6	93.5 (14.8)	95.0 (17.2)	98.6 (4.3)
T1 + T2 trials:			
Lag 1	70.0 (40.3)	79.1 (34.2)	76.7 (34.6)
Lag 3	86.5 (25.8)	90.7 (19.6)	91.7 (16.3)
Lag 6	88.7 (19.0)	93.3 (16.9)	93.3 (14.1)

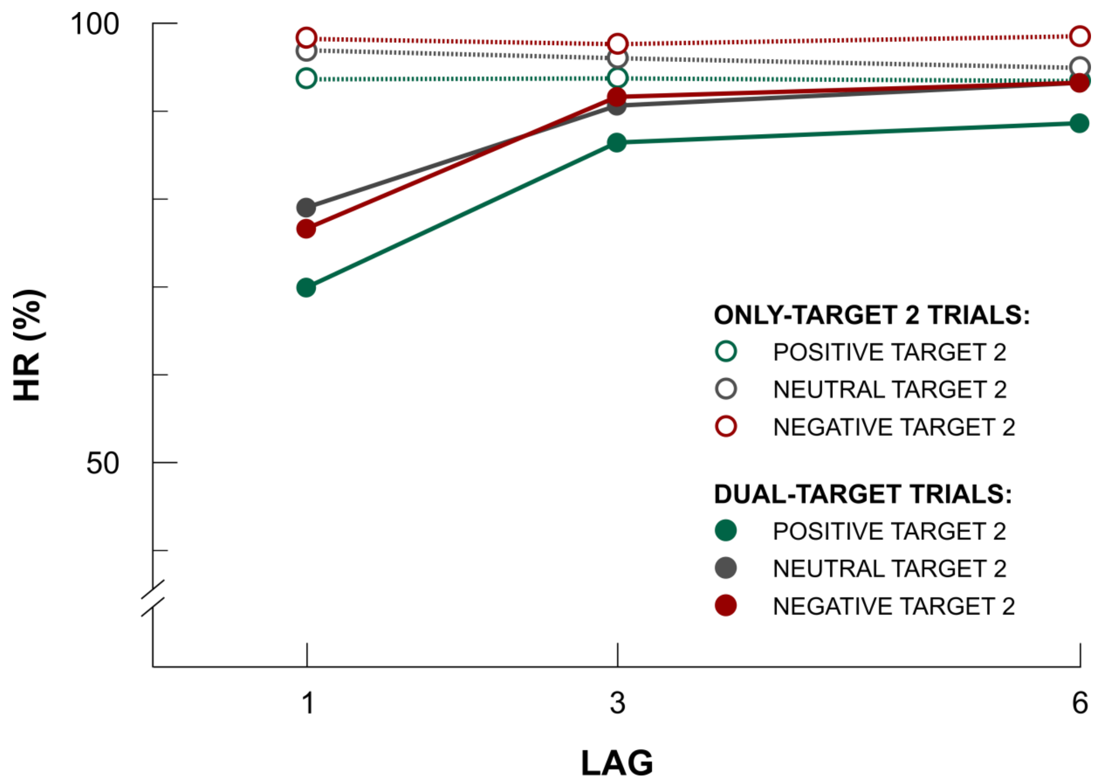


Figure 11. Mean HRs for Target 2 in Experiment 7.

Mean HRs for Target 2 given correct Target 1-performance as a function of the serial position (lag condition) and affective valence of Target 2 on only-Target 2 (open circles) and dual-target trials (solid circles).

As can be seen in *Figure 11*, matching expectations, on dual-target trials there was a lag-dependent impairment of T2-performance. Similarly to the results of Experiment 6, this impairment appeared most pronounced when the two targets were presented in direct succession. The 2 × 3 × 3 × 3 MANOVA for repeated measures with trial type (only-T2 vs. T1 + T2), lag (1 vs. 3 vs. 6), and T2-valence (positive vs. negative vs. neutral) as within-participants factors and counterbalancing group as a between-participants factor of T2-HRs showed a trial type main effect, according to expectations, with higher HRs when only T2 was presented without preceding T1, $F(1,26) = 12.83$, $p = .001$, $\eta_p^2 = .330$; and a lag main effect, $F(2,25) = 3.83$, $p = .035$,

η_p^2 = .235. As expected, the trial type main effect was significantly moderated by the temporal position of T2, $F(2,25) = 4.88$, $p = .016$, η_p^2 =.281; with a significant a priori trial type × lag 1 vs. later lags (lag 3 and 6) interaction contrast, $F(1,26) = 10.14$, $p = .004$, η_p^2 =.281, representing a short-lived dual-target deficit. The trial type × lag 3 vs. lag 6 interaction contrast was not significant, $F(1,26) = 1.41$, $p = .245$, η_p^2 =.052.

T2-valence did not show significant differences on T2-performance across lag conditions and trial types, $F(2,25) = 1.83$, $p = .181$, η_p^2 = .128. The a priori contrast of positive vs. negative conditions indicated a tendency for better detection performance for negative compared with positive T2, $F(1,26) = 3.75$, $p = .064$, η_p^2 = .126 (M = 92.7%, SD = 10.0% for negative, and M = 87.7%, SD = 17.0% for positive T2). The valenced vs. neutral comparison was not significant, $F(1,26) = 0.63$ $p = .436$, η_p^2 = .024 (M = 90.2%, SD = 12.2% for valenced, and M = 91.9%, SD = 11.6% for neutral T2).

The 2 (trial type: only-T2 vs. T1 + T2) × 3 (lag: 1 vs. 3 vs. 6) × 3 (T2-valence: positive vs. negative vs. neutral) interaction was not significant, $F(4,23) = 0.34$, p = .849, η_p^2 =.056. Similarly, the a priori interaction contrasts were not significant: (a) trial type (only-T2 vs. T1 + T2) × valenced (positive and negative) vs. neutral T2 × lag 1 vs. later lags (lag 3 and 6) contrast representing a general relevance bias in the short-range auditory dual-target deficit, $F(1,26) = 0.21$, $p = .648$, η_p^2 = .008; and (b) trial type (only-T2 vs. T1 + T2) × positive vs. negative T2 × lag 1 vs. later lags (lag 3 and 6) contrast representing a bias for a specific valence category in the short-range auditory dual-target deficit, $F(1,26) = 0.13$, $p = .722$, η_p^2 = .005. ($F[1,26] = 0.34$, p = .564, η_p^2 = .013, for the corresponding trial type × valenced vs. neutral T2 × lag 3 vs. lag 6 contrast; $F[1,26] = 0.17$, $p = .684$, η_p^2 = .006, for the trial type × positive vs. negative T2 × lag 3 vs. lag 6 contrast). No other significant interaction emerged in the present design, Fs < 0.55, *n.s.*

Performance for T1

In order to reveal possible valence effects in the auditory dual-target deficit related to a successfully detected T2, we conducted an analysis on the HRs for T1 conditioned on correct T2-performance. *Table 19* presents the HRs for T1, as a function of the serial position and valence of T2 on dual-target trials; and on only-T1 trials (see also *Figure 12* as a visual aid for interpreting the pattern of means). Similarly to the results found in Experiment 6, *Figure 12* indicates an impairment of detection performance for T1 when it was followed by a correctly detected T2, with the most marked impairment at lag 1.

Table 19. Mean HRs (%) for T1 given correct T2-performance in Experiment 7; as a function of the serial position (lag condition) and affective valence of T2 on dual-target trials; and on only-T1 trials; *SD* in parentheses.

	T2-valence		
	Positive	**Neutral**	**Negative**
T1 + T2 trials:			
Lag 1	54.2 (36.2)	59.7 (35.4)	53.0 (32.6)
Lag 3	86.1 (22.7)	84.4 (24.7)	85.3 (25.2)
Lag 6	86.0 (20.4)	84.9 (22.8)	90.7 (18.8)
only-T1 trials:			
87.3 (19.0)			

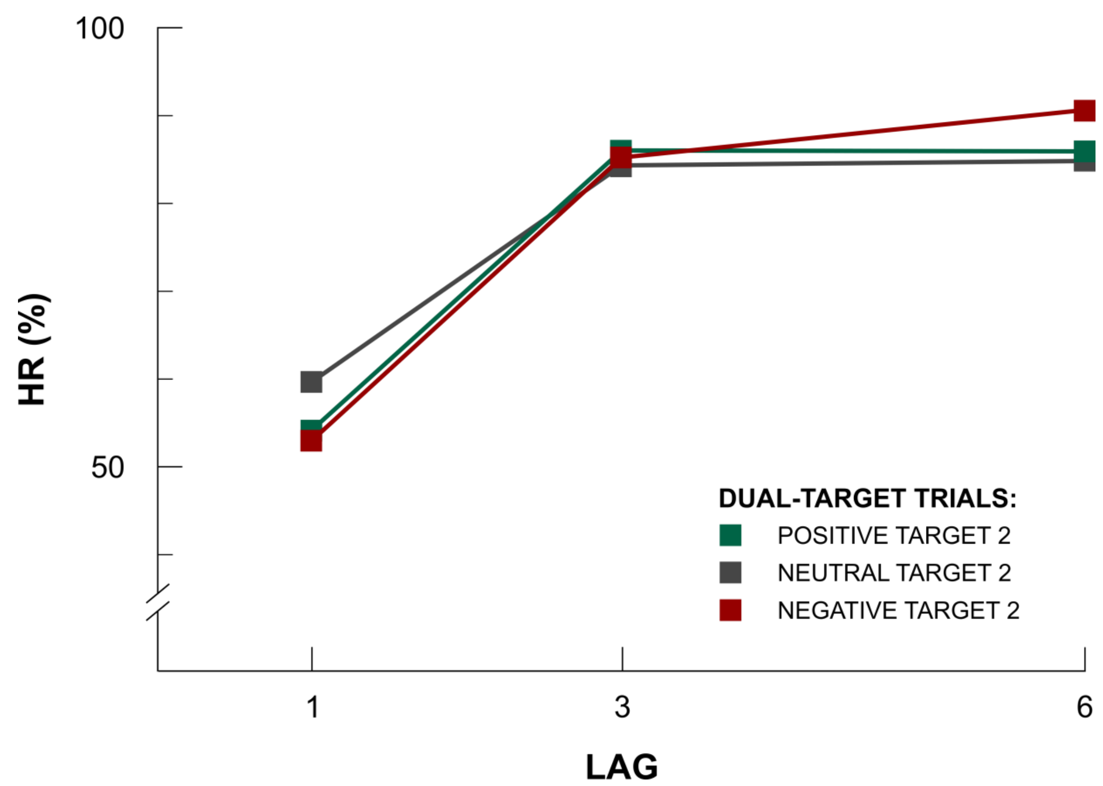


Figure 12. Mean HRs for Target 1 in Experiment 7.

Mean HRs for Target 1 given correct Target 2-performance as a function of the serial position (lag condition) and affective valence of Target 2 on dual-target trials.

The 3 × 3 × 3 MANOVA for repeated measures with lag (1 vs. 3 vs. 6), and T2-valence (positive vs. negative vs. neutral) as within-participants factors and counterbalancing group as a between-participants factor of T1-HRs showed the expected lag main effect, $F(2,25) = 34.90$, $p < .001$, $\eta_p^2 = .736$. The a priori contrast of lag 1 vs. longer lags (lag 3 and 6) representing the mutual auditory dual-target deficit was significant, $F(1,26) = 70.73$, $p < .001$, $\eta_p^2 = .731$, indicating a pronounced degradation of detection performance for T1 when it was followed by a T2 directly. The contrast of lag 3 vs. lag 6 conditions was not significant, $F(1,26) = 0.71$, $p = .408$, $\eta_p^2 = .026$.

There was no significant difference as a function of T2-valence across lag conditions, $F(2,25) = 0.28$, $p = .757$, $\eta_p^2 = .022$ ($F[1,26] = 0.25$, $p = .875$, $\eta_p^2 = .001$, for the valenced vs. neutral comparison; and $F[1,26] = 0.45$, $p = .507$, $\eta_p^2 = .017$, for the positive vs. negative comparison).

The lag main effect was not moderated significantly by the affective valence of T2, $F(4,23) = 0.93$, $p = .462$, $\eta_p^2 =.140$. The a priori interaction contrast of (a) valenced (positive and negative) vs. neutral T2 × lag 1 vs. later lags (lag 3 and 6) representing the hypothesis of a general relevance bias in the short-range auditory dual-target deficit, just missed the conventional level of significance, $F(1,26) = 4.13$, $p = .053$, $\eta_p^2 =.137$, indicating a tendency for more degraded T1-performance when it was followed directly by a valenced T2 compared with a neutral T2 ($F[1,26] = 0.51$, $p = .482$, $\eta_p^2 = .019$, for the valenced vs. neutral T2 × lag 3 vs. lag 6 contrast). The a priori interaction contrast of (b) positive vs. negative T2 × lag 1 vs. later lags (lag 3 and 6) representing the hypothesis of a bias for a specific valence category in the short-range auditory dual-target deficit was not significant, $F(1,26) = 0.14$, $p = .715$, $\eta_p^2 =.005$ ($F[1,26] = 1.32$, $p = .261$, $\eta_p^2 = .048$, for the positive vs. negative T2 × lag 3 vs. lag 6 contrast).

Inversion errors

In the present experiment, participants reported the targets in a reversed order on 3.9% (*SD* = 5.7%) of the dual-target trials. *Table 20* presents inversion ERs on dual-target trials as a function of the serial position and valence of T2.

Table 20. Mean inversion ERs (%) on dual-target trials in Experiment 7; as a function of the serial position (lag condition) and affective valence of T2; *SD* in parentheses.

	T2-valence		
	Positive	**Neutral**	**Negative**
Lag 1	8.2 (13.1)	4.7 (10.3)	5.2 (8.5)
Lag 3	5.2 (10.8)	3.5 (8.8)	3.9 (8.9)
Lag 6	1.7 (5.5)	0.9 (3.2)	2.2 (6.7)

A 3 × 3 × 3 MANOVA for repeated measures with lag (1 vs. 3 vs. 6), and T2-valence (positive vs. negative vs. neutral) as within-participants factors and counter-balancing group as a between-participants factor of inversion ERs on dual-target trials showed a lag main effect, $F(2,25) = 3.87$, $p = .034$, $\eta_p^2 = .236$. As expected, the lag 1 vs. longer lags (lag 3 and 6) contrast was significant, $F(1,26) = 5.44$, $p = .028$, $\eta_p^2 = .173$. Moreover, there was a significant difference between lag 3 and lag 6, $F(1,26) = 5.22$, $p = .031$, $\eta_p^2 = .167$, indicating more inversion errors at lag 3 compared with the longest lag.

T2-valence did not yield significant differences, $F(2,25) = 2.48$, $p = .104$, $\eta_p^2 = .166$. However, the a priori contrast of valenced (positive and negative) vs. neutral T2 was significant, $F(1,26) = 4.78$, $p = .038$, $\eta_p^2 = .155$, indicating more frequent reversed order-reports for valenced T2 compared with neutral T2 across lag conditions; while the comparison of negative vs. positive T2-valence did not show any difference, $F(1,26) = 0.87$, $p = .360$, $\eta_p^2 = .032$. The lag main effect was not moderated by the affective valence of T2, $F(4,23) = 0.46$, $p = .765$, $\eta_p^2 = .074$. None of the comparisons representing our a priori hypotheses was significant, thus, (a) valenced (positive and negative) vs. neutral T2 × lag 1 vs. later lags (lag 3 and 6) representing the hypothesis of a general relevance bias in the short-range dual-target deficit, $F(1,26) = 0.11$, $p = .746$, $\eta_p^2 = .004$ ($F[1,26] = 0.01$, $p = .941$, $\eta_p^2 < .001$, for the valenced vs. neutral T2 × lag 3 vs. lag 6 contrast); and (b) positive vs. negative T2 × lag 1 vs. later lags (lag 3 and 6) representing the hypothesis of a bias for a specific valence category in the short-range dual-target deficit, $F(1,26) = 1.18$, $p = .287$, $\eta_p^2 = .043$ ($F[1,26] = 0.57$, $p = .459$, $\eta_p^2 = .021$, for the positive vs. negative T2 × lag 3 vs. lag 6 contrast).

DISCUSSION

First, in Experiment 7, we replicated the short-range interference between the two targets that was found in Experiment 6. That is, when the two targets were presented in direct succession, there was a severe impairment on T2-performance in the case of successful T1-detection, and in turn there was also an even more pronounced deficit on T1-performance in the case of successful T2-detection. Second, the main question of Experiment 7 was whether affectively significant targets receive a detection benefit in the above-described auditory mutual dual-target deficit due to preferential attention to them. Although there was a tendency for performance benefit for negative relative to positive targets across lag conditions, in contrast to studies on affective modulation of visual AB, we did not observe more preserved performance for positive and/or negative T2s in the lag-dependent dual-target deficit. However, in turn, we found marginally more pronounced impairment on the detection performance for T1 when it was followed directly by a correctly detected valenced T2 relative to a neutral T2. Furthermore, participants reported more often that they heard T2 before T1 when it was endowed with affective valence; however, this “priority” for valenced T2 was not related to the time lag between the two targets.

Taken together, contrary to the findings on affective AB-sparing in the visual modality, we did not find any indication for a lag-dependent spared detection performance for positive and/or negative T2s in the auditory modality. However, we found a marginally significant indication that a valenced T2 more readily "wins" the direct competition between the two targets than a neutral T2, as reflected in more impaired T1-performance when it was directly followed by a valenced T2. This marginally significant pattern suggests an enhanced weighting of valenced (i.e., positive and negative) T2s compared with neutral T2s at the expense of T1-detection in the short-ranged competition between the targets.

Intermediate Summary: Preferential Selection of Valenced Tones from Temporal Sequences

Despite its key role in our auditory perception, temporal aspects of auditory processing have not been addressed yet by affective research explicitly. In the present section, we have targeted the question whether valenced tones are selected preferentially in a RASP paradigm. Previous visual and auditory studies investigating dual-target detection in rapid serial presentation paradigms often reported a performance decrement for T2 after a correctly detected T1, with the severeness of this deficit being dependent on the temporal distance between the targets. This phenomenon termed as AB is assumed to indicate a transient attentional deterioration in an about 500 ms long time window following a successfully detected T1 (for reviews, see, e.g., Dux & Marois, 2009; Martens & Wyble, 2010; Shapiro et al., 1997). It is often accompanied by the intriguing outcomes of lag 1 sparing (i.e., relatively preserved performance for T2 when it follows T1 directly) and inversion errors (i.e., participants report the presentation order of the two targets reversed at short temporal lags). Explanations of these outcomes have a critical role in more recent theoretical accounts of AB, for example by assuming integration of the targets as a common explanation for both outcomes (e.g., Akyürek et al., 2012), or an open "attentional gate" in the case of lag

1 sparing (Olivers & Meeter, 2008). Of most relevance to the present work, the visual AB is modulated by the affective significance of the targets. Valenced T2s were typically found to "break through" the visual AB deficit presumably as affectively significant stimuli attract early preferential attention that can lead to further facilitation, or as they have a more privileged access to higher stages of processing (A. K. Anderson, 2005; A. K. Anderson & Phelps, 2001; Ihssen & Keil, 2009; Keil & Ihssen, 2004; Keil et al., 2006; Oca et al., 2012; Raymond & O'Brien, 2009; Schwabe et al., 2011; D. Zhang et al., 2014).

Compared with the numerous visual studies, the notion of AB is much less established in the auditory modality, with the majority of the handful of auditory studies pointing toward the absence of the lag 1 sparing phenomenon (e.g., Arnell & Jenkins, 2004; Arnell & Jolicœur, 1999; Duncan et al., 1997; Horváth & Burgyán, 2011; Mondor, 1998; Shen & Mondor, 2006; Vachon et al., 2010; Vachon & Tremblay, 2005). As a possible explanation for this outcome, auditory AB has been suggested to reflect the sum of several effects including a "genuine" AB that lasts about 500 ms long, and more short-lived and presumably lower-level deficits that are maximal at lag 1 (e.g., deficit in temporal processing abilities, auditory distraction, switching between tasks, stimulus sets, or tone-frequency regions, see, e.g., Horváth & Burgyán, 2011; Tremblay et al., 2005).

Our main question was whether affectively significant T2 tones can break through the auditory dual-target deficit. First, in Experiment 6, we introduced a RASP paradigm in order to investigate the basic temporal characteristics of the dual-target deficit in rapid sound sequences. Second, in Experiment 7, we introduced an affective version of this paradigm in order to reveal possible preferential selection of valenced information during dual-target detection. Given the vital importance of selection of significant cues from the transient auditory input, we expected a marked sparing of affectively significant T2s in the auditory dual-target deficit.

In both experiments, in accordance with the majority of auditory AB studies, we did not find a U-shaped pattern of a “genuine” AB that is predicted by many (visual) AB theories (e.g., Olivers & Meeter, 2008) and was demonstrated previously mainly by visual studies (i.e., spared performance when the two targets follow each other directly without intervening distractors, diminished performance when distractors intervene between the targets within the AB-time window, and recovered performance after about 500 ms). On the contrary, in our RASP paradigm, the dual-target deficit was attributable mainly to the lag 1 condition, and it was reflected not only in a T1-dependent T2-deficit, but also in an even more pronounced T2-dependent deficit on T1. A short-range competition was also reflected in the inversion errors. By and large, this pattern of results does not likely reflect a difficulty of engaging voluntary attention repeatedly to successive target events within a brief time period and/or when non-target events intervene between the targets, but rather a short-range competition between the targets. Thus, we interpret the present results as a mutual competition between the targets that is likely, at least partly, of perceptual origin (i.e., informational masking as a short-range competition for auditory perceptual awareness) rather than a “genuine” AB that is predicted mainly by visual theories.

This observation lead us to the main question of the present section: Is there a detection benefit for valenced tones in the short-lived competition between the targets? In a nutshell, we did not find indications for the expected sparing of valenced T2s in the T1-related detection deficit. However, in turn, we found a marginally significant benefit for valenced compared with neutral T2s in the short-range competition with T1 as reflected in more impaired T1-performace when directly followed by a valenced compared with neutral T2. These results indicate a weak evidence for a preferential bias to affectively significant tones in the short-ranged competition between the targets for auditory perceptual awareness.

The present pattern of results raises several further questions: What perceptual and/or cognitive processes are involved in this short-range competition effect, and by what mechanism can affectively significant information win this competition for auditory perceptual awareness? Do our results reflect different processes than the deficit termed as AB in the auditory domain? As we have already suggested, the brief time course of the effect and the trade-off between the targets in detection performance suggest at least partly perceptually-based interference between the targets rather than a "genuine" AB. However, the mechanism of informational masking is similarly scarcely understood as the mechanism of auditory AB. Moreover, the phenomenon of informational masking is assumed to include cognitive processes that can be also involved in the AB, like attentional (mis)selection (for reviews, see, Kidd et al., 2008; B. C. J. Moore, 2012). Notably, recent evidence indicates that informational masking can occur as a suppression at a comparable level of auditory processing as the early differential attentional enhancement for valenced tones found in Experiments 3-5 (as reflected in reduced auditory N1 during informational masking, C. Zhang et al., 2014; or at comparable source location in the secondary auditory cortex as the location of the temporal generators of the N1, Gutschalk et al., 2008). Moreover, voluntary attention focused on the target in an informational masking paradigm results in enhanced N1 amplitude elicited by the target (C. Zhang et al., 2014). In light of these findings, we can speculate that valenced target tones could receive facilitation already at the level of perceptual analysis due to preferential attention to these tones in the present paradigm, in line with the results of Experiments 3-5, leading to relatively (and modestly) enhanced representation for these tones in the competition between the two target tones during informational masking.

In the present paradigm, it is especially difficult to interpret preferential valence effects that are independent of the temporal distance between two targets (see the slightly better detection performance for negative compared with positive T2s and the higher proportion of inversion errors for valenced T2s across lag conditions). For example, as the report of presentation order was operationalized in the present experiments as two separate response options ("T1 + T2" and "T2 + T1"), a bias for choosing the second, otherwise infrequently used response option in the case of a more salient T2 could also lead to higher proportion of inversion errors for valenced tones.

AFTER ALL, STILL AN AUDITORY ATTENTIONAL BLINK?

While we interpreted the above described short-lived dual-target performance decrement as an at least partly perceptually-based interference (i.e., informational masking), one can argue that this pattern of results – the absence of lag 1 sparing, mutual interference between the targets, and the brief time course – can be a special characteristic of a "genuine" AB in the auditory modality. As the auditory system has generally better temporal resolution than the visual system, it is possible that two successive targets are processed more likely sequentially than in a single episode in the auditory domain compared with vision (see also the argumentation of Shen & Mondor, 2006). Thus, as targets are not likely to merge into a single episode when presented in direct succession, lag 1 sparing is not likely to occur in the auditory modality, and the two initial target representations can mutually compete with each other for higher stages of processing. Moreover, auditory attention might be recovered faster from an AB-like deficit compared with visual attention as an explanation for the rapid time course of our effect. Notably, this interpretation resembles resource depletion accounts of the "genuine" (visual) AB at a faster time scale. Importantly, a trade-off between the identification of the two targets depending on their temporal distance and relative salience has been also suggested as an explanation of a "genuine" AB at short lags (see the two-stage competition model of visual attention as an extension of the original two-stage model of AB, Potter, Staub, & O'Connor, 2002; and also Hommel & Akyürek, 2005). In line with this points, most of the auditory AB studies indeed did not demonstrate lag 1 sparing but rather found the most pronounced effects at lag 1, and also a T2-related T1-degradation is not an uncommon outcome in the auditory modality (see, e.g., Duncan et al., 1997; Horváth & Burgyán, 2011).

Presumably, deficits in auditory perceptual abilities and attentional processes are interwoven in a complex way during rapid multiple target detection in the auditory modality. However, taken together, this phenomenon does not appear as a direct equivalent of the visual AB but likely shaped by unique features of the auditory modality like its fine temporal resolution. As informational masking has an earlier locus compared with a typical "genuine" AB effect (i.e., involving the time range of N1 and even earlier processes in the primary auditory cortex, likely related to auditory perceptual awareness; Gutschalk et al., 2008; Wiegand & Gutschalk, 2012), more direct electrophysiological measures could clarify the above outlined questions about the underlying processes.

CONCLUSIONS

To sum it up, our knowledge about the underlying mechanisms of auditory AB and auditory informational masking is limited, and there is no clear line between these two phenomena. Moreover, our understanding is the modest about whether detection of affectively significant auditory stimuli can be prioritized during these deficits. The two studies presented in this section demonstrated (1) a short-lived deficit during multiple target detection in a RASP paradigm that cannot be considered as a direct equivalent of a visual AB; (2) a weak indication for preferential selection of affectively significant tones under the conditions in which temporal abilities of the auditory system are limited; and (3) this weak indication occurred distinctively from the typical visual affective AB-sparing, as it was manifested in an enhanced weighting of the valenced relative to neutral T2s at the expense of T1-detection.

General Discussion

Without doubt, efficient selection of affectively significant auditory stimuli from the rapidly changing acoustic environment is essential for our survival and well-being. However, till now, our understanding on auditory affective processing is relatively sparse compared with the vast number of studies considering only visual modality. The present work aimed to provide a systematic investigation of automatic sound evaluation and preferential attention for affectively significant tones, while emphasizing one of the most conspicuous characteristics of auditory modality, that is, the importance of the temporal aspect.

Time plays a central role in the auditory domain, as sounds are characterized by rapid changes and complex temporal structure. Correspondingly to the basic importance of time and refined temporal abilities in the auditory modality, we argued that our auditory system should allow for fast and efficient evaluation and selection of significant cues from our dynamic acoustic environment. However, gradual evolving of auditory information due to the complex and extended temporal structure of sounds calls for some caution (cf. the immediately available affective content of emotional pictures), and we argued that this caution could contribute to the relative neglect of investigation into rapid automatic affective processes in the auditory modality.

Concerning the question of automaticity, we adopted a recent decompositional approach (Moors, 2015; Moors & De Houwer, 2006). Accordingly, we specified certain aspects of automaticity that we targeted by our investigations, such as fast and unintentional affective processes. Moreover, we asked the question whether preferential attention for affectively significant stimuli can arise involuntarily; thereby valenced tones can receive (1) preferential processing at an early, perceptual level of sound encoding even outside of the focus of voluntary attention, and (2) prioritized detection when task-relevant attentional processes are limited in time.

Throughout this work, we considered the basic positive versus negative affective valence when investigating evaluations and attentional prioritization of affective auditory stimuli. Given this backdrop, we targeted the question whether basic affective biases in auditory attention are driven by a specific valence category (especially by negative valence), the relevance of affectively significant information in general, or, instead of fixed and rigid biases, affective attentional prioritization varies flexibly regarding the challenges of different affective-motivational contexts. In the following, we will give a summary of the results concerning these aspects of investigation, and we will elaborate on the limitations of the conducted studies. Furthermore, we will delineate possible directions for future research.

Automatic Affective Processes in the Auditory Modality

FAST EXPLICIT AND UNINTENTIONAL EVALUATION OF NATURAL EMOTIONAL SOUNDS

The first research route of this thesis investigated the boundaries for evaluation of complex, natural emotional sounds concerning the available time and the intentionality of the evaluation process. The results of Experiment 1 indicated that complex, natural sounds can be evaluated already after very brief exposure time despite the apparent drawback of gradually evolving sound information. First, explicit valence ratings revealed that valence of natural emotional sounds can be evaluated validly even if sound segments of only 400 and 600 ms long duration – and with some limitation even sound segments as short as 200 ms – are available to base the judgment on. Second, we found evidence that early evaluations, at least partly, can be driven by rapid semantic identification of the sounds.

Furthermore, Experiment 2 demonstrated that valence of natural emotional sounds can be evaluated implicitly in a speeded, RT-based paradigm. We introduced an auditory version of the affective Simon task, in which participants had to give intrinsically valent responses (i.e., uttering "good" or "bad") to a valence-neutral perceptual stimulus feature of emotional sounds (i.e., illusory movement of the sound source to the left or right side) while they were asked to ignore every further feature of the presented sounds except for the task-relevant manipulation. We found that participants responded slower if the valence of the response and the valence of the sound mismatched. As this effect emerged even though participants were instructed to ignore stimulus valence, and even though the task-relevant feature was varied orthogonally to valence and was purely perceptual (i.e., no semantic encoding of the sound was necessary), we can conclude that sound valence was processed automatically in the sense as unintentionally.

Taken together, the results of Experiments 1-2 suggest that affectively significant natural sounds can be evaluated (1) rapidly with an extremely high precision, and (2) even without (conscious) intention. Furthermore, fast sound evaluations can be based, at least partly, on the complex semantic meaning of natural sounds instead of mere combinations of low-level acoustic cues (e.g., loudness or spectral entropy contour; e.g., Weninger et al., 2013). Although information content of complex, natural sounds is typically distributed in time, valence information can be obtained after short presentation time in contrast to the assumption of a slow affective evaluation in the auditory modality. However, while a relatively short exposure time was already sufficient for task-irrelevant sound valence to influence behavioral responses to a task-relevant feature in the auditory affective Simon paradigm, results of Experiment 2 also suggest that unintentional evaluation of complex sounds may occur somewhat slower compared with the situation when explicit instructions were given for evaluation, indicating that the lack of intentionality might have been compensated by longer evaluation.

INVOLUNTARY ATTENTIONAL ENHANCEMENT FOR VALENCED TONES

Given the crucial importance of rapid and efficient selection of auditory stimuli that are relevant for our survival and well-being, our second research route targeted the following question: Can affectively significant tones receive prioritized processing already at an early, perceptual stage of sound encoding in a relatively automatic manner? While the first research route examined evaluation of complex, natural emotional sounds with strong, well-defined intrinsic valence, the second and third research routes can be considered as a complementary approach: We assigned affective valence to tone-frequencies experimentally during a learning phase in a balanced design in order to avoid stimulus confounds and to have a strict control over the time of availability of the valance-related information.

The main finding of Experiment 3 is that valenced tones were processed preferentially compared with neutral ones already at an early, perceptual stage of sound encoding (within about 100 ms after tone onset) as reflected in the enhanced N1 amplitude for valenced tones, suggesting that processing of affectively significant tones can be facilitated by rapid attention. Furthermore, this early preferential enhancement for valenced relative to neutral tones emerged when these tones were entirely task-irrelevant, and participants were engaged in a demanding concurrent task, suggesting that encoding of valenced tones has been facilitated relatively automatically by “natural” selective attention.

In Experiment 4, we employed a stricter control over the direction of voluntary attention in order to investigate whether a differential attentional enhancement for valenced relative to neutral tones can occur strictly outside of the focus of voluntary attention. We found a marginally significant valence effect in our standard analysis on the N1 amplitude in this experiment, and we found indication for a very early valence effect on the P1-N1 complex. As this pattern emerged when valenced tones were not only task-irrelevant and to-be-ignored but participants’ voluntary attention was devoted to a continuous task-relevant stimulation, we interpret this pattern of results as an indication for an involuntary attentional capture by valenced tones. In Experiment 5, we targeted the question whether in salient positive or negative motivational contexts negative and positive tones can receive early differential attentional enhancement in line with the assumption of flexible affective attentional biases. Experiment 5 provided a replication of the early attentional enhancement on the N1 for valenced relative to neutral tones; however, this effect was not moderated by anticipating positive or negative future outcomes.

Although the differential attentional effect on the N1 amplitude for valenced compared with neutral tones appeared admittedly weaker in the single Experiment 4 and also in Experiment 5 relative to Experiment 3, a weakening of the valence effect was not confirmed statistically by a combined analysis of the three experiments. In this analysis, we found a clear enhancement of N1 amplitudes for valenced relative to neutral tones across experiments, and, importantly, this attentional enhancement was comparable across the three experiments. This pattern of results occurred despite that the single experiments employed different degrees of attentional and motivational demands on the concurrent task. This "immunity" to the changes of the competing task demands also supports our interpretation that the early preferential attentional effect for valenced tones occurred efficiently, thus relatively automatically.

In conclusion, the pattern of our results is in accordance with the assumption that affectively significant stimuli can receive attentional enhancement in a reflexive way independently from more voluntary controlled attentional processes (at least under such circumstances when the concurrent attentional load is not extreme, Pourtois et al., 2013; Vuilleumier, 2005; Vuilleumier & Huang, 2009). We can conclude that preferential attention to valenced tones occurred automatically in the terms as it is fast, unintentional, and can emerge in the absence of voluntary attention. Moreover, the combined analysis across experiments suggests that preferential attention to task-irrelevant valenced tones can occur relatively independently of changes in concurrent task demands.

We would like to highlight that a strength of these experiments is the strict control over arbitrary physical differences that was achieved by assigning a priori neutral tones with positive, negative, and neutral meaning through a learning phase. It is important to note that the enhancement of the N1 amplitude for valenced compared with neutral stimuli across the three experiments cannot be explained by the variable physical feature of tone-frequency, as it was counterbalanced between participants in each experiment. Thus, contrary to the typical approach of affective studies comparing physically complex stimuli, we can conclude that the preferential processing of valenced relative to neutral tones occurred due to their affective relevance in the present experiments and not as a response to certain salient perceptual features.

Taken together, the three experiments suggest that affectively significant auditory stimuli can receive preferential attention relatively automatically at an early level of sound encoding. Notably, this “natural” affective attentional effect to valenced tones closely resembles the effect of highly focused attention on early auditory processing (i.e., enhanced N1 and even P1 component for attended compared with ignored tones; for a review, see Giard et al., 2000), even if it occurred presumably outside of the focus of voluntary attention in the present experiments. This pattern is in line with the assumption – based on visual studies – that involuntary attention to affective information can produce similar enhancement of perceptual processing as further, “cold” forms of attention; however, in the case of affectively significant stimuli, this amplification is presumably mediated by distinct “emotional attention” than a gain control exerted by attention networks under voluntary control, implying that these different forms of attention can produce relatively independent facilitation on perceptual processes (for reviews, see Pourtois et al., 2013; Vuilleumier, 2005; Vuilleumier & Driver, 2007; Vuilleumier & Huang, 2009). However, while the here applied ERP-method gives precise information about the stage of auditory processing at which preferential processing of valenced tones occur, the exact underlying mechanism of this facilitation could not be answered directly in the present experiments. By employing different approaches, gain control by amygdala signals that can bias ongoing perceptual analysis in sensory areas (e.g., Pourtois et al., 2013), and short-term plasticity in the sensory cortices enabling selective tuning to specific features of the affectively significant stimuli (e.g., Bröckelmann et al., 2011, 2013) have been proposed as underlying mechanisms for rapid prioritization of affective stimuli.

In sum, “natural” selective attention can amplify initial perceptual representation of affectively significant tones in a relatively automatic way. This outcome also resembles affective attentional effects on early visual processing (for reviews, see Pourtois et al., 2013; Vuilleumier & Driver, 2007; Vuilleumier & Huang, 2009), suggesting that similar or common mechanism might be involved across sensory domains.

PREFERENTIAL SELECTION OF VALENCED TONES FROM TEMPORAL SEQUENCES?

Given the basic importance of the temporal aspect in the auditory modality, in the third research route, we investigated whether valenced tones can be selected preferentially from rapid sound sequences under the circumstances when temporal abilities of the auditory system are limited. Visual studies suggest a functional link between early attentional enhancement for affectively significant pictures and a benefit in target detection performance in visual space (e.g., Pourtois et al., 2013; Vuilleumier & Driver, 2007). Thus, in line with the early attentional enhancement found in Experiments 3-5, we expected that a target detection benefit can arise in the temporal dimension of auditory attention for valenced tones.

The attentional blink, thus a performance deficit during multiple target detection in rapid stimulus sequences has been extensively investigated in the visual modality. Of most relevance, affectively significant visual targets are relatively preserved from this constraint of visual attention, indicating their special attentional status. We examined whether early preferential attention for affectively significant tones can provide a detection benefit for these tones, leading to relatively preserved performance during dual-target detection in rapid sound sequences.

As the auditory dual-target deficit is much less understood compared with the visual attentional blink, first we investigated the basic phenomenon in Experiment 6.

We found a short-lived deficit during multiple target detection that was characterized by mutual degradation of the targets. This pattern cannot be considered as a direct equivalent of a visual attentional blink, but more compatible with a short-range mutual competition for perceptual awareness between potentially relevant auditory stimuli. This deficit that is likely, at least partly, of perceptual origin is referred as informational masking. In Experiment 7, we explored the possible affective modulation of the dual-target deficit. We found a weak indication of a bias toward valenced tones as a preferential weighting of valenced second targets over the first target in their short-range competition. Notably, this weak indication occurred distinctively from the affective attentional blink-sparing that was demonstrated in the visual modality, supporting our suggestion about the distinctiveness of the auditory dual-target deficit from the visual attentional blink.

Taken together, in Experiment 7 we found a weak (i.e., marginally significant) evidence for prioritization of valenced tones during target detection. However, while the approach of inducing valence experimentally possesses the incontrovertible advantage of strict control over arbitrary physical differences and the time of availability of the valance-related information, we have to acknowledge that the newly acquired valence of simple tones is expectedly and reasonably less intense than a priori valence of natural emotional sounds (e.g., biologically relevant stimuli, such as erotic sounds, screaming, growling). Thus, it is possible that the relatively mild affective valence lead to facilitation of sensory encoding of positive and negative tones as reflected in the results of Experiments 3-5, but a rather slight facilitation did not allow for pronounced detection benefit in Experiment 7.

General Relevance Principle in Attentional Biases to Affectively Significant Tones

In Experiments 3-5 and Experiment 7, we targeted the question whether rapid attentional biases to valenced auditory stimuli are driven by their motivational relevance in general (thus both positive and negative tones are prioritized given their general relevance regardless of their valence category; in the visual modality, see, e.g., Brosch et al., 2008; Rothermund et al., 2008; Wentura et al., 2014); or, based on the assumption that detection of dangers possesses arguably higher survival value compared with detection of opportunities, our auditory attention is tuned to selectively prioritize negative information (in the visual modality, see, e.g., Öhman et al., 2001; Öhman, 2005).

The pattern of results concerning affective attentional biases at the level of sound encoding appeared unequivocal: In Experiments 3-5, we found only indications for valenced versus neutral differential enhancement at an early stage of sound encoding, but we found no indication for prioritization of a specific valence category over another. This result is in accordance with a general relevance principle, that is, it is the relevance of the affective stimulus concerning our goals, needs, and well-being that governs early attentional processes rather than a specific valence category. Although we found only a weak indication for preferential selection of valenced targets in Experiment 7, this indication also occurred for valenced tones in general relative to neutral tones.

Collectively, our results are in accordance with findings from the visual domain (e.g., Brosch et al., 2008; Rothermund et al., 2008; Wentura et al., 2014), supporting a general relevance principle in basic affective attentional biases. Moreover, comparable preferential attention across valence categories has been also demonstrated in visual studies employing conditioned valence based on rewards and losses (see Müller et al., 2015; Wentura et al., 2014). Note that this approach possesses the advantage of “objectively” balanced relevance between positive and negative valance category (i.e., positive and negative valence is induced by the win and loss of the same value). However, studies demonstrating negativity bias by employing natural emotional stimuli often simply compare negative stimuli with neutral ones, or when using both valence categories, negative and positive stimuli are often not equated properly in their motivational relevance (e.g., an angry face directed toward the observer may signal severe consequences and prompt for immediate reaction, while a happy face has weaker relevance for the observer, see the argumentation of Brosch et al., 2008).

Experiment 5 targeted a third (not exclusive) possibility, that is that affective biases in early auditory attention can promote asymmetric facilitation between positive and negative valence in different motivational-affective contexts. However, in Experiment 5 we did not find indication of a counter-regulation principle on the attentional biases to valent information. That is, while behavioral measures of the learning phase indicated a differentiation between positive and negative valence, there was no such differentiation on the N1 amplitude. The pattern of results is consistent with a rather fixed general relevance bias hypothesis. However, based on the absence of moderation of the early preferential valence effect by motivational outcome focus in Experiment 5, it is still unclear whether there are such more extreme affective-motivational contexts and personal factors that can promote differential attentional enhancement for specific valence categories during sound encoding.

Limitations

As a possible criticism, one can argue that the statistical power of some of our single experiments necessitates a rather cautious interpretation of the results. Experiments

3-5 had a relatively modest sample size of N = 24 each and showed the critical valenced vs. neutral difference on the N1 amplitudes (i.e., the most relevant comparison concerning our research question) with effect sizes in the lower middle range ($|d_z|$s were in the range of 0.30-0.51 for the critical valenced vs. neutral difference).

First, we have to acknowledge that when determining sample sizes of the single experiments, besides attaining high test power, we also considered keeping the relatively costly EEG data collection in a reasonable range. Second, as an answer to the possible criticism, a combined analysis of Experiments 3-5 proved that the three experiments reflect comparable valence effect across the pooled sample of 72 participants: (1) A clear early attentional effect for valenced compared with neutral tones as reflected in enhanced N1 amplitudes for positive and negative tones relative to neutral ones; (2) with no indication of difference between positive and negative valences. This pattern of findings was consistent across the three experiments as no interaction between the factor of valence and experiment occurred (F < 0.16, *n.s.*). Across the three experiments, the most relevant valence-minus-neutral difference on the N1 amplitude was associated with an effect size of $|d_z|$ = 0.38 and achieved a statistical power of .89. Furthermore, the comparison of the positive vs. negative valence conditions was associated with the very small effect size of $|d_z|$ = 0.03, thus, it is highly unlikely that this negative finding occurred due to the limited sample size.

Taken together, although the same pattern of valence results emerged across the three experiments, we have to acknowledge that the sample sizes of the single Experiments 3-5 were rather modest. Thus, limited statistical power could have played a role in limiting significance of some of the statistical analyses, and therefore, we recommend using higher sample sizes for future investigations.

Similarly, a critical point concerns the power of Experiment 7, which were based on a relatively modest sample of 29 participants. This sample size was suited for detecting medium sized effects with a probability of 1 - β = .80 for the most relevant comparisons concerning our research questions in this rather exploratory work. However, the most important valenced vs. neutral difference on the T2-related interference effect on T1 was rather small, and it just missed the conventional level of statistical significance. Consequently, limited statistical power could have played a role in limiting significance in our analyses. Further studies with higher sample sizes could clarify this issue exhaustively.

Furthermore, as a possible criticism, preparedness for prima facie automatic affective evaluation might depend to some degree on attentional biases introduced by the employed task. This possible limitation concerns primarily the auditory affective Simon paradigm, as we cannot preclude the possibility that attention allocation on evaluative stimulus features could contribute to the evaluation of emotional sounds. Numerous evidences indicate that processing of a task-irrelevant stimulus can be modulated by selective attention allocation to specific stimulus features or dimensions promoted by the task (in the visual modality, see Folk, Remington, & Johnston, 1992; Kiefer & Martens, 2010; Spruyt, De Houwer, Everaert, & Hermans, 2012; Spruyt, De Houwer, Hermans, & Eelen, 2007; Spruyt, De Houwer, & Hermans, 2009; Spruyt, Klauer, Gast, Schryver, & De Houwer, 2014; Everaert, Spruyt, Rossi, Pourtois, & De Houwer, 2014; see also Bermeitinger, Wentura, & Frings, 2011). Moreover, feature-specific attention allocation can be an important factor in the evaluative domain: Several evidences suggest that prima facie automatic evaluative effects can be reduced when participants assign selective attention to a non-affective rather than an affective stimulus dimension (in the visual modality, see Everaert, Spruyt, & De Houwer, 2013; Everaert et al., 2014; Spruyt et al., 2012, 2007, 2009). As in Experiment 2 strongly positively or negatively connoted verbal responses had to be uttered throughout the experiment, it is possible that the responses promoted selective attention to the affective dimension. Further research can elucidate whether feature-specific attention allocation contributes to the auditory affective Simon effect.

For Experiments 3-4 and 7, we can rule out the possibility that attention allocation to the evaluative dimension was introduced by the concomitant task, as the tasks applied in the test phase (i.e., selective listening; RASP) was purely perceptual without any reference to affective features. In Experiment 5, although the task was again purely perceptual, the motivational context of anticipating positive or negative future outcomes could promote an "evaluative set" that in turn could support selective attention to the affective dimension in general. Nonetheless, the actual pattern of results (i.e., comparable valence effects across Experiments 3-5) appears to be contradictory with this concern.

Open Questions and Future Directions

Open questions and future directions derived from the present work can be delineated at two levels: First, the more specific open questions were already discussed concerning our three research routes. Second, much more remains to be investigated at an overarching level. In the present and rather exploratory work, we focused primarily on one of the most conspicuous characteristics of auditory modality, that is, the importance of the temporal dimension, given the transient nature of the auditory world and complex and extended temporal structure of sounds. However, much remains to be understood regarding other important characteristics of the auditory domain which can be potentially relevant concerning automatic affective processing. For example, the omnidirectionality and transparency of hearing (i.e., it covers 360-degree in space and sounds are perceived even if the source is obscured), its "obligatory" nature, intrusiveness, and urging behavioral relevance (e.g., in terminating ongoing actions) can be potentially important for affective research.

In the presented experiments, we found evidence for remarkably fast intentional and unintentional affective sound evaluation and rapid, involuntary preferential attention for valenced auditory stimuli. These results delineate notably fast and efficient affective processing – rather similar to that found in visual modality despite the obvious differences between the two domains. An important question for future research is that to which extent the similar phenomena observed in visual and auditory domains are brought by general mechanisms that are present across stimulus modalities during affective processing.

Furthermore, we emphasized primarily certain features of automaticity (e.g., fast and unintentional evaluations and involuntary attentional effects) from a decompositional view, while further aspects have not been emphasized particularly due to the limited scope of the present thesis. However, a more comprehensive investigation of the interplay between different automaticity features would be beneficial in the future. In addition, as in Experiments 3-5 and Experiment 7 we employed simple stimuli endowed with affective valence during a "value-based" valence induction, much also remains to be explore concerning the stimulus specificity of the presented effects, and whether and how further dimensions (e.g., biological relevance) can shape preferential attentional effects in the auditory domain. Likewise, as the present work is rather exploratory, we did not target possible moderation of the presented effects by more enduring affective states (e.g., positive or negative mood) and personal factors (e.g., trait anxiety). Future research can elucidate the possible role of longer-term affective states and personal differences.

Finally, and importantly, we live and act in a multisensory environment. While the present thesis targeted the relatively neglected field of affective processing in the auditory domain compared to our vast knowledge from visual modality, our work still considered unimodal situations. Future research should put more emphasize on cross-modal interactions in affective evaluations and affective attention.

Conclusions

In the present work, we investigated affective evaluation of auditory stimuli and its interplay with auditory attention concerning different aspects of automaticity (like fast and unintentional affective processes), while highlighting the basic importance of the temporal dimension in the auditory domain. Our first research route provided evidence that affective valence of complex natural sounds can be extracted automatically in the sense of rapid and implicit evaluations. The second research route demonstrated that affectively significant tones can attract preferential attention at an early, perceptual level of sound encoding (i.e. within about 100 ms after sound onset), even outside of the focus of voluntary attention. In the third research route, we found a weak indication for preferential selection of affectively significant tones under the conditions in which temporal abilities of the auditory system are limited; however, this weak benefit appeared in a distinctive manner as in the visual modality. We suggest that this result can reflect, at least partly, the superior temporal resolution of the auditory system compared with vision. Our results outline impressively rapid auditory affective processing that can analyze potentially significant information in our acoustic environment automatically, not only in the sense as fast but also without intention, voluntary attention, and relatively independently of concurrent tasks.

Auditory perception, attention, and, as the present work highlighted, affective processing are strongly connected. On the one hand, affective information can bias perception and attention. In the second research route, we found enhanced early auditory processing that can be considered as the hallmark of “natural” affective attention. In the third research route, at the behavioral level, we found a weak evidence for preferential selection of valenced tones under the conditions when temporal abilities of the auditory system are limited. On the other hand, the observed affective evaluations and preferential attention appeared relatively unaffected by the (un)intentionality of the process and concurrent attentional and motivational demands, delineating rather “reflexive” affective processes in the auditory domain. However, as a consequence of its exploratory nature, the present work opened numerous new questions. We hope that investigation of auditory affective processing with a consideration of modality-specific factors attracts more attention in the future.

References

Akyürek, E. G., H, A., Nieuwenstein, M. R., Saija, J. D., Başkent, D., & Hommel, B. (2012). Temporal target integration underlies performance at lag 1 in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1448–1464. doi:10.1037/a0027610

Alho, K., Töttölä, K., Reinikainen, K., Sams, M., & Näätänen, R. (1987). Brain mechanism of selective listening reflected by event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(6), 458–470. doi:10.1016/0168-5597(87)90057-8

Alvarez, W., Fuente, A., Coloma, C. J., & Quezada, C. (2015). Association between temporal resolution and Specific Language Impairment: The role of nonsensory processing. *International Journal of Pediatric Otorhinolaryngology*, *79*(10), 1702–1707. doi:10.1016/j.ijporl.2015.07.029

Anderson, A. K. (2005). Affective Influences on the Attentional Dynamics Supporting Awareness. *Journal of Experimental Psychology: General*, *134*(2), 258–281. doi:10.1037/0096-3445.134.2.258

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*(6835), 305–309. doi:10.1038/35077083

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011a). Learned Value Magnifies Salience-Based Attentional Capture. *PLoS ONE*, *6*(11), e27926. doi:10.1371/journal.pone.0027926

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011b). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367–10371. doi:10.1073/pnas.1104047108

Anderson, B. A., & Yantis, S. (2013). Persistence of value-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 6–9. doi:10.1037/a0030860

Armony, J. L., Aubé, W., Angulo-Perkins, A., Peretz, I., & Concha, L. (2015). The specificity of neural responses to music and their relation to voice processing: An fMRI-adaptation study. *Neuroscience Letters*, *593*, 35–39. doi:10.1016/j.neulet.2015.03.011

Armony, J. L., & Dolan, R. J. (2002). Modulation of spatial attention by fear-conditioned stimuli: an event-related fMRI study. *Neuropsychologia*, *40*(7), 817–826. doi:10.1016/S0028-3932(01)00178-6

Arnell, K. M., & Jenkins, R. (2004). Revisiting within-modality and cross-modality attentional blinks: Effects of target—distractor similarity. *Perception & Psychophysics*, *66*(7), 1147–1161. doi:10.3758/BF03196842

Arnell, K. M., & Jolicœur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 630–648. doi:10.1037/0096-1523.25.3.630

Arnell, K. M., Killman, K. V., & Fijavz, D. (2007). Blinded by emotion: Target misses follow attention capture by arousing distractors in RSVP. *Emotion*, *7*(3), 465–477. doi:10.1037/1528-3542.7.3.465

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Bargh, J. A. (1992). The Ecology of Automaticity: Toward Establishing the Conditions Needed to Produce Automatic Processing Effects. *The American Journal of Psychology*, *105*(2), 181–199. doi:10.2307/1423027

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & Srull (Ed.), *Handbook of social cognition* (pp. 1–40). Hillsdale, NJ: Erlbaum.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. doi:10.1037/0033-2909.133.1.1

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality*, *40*(1), 35–55. doi:10.1016/j.jrp.2005.08.006

Barry, R. J., Cocker, K. I., Anderson, J. W., Gordon, E., & Rennie, C. (1992). Does the N100 evoked potential really habituate? Evidence from a paradigm appropriate to a clinical setting. *International Journal of Psychophysiology*, *13*(1), 9–16. doi:10.1016/0167-8760(92)90014-3

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Lme4: Linear Mixed-Effects Models using “Eigen” and S4. R package version 1.1-7. Retrieved from http://CRAN.R-project.org/package=lme4

Batty, M. J., Cave, K. R., & Pauli, P. (2005). Abstract Stimuli Associated With Threat Through Conditioning Cannot Be Detected Preattentively. *Emotion*, *5*(4), 418–430. doi:10.1037/1528-3542.5.4.418

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*(2), 120–131. doi:10.1016/j.ijpsycho.2011.08.003

Bermeitinger, C., Wentura, D., & Frings, C. (2011). How to switch on and switch off semantic priming effects for natural and artifactual categories: activation processes in category memory depend on focusing specific feature dimensions. *Psychonomic Bulletin & Review*, *18*(3), 579–585. doi:10.3758/s13423-011-0067-z

Bertels, J., Kolinsky, R., & Morais, J. (2012). Lack of habituation to shocking words: The attentional bias to their spatial origin is context free. *Cognition and Emotion*, *26*(8), 1345–1358. doi:10.1080/02699931.2012.656582

Bestelmeyer, P. E. G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to Vocal Expressions Reveals Multistep Perception of Auditory Emotion. *The Journal of Neuroscience*, *34*(24), 8098–8105. doi:10.1523/JNEUROSCI.4820-13.2014

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M. H., & Bertrand, O. (2007). Effects of Selective Attention on the Electrophysiological Representation of Concurrent Sounds in the Human Auditory Cortex. *The Journal of Neuroscience*, *27*(35), 9252–9261. doi:10.1523/JNEUROSCI.1402-07.2007

Bishop, S. J., Jenkins, R., & Lawrence, A. D. (2007). Neural Processing of Fearful Faces: Effects of Anxiety are Gated by Perceptual Capacity Limitations. *Cerebral Cortex*, *17*(7), 1595–1603. doi:10.1093/cercor/bhl070

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395. doi:10.1038/nn1409

Bradley, M. M., Hamby, S., Löw, A., & Lang, P. J. (2007). Brain potentials in perception: Picture complexity and emotional arousal. *Psychophysiology*, *44*(3), 364–373. doi:10.1111/j.1469-8986.2007.00520.x

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, *37*(02), 204–215. doi:null

Bradley, M. M., & Lang, P. J. (2007). The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. *University of Florida, Gainesville, FL, Tech. Rep. B-3*.

Bregman, A. S. (1990). *Auditory scene analysis. The perceptual organization of sound*. Cambridge, MA: The MIT Press.

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *44*(3), 400–413. doi:10.1037/h0084255

Broadbent, D. E. (1958). *Perception and Communication*. Oxford: Oxford University Press.

Bröckelmann, A.-K., Steinberg, C., Dobel, C., Elling, L., Zwanzger, P., Pantev, C., & Junghöfer, M. (2013). Affect-specific modulation of the N1m to shock-conditioned tones: magnetoencephalographic correlates. *European Journal of Neuroscience*, *37*(2), 303–315. doi:10.1111/ejn.12043

Bröckelmann, A.-K., Steinberg, C., Elling, L., Zwanzger, P., Pantev, C., & Junghöfer, M. (2011). Emotion-Associated Tones Attract Enhanced Attention at Early Auditory Processing: Magnetoencephalographic Correlates. *The Journal of Neuroscience*, *31*(21), 7801–7810. doi:10.1523/JNEUROSCI.6236-10.2011

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2009). Cross-modal Emotional Attention: Emotional Voices Modulate Early Stages of Visual Processing. *Journal of Cognitive Neuroscience*, *21*(9), 1670–1679. doi:10.1162/jocn.2009.21110

Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond Fear Rapid Spatial Orienting Toward Positive Emotional Stimuli. *Psychological Science*, *19*(4), 362–370. doi:10.1111/j.1467-9280.2008.02094.x

Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye... Attention capture by infant faces. *Emotion*, *7*(3), 685–689. doi:10.1037/1528-3542.7.3.685

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527–2538. doi:10.1121/1.1408946

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *International Journal of Psychophysiology*, *31*(1), 51–68. doi:10.1016/S0167-8760(98)00040-3

Burkard, R. F., Eggermont, J. J., & Don, M. (2007). *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Lippincott Williams & Wilkins.

Carlson, J. M., Beacher, F., Reinke, K. S., Habib, R., Harmon-Jones, E., Mujica-Parodi, L. R., & Hajcak, G. (2012). Nonconscious attention bias to threat is correlated with anterior cingulate cortex gray matter volume: A voxel-based morphometry result and replication. *NeuroImage*, *59*(2), 1713–1718. doi:10.1016/j.neuroimage.2011.09.040

Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J. A. (2001). Emotion, attention, and the "negativity bias", studied through event-related potentials. *International Journal of Psychophysiology*, *41*(1), 75–85. doi:10.1016/S0167-8760(00)00195-1

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(7), 1173–1197. doi:10.1080/02724980443000539

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109–127. doi:10.1037/0096-1523.21.1.109

Chun, M. M., & Potter, M. C. (2001). The attentional blink and task switching within and across modalities. In K. L. Shapiro (Ed.), *The limits of attention: Temporal constraints in human information processing* (pp. 20–35). Oxford: Oxford University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331–335. doi:10.3758/BF03196169

Craft, J. L., & Simon, J. R. (1970). Processing symbolic information from a visual display: Interference from an irrelevant directional cue. *Journal of Experimental Psychology*, *83*(3, Pt.1), 415–420. doi:10.1037/h0028843

Crottaz-Herbette, S., & Ragot, R. (2000). Perception of complex sounds: N1 latency codes pitch and topography codes spectra. *Clinical Neurophysiology*, *111*(10), 1759–1766. doi:10.1016/S1388-2457(00)00422-3

Czigler, I., Cox, T. J., Gyimesi, K., & Horváth, J. (2007). Event-related potential study to aversive auditory stimuli. *Neuroscience Letters*, *420*(3), 251–256. doi:10.1016/j.neulet.2007.05.007

Degner, J. (2011). Affective priming with auditory speech stimuli. *Language and Cognitive Processes*, *26*(10), 1710–1735. doi:10.1080/01690965.2010.532625

De Houwer, J., & Eelen, P. (1998). An Affective Variant of the Simon Paradigm. *Cognition & Emotion*, *12*(1), 45–62. doi:10.1080/026999398379772

Dell'Acqua, R., Dux, P. E., Wyble, B., Doro, M., Sessa, P., Meconi, F., & Jolicœur, P. (2015). The Attentional Blink Impairs Detection and Delays Encoding of Visual Information: Evidence from Human Electrophysiology. *Journal of Cognitive Neuroscience*, *27*(*4*), 720–735. doi:10.1162/jocn_a_00752

Dimitrijevic, A., Michalewski, H. J., Zeng, F.-G., Pratt, H., & Starr, A. (2008). Frequency changes in a continuous tone: Auditory cortical potentials. *Clinical Neurophysiology*, *119*(9), 2111–2124. doi:10.1016/j.clinph.2008.06.002

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The Automatic Evaluation of Novel Stimuli. *Psychological Science*, *13*(6), 513–519. doi:10.1111/1467-9280.00490

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*, 808–810. doi:10.1038/42947

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics*, *71*(8), 1683–1700. doi:10.3758/APP.71.8.1683

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*(3), 183–201. doi:10.1037/h0047707

Eddins, D. A., & Green, D. M. (1995). Temporal integration and temporal resolution. In B. C. J. Moore (Ed.), *Hearing* (pp. 206–242). New York: Academic Press.

Eggermont, J. J. (2015). *Auditory Temporal Processing and Its Disorders*. Oxford University Press.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, *45*(1), 15–31. doi:10.1016/j.neuropsychologia.2006.04.022

Eimer, M., Holmes, A., & McGlone, F. P. (2003). The role of spatial attention in the processing of facial expression: An ERP study of rapid brain responses to six basic emotions. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(2), 97–110. doi:10.3758/CABN.3.2.97

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, H. H. Goldsmith, & K. R. Scherer, *Handbook of affective sciences* (pp. 572–595). Oxford: Oxford University Press.

Engen, H. G., Smallwood, J., & Singer, T. (2015). Differential impact of emotional task relevance on three indices of prioritised processing for fearful and angry facial expressions. *Cognition and Emotion*, *0*(0), 1–10. doi:10.1080/02699931.2015.1081873

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. doi:10.1162/089892998562997

Esteves, F., Dimberg, U., & Öhman, A. (1994). Automatically elicited fear: Conditioned skin conductance responses to masked facial expressions. *Cognition & Emotion*, *8*(5), 393–413. doi:10.1080/02699939408408949

Everaert, T., Spruyt, A., & De Houwer, J. (2013). On the malleability of automatic attentional biases: Effects of feature-specific attention allocation. *Cognition and Emotion*, *27*(3), 385–400. doi:10.1080/02699931.2012.712949

Everaert, T., Spruyt, A., Rossi, V., Pourtois, G., & De Houwer, J. (2014). Feature-specific attention allocation overrules the orienting response to emotional stimuli. *Social Cognitive and Affective Neuroscience*, *9*(9), 1351–1359. doi:10.1093/scan/nst121

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2). 175–191.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *NeuroImage*, *36*(2), 480–487. doi:10.1016/j.neuroimage.2007.02.043

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *NeuroImage*, *30*(4), 1441–1448. doi:10.1016/j.neuroimage.2005.11.003

Fitzgibbons, P. J., & Gordon-Salant, S. (1998). Auditory Temporal Order Perception in Younger and Older Adults. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1052–1060. doi:10.1044/jslhr.4105.1052

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. doi:10.1037/0096-1523.18.4.1030

Folyi, T., Fehér, B., & Horváth, J. (2012). Stimulus-focused attention speeds up auditory processing. *International Journal of Psychophysiology*, *84*(2), 155–163. doi:10.1016/j.ijpsycho.2012.02.001

Folyi, T., Liesefeld, H. R., & Wentura, D. (2016). Attentional enhancement for positive and negative tones at an early stage of auditory processing. *Biological Psychology*, *114*, 23-32. doi:10.1016/j.biopsycho.2015.12.001

Folyi, T., & Wentura, D. (2015). Fast and unintentional evaluation of emotional sounds: evidence from brief segment ratings and the affective Simon task. *Cognition and Emotion*, Advance online publication. doi:10.1080/02699931.2015.1110514

Forster, S., & Lavie, N. (2007). High Perceptual Load Makes Everybody Equal Eliminating Individual Differences in Distractibility With Load. *Psychological Science*, *18*(5), 377–381. doi:10.1111/j.1467-9280.2007.01908.x

Forster, S., & Lavie, N. (2008). Failures to ignore entirely irrelevant distractors: The role of load. *Journal of Experimental Psychology: Applied*, *14*(1), 73–83. doi:10.1037/1076-898X.14.1.73

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention — focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*(4), 437–455. doi:10.1016/j.conb.2007.07.011

Frühholz, S., & Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, *37*(1), 24–35. doi:10.1016/j.neubiorev.2012.11.002

Frühholz, S., Hofstetter, C., Cristinzio, C., Saj, A., Seeck, M., Vuilleumier, P., & Grandjean, D. (2015). Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proceedings of the National Academy of Sciences*, *112*(5), 1583–1588. doi:10.1073/pnas.1411315112

Furl, N., Henson, R. N., Friston, K. J., & Calder, A. J. (2013). Top-Down Control of Visual Responses to Fear by the Amygdala. *The Journal of Neuroscience*, *33*(44), 17435–17443. doi:10.1523/JNEUROSCI.2992-13.2013

Giard, M. H., Collet, L., Bouchet, P., & Pernier, J. (1994). Auditory selective attention in the human cochlea. *Brain Research*, *633*(1), 353–356. doi:10.1016/0006-8993(94)91561-X

Giard, M. H., Fort, A., Mouchetant-Rostaing, Y., & Pernier, J. (2000). Neurophysiological mechanisms of auditory selective attention in humans. *Frontiers in Bioscience*, *5*, D84–D94.

Giard, M. H., Perrin, F., Echallier, J. F., Thévenet, M., Froment, J. C., & Pernier, J. (1994). Dissociation of temporal and frontal components in the human auditory N1 wave: a scalp current density and dipole model analysis. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(3), 238–252. doi:10.1016/0168-5597(94)90067-1

Gilmore, C. S., Clementz, B. A., & Berg, P. (2009). Hemispheric differences in auditory oddball responses during monaural versus binaural stimulation. *International Journal of Psychophysiology*, *73*(3), 326–333. doi:10.1016/j.ijpsycho.2009.05.005

Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2012). The Nature of Affective Priming in Music and Speech. *Journal of Cognitive Neuroscience*, *24*(8), 1725–1741. doi:10.1162/jocn_a_00213

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146. doi:10.1038/nn1392

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(11), 887–892. doi:10.1038/nrn1538

Grube, M., Cooper, F. E., Kumar, S., Kelly, T., & Griffiths, T. D. (2014). Exploring the role of auditory analysis in atypical compared to typical language development. *Hearing Research*, *308*, 129–140. doi:10.1016/j.heares.2013.09.015

Grube, M., Kumar, S., Cooper, F. E., Turton, S., & Griffiths, T. D. (2012). Auditory sequence analysis and phonological skill. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1746), 4496–4504. doi:10.1098/rspb.2012.1817

Gupta, R., Hur, Y.-J., & Lavie, N. (2015). Distracted by pleasure: Effects of positive versus negative valence on emotional capture under load. *APA PsycNET*. doi:10.1037/emo0000112

Gutschalk, A., Micheyl, C., & Oxenham, A. J. (2008). Neural Correlates of Auditory Perceptual Awareness under Informational Masking. *PLoS Biol*, *6*(6), e138. doi:10.1371/journal.pbio.0060138

Hahn, S., & Gronlund, S. D. (2007). Top-down guidance in visual search for facial expressions. *Psychonomic Bulletin & Review*, *14*(1), 159–165. doi:10.3758/BF03194044

Hajcak, G., & Olvet, D. M. (2008). The persistence of attention to emotion: Brain potentials during and after picture presentation. *Emotion*, *8*(2), 250–255. doi:10.1037/1528-3542.8.2.250

Handy, T. C. (2005). *Event-related Potentials: A Methods Handbook*. Cambridge, MA: MIT Press.

Handy, T. C., Soltani, M., & Mangun, G. R. (2001). Perceptual Load and Visuocortical Processing: Event-Related Potentials Reveal Sensory-Level Selection. *Psychological Science*, *12*(3), 213–218. doi:10.1111/1467-9280.00338

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, *54*(6), 917–924. doi:10.1037/0022-3514.54.6.917

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*(3), 356. doi:10.1037/0096-3445.108.3.356

Hawk, S. T., Fischer, A. H., & Van Kleef, G. A. (2012). Face the noise: Embodied responses to nonverbal vocalizations of discrete emotions. *Journal of Personality and Social Psychology*, *102*(4), 796–814. doi:10.1037/a0026234

Hein, G., Alink, A., Kleinschmidt, A., & Müller, N. G. (2009). The Attentional Blink Modulates Activity in the Early Visual Cortex. *Journal of Cognitive Neuroscience*, *21*(1), 197–206. doi:10.1162/jocn.2008.21026

Hein, G., Parr, A., & Duncan, J. (2006). Within-modality and cross-modality attentional blinks in a simple discrimination task. *Perception & Psychophysics*, *68*(1), 54–61. doi:10.3758/BF03193655

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion*, *15*(2), 143–165. doi:10.1080/02699930125768

Herrmann, C. S., & Knight, R. T. (2001). Mechanisms of human attention: event-related potentials and oscillations. *Neuroscience & Biobehavioral Reviews*, *25*(6), 465–476. doi:10.1016/S0149-7634(01)00027-6

Hilgard, J., Weinberg, A., Hajcak Proudfit, G., & Bartholow, B. D. (2014). The negativity bias in affective picture processing depends on top-down and bottom-up motivational significance. *Emotion*, *14*(5), 940–949. doi:10.1037/a0036791

Hilkenmeier, F., Olivers, C. N. L., & Scharlau, I. (2012). Prior entry and temporal attention: Cueing affects order errors in RSVP. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 180–190. doi:10.1037/a0025978

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science*, *182*(4108), 177–180. doi:10.1126/science.182.4108.177

Hillyard, S. A., Mangun, G. R., Woldorff, M. G., & Luck, S. J. (1995). Neural systems mediating selective attention. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 665–681). Cambridge, MA, US: The MIT Press.

Hillyard, S. A., Teder-Sälejärvi, W. A., & Münte, T. F. (1998). Temporal dynamics of early perceptual processing. *Current Opinion in Neurobiology*, *8*(2), 202–210. doi:10.1016/S0959-4388(98)80141-4

Hintze, P., Junghöfer, M., & Bruchmann, M. (2014). Evidence for rapid prefrontal emotional evaluation from visual evoked responses to conditioned gratings. *Biological Psychology*, *99*, 125–136. doi:10.1016/j.biopsycho.2014.03.010

Hodsoll, S., Viding, E., & Lavie, N. (2011). Attentional capture by irrelevant emotional distractor faces. *Emotion*, *11*(2), 346–353. doi:10.1037/a0022771

Holmes, A., Vuilleumier, P., & Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: evidence from event-related brain potentials. *Cognitive Brain Research*, *16*(2), 174–184. doi:10.1016/S0926-6410(02)00268-9

Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *The Quarterly Journal of Experimental Psychology Section A*, *58*(8), 1415–1433. doi: 10.1080/02724980443000647

Horváth, J., & Burgyán, A. (2011). Distraction and the auditory attentional blink. *Attention, Perception, & Psychophysics*, *73*(3), 695–701. doi:10.3758/s13414-010-0077-3

Horváth, J., Maess, B., Baess, P., & Tóth, A. (2012). Action–Sound Coincidences Suppress Evoked Responses of the Human Auditory Cortex in EEG and MEG. *Journal of Cognitive Neuroscience*, *24*(9), 1919–1931. doi:10.1162/jocn_a_00215

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, *79*(2), 139–147. doi:10.1016/j.biopsycho.2008.04.001

Huang, Y.-M., Baddeley, A., & Young, A. W. (2008). Attentional capture by emotional stimuli is modulated by semantic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 328–339. doi:10.1037/0096-1523.34.2.328

Hughes, R. W., & Jones, D. M. (2001). The intrusiveness of sound : Laboratory findings and their implications for noise abatement. *Noise and Health*, *4*(13), 51.

Ihssen, N., & Keil, A. (2009). The costs and benefits of processing emotional stimuli during rapid serial visual presentation. *Cognition and Emotion*, *23*(2), 296–326. doi:10.1080/02699930801987504

Isaak, M. I., Shapiro, K. L., & Martin, J. (1999). The attentional blink reflects retrieval competition among multiple rapid serial visual presentation items: Tests of an interference model. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1774–1792. doi:10.1037/0096-1523.25.6.1774

Ito, T. A., Larsen, J. T., Kyle, N., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887–900. doi:10.1037/0022-3514.75.4.887

Jääskeläinen, I. P., & Ahveninen, J. (2014). Auditory-Cortex Short-Term Plasticity Induced by Selective Attention. *Neural Plasticity*, *2014*, e216731. doi:10.1155/2014/216731

Jacobsen, T., Schröger, E., Horenkamp, T., & Winkler, I. (2003). Mismatch negativity to pitch change: varied stimulus proportions in controlling effects of neural refractoriness on human auditory event-related brain potentials. *Neuroscience Letters*, *344*(2), 79–82. doi:10.1016/S0304-3940(03)00408-7

Jacobson, G. P., Lombardi, D. M., Gibbens, N. D., Ahmad, B. K., & Newman, C. W. (1992). The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. *Ear and Hearing*, *13*, 300–306.

Jacoby, O., Visser, T. A. W., Hart, B. C., Cunnington, R., & Mattingley, J. B. (2011). No Evidence for Early Modulation of Evoked Responses in Primary Visual Cortex to Irrelevant Probe Stimuli Presented during the Attentional Blink. *PLoS ONE*, *6*(8), e24255. doi:10.1371/journal.pone.0024255

Jansma, H., Roebroeck, A., & Münte, T. F. (2014). A network analysis of audiovisual affective speech perception. *Neuroscience*, *256*, 230–241. http://doi.org/10.1016/j.neuroscience.2013.10.047

Jolicœur, P., & Dell'Acqua, R. (1998). The Demonstration of Short-Term Consolidation. *Cognitive Psychology*, *36*(2), 138–202. doi:10.1006/cogp.1998.0684

Jolicœur, P., & Dell'Acqua, R. (1999). Attentional and structural constraints on visual encoding. *Psychological Research*, *62*(2-3), 154–164. doi:10.1007/s004260050048

Junghöfer, M., Sabatinelli, D., Bradley, M. M., Schupp, H. T., Elbert, T. R., & Lang, P. J. (2006). Fleeting images: Rapid affect discrimination in the visual cortex. *NeuroReport*, *17*, 225–229.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(05), 559–575. doi:10.1017/S0140525X08005293

Kauramäki, J., Jääskeläinen, I. P., Hänninen, J. L., Auranen, T., Nummenmaa, A., Lampinen, J., & Sams, M. (2012). Two-Stage Processing of Sounds Explains Behavioral Performance Variations due to Changes in Stimulus Contrast and Selective Attention: An MEG Study. *PLoS ONE*, *7*(10), e46872. doi:10.1371/journal.pone.0046872

Keil, A., Gruber, T., Müller, M. M., Moratti, S., Stolarova, M., Bradley, M. M., & Lang, P. J. (2003). Early modulation of visual perception by emotional arousal: Evidence from steady-state visual evoked brain potentials. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(3), 195–206. doi:10.3758/CABN.3.3.195

Keil, A., & Ihssen, N. (2004). Identification Facilitation for Emotionally Arousing Verbs During the Attentional Blink. *Emotion*, *4*(1), 23–35. doi:10.1037/1528-3542.4.1.23

Keil, A., Ihssen, N., & Heim, S. (2006). Early cortical facilitation for emotionally arousing targets during the attentional blink. *BMC Biology*, *4*(1), 23. doi:10.1186/1741-7007-4-23

Keil, A., Moratti, S., Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2005). Additive Effects of Emotional Content and Spatial Selective Attention on Electrocortical Facilitation. *Cerebral Cortex*, *15*(8), 1187–1197. doi:10.1093/cercor/bhi001

Keil, A., Müller, M. M., Gruber, T., Wienbruch, C., Stolarova, M., & Elbert, T. (2001). Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials. *Clinical Neurophysiology*, *112*(11), 2057–2068. doi:10.1016/S1388-2457(01)00654-X

Khalfa, S., Schon, D., Anton, J.-L., & Liégeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *NeuroReport*, *16*, 1981–1984.

Kidd, G. J., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory Perception of Sound Sources* (pp. 143–189). Springer US.

Kiefer, M., & Martens, U. (2010). Attentional sensitization of unconscious cognition: Task sets modulate subsequent masked semantic priming. *Journal of Experimental Psychology: General*, *139*(3), 464–489. doi:10.1037/a0019561

King, A. J., & Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? : Abstract : Nature Neuroscience. *Nature Neuroscience*, *12*, 698–701. doi:10.1038/nn.2308.

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7– 50). Mahwah, NJ: Erlbaum.

Klauer, K. C., Roßnagel, C., & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 246–255. doi:10.1037/0278-7393.23.1.246

Kluge, C., Bauer, M., Leff, A. P., Heinze, H.-J., Dolan, R. J., & Driver, J. (2011). Plasticity of human auditory-evoked fields induced by shock conditioning and contingency reversal. *Proceedings of the National Academy of Sciences*, *108*(30), 12545–12550. doi:10.1073/pnas.1016124108

Koelewijn, T., Burg, E. V. der, Bronkhorst, A., & Theeuwes, J. (2008). Priming T2 in a visual and auditory attentional blink task. *Perception & Psychophysics*, *70*(4), 658–666. doi:10.3758/PP.70.4.658

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*(3), 170–180. doi:10.1038/nrn3666

Koelsch, S., Fritz, T., von Cramon, D. Y., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: An fMRI study. *Human Brain Mapping*, *27*, 239 –250. doi:10.1002/hbm.20180

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J.-D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping*, *34*(8), 1971–1981. doi:10.1002/hbm.22041

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, *86*(3), 366–376. doi:10.1016/S0093-934X(02)00532-1

Kotz, S. A., & Paulmann, S. (2011). Emotion, Language, and the Brain. *Language and Linguistics Compass*, *5*(3), 108–125. doi:10.1111/j.1749-818X.2010.00267.x

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). LmerTest: Tests in linear mixed effects models. R package version 2.0-20. Retrieved from http://CRAN.R-project.org/package=lmerTest

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-Five Years After Broadbent (1958): Still No Identification Without Attention. *Psychological Review*, *111*(4), 880–913. doi:10.1037/0033-295X.111.4.880

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The Spectrotemporal Filter Mechanism of Auditory Selective Attention. *Neuron*, *77*(4), 750–761. doi:10.1016/j.neuron.2012.11.034

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, *50*(5), 372–385. doi:10.1037/0003-066X.50.5.372

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

Lavie, N. (2010). Attention, Distraction, and Cognitive Control Under Load. *Current Directions in Psychological Science*, *19*(3), 143–148. doi:10.1177/0963721410370295

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1641), 20130205. doi:10.1098/rstb.2013.0205

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, *133*(3), 339–354. doi:10.1037/0096-3445.133.3.339

Leek, M. R., Brown, M. E., & Dorman, M. F. (1991). Informational masking and auditory attention. *Perception & Psychophysics*, *50*(3), 205–214. doi:10.3758/BF03206743

Lehmann, A., & Schönwiesner, M. (2014). Selective Attention Modulates Human Auditory Brainstem Responses: Relative Contributions of Frequency and Spatial Cues. *PLoS ONE*, *9*(1), e85442. doi:10.1371/journal.pone.0085442

Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(2), 189–207. doi:10.1037/0096-1523.5.2.189

Lollo, V. D., Kawahara, J., Ghorashi, S. M. S., & Enns, J. T. (2005). The attentional blink: Resource depletion or temporary loss of control? *Psychological Research*, *69*(3), 191–200. doi:10.1007/s00426-004-0173-x

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616–618. doi:10.1038/383616a0

Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception, & Psychophysics*, *73*(6), 1780–1789. doi:10.3758/s13414-011-0144-4

Marin, M. M., Gingras, B., & Bhattacharya, J. (2012). Crossmodal transfer of arousal, but not pleasantness, from the musical to the visual domain. *Emotion*, *12*(3), 618–631. doi:10.1037/a0025020

Martens, S., Johnson, A., Bolle, M., & Borst, J. (2008). A Quick Visual Mind Can be a Slow Auditory Mind. *Experimental Psychology*, *56*(1), 33–40. doi:10.1027/1618-3169.56.1.33

Martens, S., & Wyble, B. (2010). The attentional blink: Past, present, and future of a blind spot in perceptual awareness. *Neuroscience & Biobehavioral Reviews*, *34*(6), 947–957. doi:10.1016/j.neubiorev.2009.12.005

Massaro, D. W. (1975). Backward recognition masking. *The Journal of the Acoustical Society of America*, *58*(5), 1059–1065. doi:10.1121/1.380765

Mathews, A., & MacLeod, C. (2005). Cognitive Vulnerability to Emotional Disorders. *Annual Review of Clinical Psychology*, *1*(1), 167–195. doi:10.1146/annurev.clinpsy.1.102803.143916

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison approach*. Belmont, CA: Wadsworth.

McHugo, M., Olatunji, B. O., & Zald, D. H. (2013). The emotional attentional blink: what we know so far. *Frontiers in Human Neuroscience*, *7*. doi:10.3389/fnhum.2013.00151

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*(10), 1410–1421. doi:10.1016/S0028-3932(03)00017-4

Mondor, T. A. (1998). A transient processing deficit following selection of an auditory target. *Psychonomic Bulletin & Review*, *5*(2), 305–311. doi:10.3758/BF03212956

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing (6th ed.)*. Bingley: Emerald.

Moore, D. R., Halliday, L. F., & Amitay, S. (2009). Use of auditory learning to manage listening problems in children. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1515), 409–420. doi:10.1098/rstb.2008.0187

Moors, A. (2015). Automaticity: Componential, Causal, and Mechanistic Explanations. *Annual Review of Psychology*, Advance online publication. doi:10.1146/annurev-psych-122414-033550

Moors, A., & De Houwer, J. (2001). Automatic appraisal of motivational valence: Motivational affective priming and Simon effects. *Cognition & Emotion*, *15*(6), 749–766. doi:10.1080/02699930143000293

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*(2), 297–326. doi:10.1037/0033-2909.132.2.297

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. doi:10.1080/17470215908416289

Morris, J. S., Friston, K. J., Buchel, C., Frith, C. D., Young, A. W., Calder, A. J., & Dolan, R. J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain*, *121*, 47–57.

Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review*, *12*(4), 654–661. doi:10.3758/BF03196754

Müller, S., Rothermund, K., & Wentura, D. (2015). Relevance drives attention: Attentional bias for gain- and loss-related stimuli is driven by delayed disengagement. *The Quarterly Journal of Experimental Psychology*, Advance online publication. doi:10.1080/17470218.2015.1049624

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection. *Psychological Bulletin*, *92*(3), 605–640. doi:10.1037/0033-2909.92.3.605

Näätänen, R., Astikainen, P., Ruusuvirta, T., & Huotilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory–cognitive core of cognitive processes. *Brain Research Reviews*, *64*(1), 123–136. doi:10.1016/j.brainresrev.2010.03.001

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. doi:10.1016/0001-6918(78)90006-9

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. doi:10.1016/j.clinph.2007.04.026

Näätänen, R., & Picton, T. W. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. doi:10.1111/j.1469-8986.1987.tb00311.x

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826–859. doi:10.1037/0033-2909.125.6.826

Nielsen, S. L., & Sarason, I. G. (1981). Emotion, personality, and selective attention. *Journal of Personality and Social Psychology*, *41*(5), 945–960. doi:10.1037/0022-3514.41.5.945

Notebaert, L., Crombez, G., Van Damme, S., De Houwer, J., & Theeuwes, J. (2011). Signals of threat do not capture, but prioritize, attention: A conditioning approach. *Emotion*, *11*(1), 81–89. doi:10.1037/a0021286

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2009). Emotional scene content drives the saccade generation system reflexively. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 305–323. doi:10.1037/a0013626

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, *97*(2), 316–333. doi:10.1037/0033-2909.97.2.316

Oca, B. M. de, Villa, M., Cervantes, M., & Welbourne, T. (2012). Emotional Modulation of the Attentional Blink by Pleasant and Unpleasant Pictures. *The Journal of General Psychology*, *139*(4), 289–314. doi:10.1080/00221309.2012.708681

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, *30*(10), 953–958. doi:10.1016/j.psyneuen.2005.03.019

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466–478. doi:10.1037/0096-3445.130.3.466

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. doi:10.1037/0033-295X.108.3.483

Öhman, A., Soares, S. C., Juth, P., Lindström, B., & Esteves, F. (2012). Evolutionary derived modulations of attention to two common fear stimuli: Serpents and hostile humans. *Journal of Cognitive Psychology*, *24*(1), 17–32. doi:10.1080/20445911.2011.629603

Okon-Singer, H., Lichtenstein-Vidne, L., & Cohen, N. (2013). Dynamic modulation of emotional processing. *Biological Psychology*, *92*(3), 480–491. doi:10.1016/j.biopsycho.2012.05.010

Olivers, C. N. L., Hilkenmeier, F., & Scharlau, I. (2011). Prior entry explains order reversals in the attentional blink. *Attention, Perception, & Psychophysics*, *73*(1), 53–67. doi:10.3758/s13414-010-0004-7

Olivers, C. N. L., & Meeter, M. (2008). A boost and bounce theory of temporal attention. *Psychological Review*, *115*(4), 836–863. doi:10.1037/a0013395

Olivers, C. N. L., Stigchel, S. van der, & Hulleman, J. (2007). Spreading the sparing: against a limited-capacity account of the attentional blink. *Psychological Research*, *71*(2), 126–139. doi:10.1007/s00426-005-0029-z

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, *77*(3), 247–265. doi:10.1016/j.biopsycho.2007.11.006

Oxenham, A. J., Fligor, B. J., Mason, C. R., & Jr, G. K. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America*, *114*(3), 1543–1549. doi:10.1121/1.1598197

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00345

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Current Opinion in Neurobiology*, *15*(2), 188–196. doi:10.1016/j.conb.2005.03.002

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, *99*(17), 11458–11463. doi:10.1073/pnas.172403899

Pessoa, L., Padmala, S., & Morland, T. (2005). Fate of unattended fearful faces in the amygdala is determined by both attentional resources and cognitive modulation. *NeuroImage*, *28*(1), 249–255. doi:10.1016/j.neuroimage.2005.05.048

Petrova, K., & Wentura, D. (2012). Upper–lower visual field asymmetries in oculomotor inhibition of emotional distractors. *Vision Research*, *62*, 209–219. doi:10.1016/j.visres.2012.04.010

Phelps, E. A. (2006). Emotion and Cognition: Insights from Studies of the Human Amygdala. *Annual Review of Psychology*, *57*(1), 27–53. doi:10.1146/annurev.psych.56.091103.070234

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, *48*(2), 175–187. doi:10.1016/j.neuron.2005.09.025

Picton, T. W. (2013). Hearing in time: evoked potential studies of temporal processing. *Ear and Hearing*, *34*, 385–401.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 785–794. doi:10.1037/0278-7393.21.3.785

Potter, M. C., Staub, A., & O'Connor, D. H. (2002). The time course of competition for attention: Attention is initially labile. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1149–1162. doi:10.1037/0096-1523.28.5.1149

Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, *41*(1), 49–59. doi:10.1016/ S0010-9452(08)70177-1

Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological Correlates of Rapid Spatial Orienting Towards Fearful Faces. *Cerebral Cortex*, *14*(6), 619–633. doi:10.1093/cercor/bhh023

Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biological Psychology*, *92*(3), 492–512. doi:10.1016/j.biopsycho.2012.02.007

Pourtois, G., Spinelli, L., Seeck, M., & Vuilleumier, P. (2010). Temporal precedence of emotion over attention modulations in the lateral amygdala: Intracranial ERP evidence from a patient with temporal lobe epilepsy. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(1), 83–93. doi:10.3758/CABN.10.1.83

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*(3), 380–391. doi:10.1037/0022-3514.61.3.380

Raveh, D., & Lavie, N. (2014). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics*, *77*(2), 483–492. doi:10.3758/s13414-014-0776-2

Raymond, J. E., & O'Brien, J. L. (2009). Selective Visual Attention and Motivation The Consequences of Value Learning in an Attentional Blink Task. *Psychological Science*, *20*(8), 981–988. doi:10.1111/j.1467-9280.2009.02391.x

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. doi:10.1037/0096-1523.18.3.849

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating Irrelevant Motion Perception by Varying Attentional Load in an Unrelated Task. *Science*, *278*(5343), 1616–1619. doi:10.1126/science.278.5343.1616

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1085–1099. doi:10.1037/0096-1523.28.5.1085

Richards, V. M., & Neff, D. L. (2004). Cuing effects for informational masking. *The Journal of the Acoustical Society of America*, *115*(1), 289–300. doi:10.1121/1.1631942

Roberts, T. P., Ferrari, P., Stufflebeam, S. M., & Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: Stimulus dependence and insights toward perception. *Journal of Clinical Neurophysiology*, *17*, 114–129.

Rohr, L., & Abdel Rahman, R. (2015). Affective responses to emotional words are boosted in communicative situations. *NeuroImage*, *109*, 273–282. doi:10.1016/j.neuroimage.2015.01.031

Rohr, M., Degner, J., & Wentura, D. (2012). Masked emotional priming beyond global valence activations. *Cognition and Emotion*, *26*(2), 224–244. doi:10.1080/02699931.2011.576852

Rolke, B., Heil, M., Streb, J., & Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology*, *38*(02), 165–174. doi:null

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception*, *16*(2), 175–186.

Ross, B., Hillyard, S. A., & Picton, T. W. (2010). Temporal Dynamics of Selective Attention during Dichotic Listening. *Cerebral Cortex*, *20*(6), 1360–1371. doi:10.1093/cercor/bhp201

Rothermund, K. (2003). Motivation and attention: Incongruent effects of feedback on the processing of valence. *Emotion*, *3*(3), 223–238. doi:10.1037/1528-3542.3.3.223

Rothermund, K. (2011). Counter-Regulation and Control-Dependency. *Social Psychology*, *42*(1), 56–66. doi:10.1027/1864-9335/a000043

Rothermund, K., Gast, A., & Wentura, D. (2011). Incongruency effects in affective processing: Automatic motivational counter-regulation or mismatch-induced salience? *Cognition and Emotion*, *25*(3), 413–425. doi:10.1080/02699931.2010.537075

Rothermund, K., Voss, A., & Wentura, D. (2008). Counter-regulation in affective attentional biases: A basic mechanism that warrants flexibility in emotion and motivation. *Emotion*, *8*(1), 34–46. doi:10.1037/1528-3542.8.1.34

Rothermund, K., Wentura, D., & Bak, P. M. (2001). Automatic attention to stimuli signalling chances and dangers: Moderating effects of positive and negative goal and action contexts. *Cognition and Emotion*, *15*(2), 231–248. doi:10.1080/02699930126213

Roye, A., Jacobsen, T., & Schröger, E. (2007). Personal significance is encoded automatically by the human brain: an event-related potential study with ringtones. *European Journal of Neuroscience*, *26*(3), 784–790. doi:10.1111/j.1460-9568.2007.05685.x

Rutherford, H. J. V., & Lindell, A. K. (2011). Thriving and Surviving: Approach and Avoidance Motivation and Lateralization. *Emotion Review*, *3*(3), 333–343. doi:10.1177/1754073911402392

Sander, D., Grafman, J., & Zalla, T. (2003). The Human Amygdala: An Evolved System for Relevance Detection. *Reviews in the Neurosciences*, *14*(4), 303–316. doi:10.1515/REVNEURO.2003.14.4.303

Sander, K., & Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Cognitive Brain Research*, *12*(2), 181–198. doi:10.1016/S0926-6410(01)00045-3

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., & Formisano, E. (2014). Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. *PLoS Computational Biology*, *10*(1). doi:10.1371/journal.pcbi.1003412

Sauter, D. A., & Eimer, M. (2009). Rapid Detection of Emotion from Human Vocalizations. *Journal of Cognitive Neuroscience*, *22*(3), 474–481. doi:10.1162/jocn.2009.21215

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.

Scherer, L. D., & Larsen, R. J. (2011). Cross-modal evaluative priming: Emotional sounds influence the processing of emotion words. *Emotion*, *11*(1), 203–208. doi:10.1037/a0022588

Schirmer, A., & Escoffier, N. (2010). Emotional MMN: Anxiety and heart rate correlate with the ERP signature for auditory change detection. *Clinical Neurophysiology*, *121*(1), 53–59. doi:10.1016/j.clinph.2009.09.029

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. doi:10.1016/j.tics.2005.11.009

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228–233. doi:10.1016/S0926-6410(02)00108-8

Schröger, E. (1996). A Neural Mechanism for Involuntary Attention Shifts to Changes in Auditory Stimulation. *Journal of Cognitive Neuroscience*, *8*(6), 527–539. doi:10.1162/jocn.1996.8.6.527

Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: a lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641–664. doi:10.1111/ejn.12816

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies, *Progress in Brain Research, 156*, 31–51.

Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2003). Emotional facilitation of sensory processing in the visual Cortex. *Psychological Science*, *14*(1), 7–13. doi:10.1111/1467-9280.01411

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: An ERP analysis. *Emotion*, *4*(2), 189–200. doi:10.1037/1528-3542.4.2.189

Schupp, H. T., Stockburger, J., Codispoti, M., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2007). Selective visual attention to emotion. *The Journal of Neuroscience*, *27*(5), 1082–1089. doi:10.1523/JNEUROSCI.3223-06.2007

Schwabe, L., Merz, C. J., Walter, B., Vaitl, D., Wolf, O. T., & Stark, R. (2011). Emotional modulation of the attentional blink: The neural structures involved in capturing and holding attention. *Neuropsychologia*, *49*(3), 416–425. doi:10.1016/j.neuropsychologia.2010.12.037

Schwager, S., & Rothermund, K. (2013a). Counter-regulation triggered by emotions: Positive/negative affective states elicit opposite valence biases in affective processing. *Cognition and Emotion*, *27*(5), 839–855. doi:10.1080/02699931.2012.750599

Schwager, S., & Rothermund, K. (2013b). Motivation and affective processing biases in risky decision making: A counter-regulation account. *Journal of Economic Psychology*, *38*, 111–126. doi:10.1016/j.joep.2012.08.005

Schwager, S., & Rothermund, K. (2014). On the dynamics of implicit emotion regulation: Counter-regulation after remembering events of high but not of low emotional intensity. *Cognition and Emotion*, *28*(6), 971–992. doi:10.1080/02699931.2013.866074

Schwent, V. L., Hillyard, S. A., & Galambos, R. (1976). Selective attention and the auditory vertex potential. II. Effects of signal intensity and masking noise. *Electroencephalography and Clinical Neurophysiology*, *40*(6), 615–622. doi:10.1016/0013-4694(76)90136-X

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391–1400. doi:10.1038/nn1549

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1994). Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 357–371. doi:10.1037/0096-1523.20.2.357

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296. doi:10.1016/S1364-6613(97)01094-2

Shen, D., & Alain, C. (2010). Neuroelectric correlates of auditory attentional blink. *Psychophysiology*, *47*(1), 184–191. doi:10.1111/j.1469-8986.2009.00924.x

Shen, D., & Mondor, T. A. (2006). Effect of distractor sounds on the auditory attentional blink. *Perception & Psychophysics*, *68*(2), 228–243. doi:10.3758/BF03193672

Shen, D., & Mondor, T. A. (2008). Object file continuity and the auditory attentional blink. *Perception & Psychophysics*, *70*(5), 896–915. doi:10.3758/PP.70.5.896

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. doi:10.1037/0033-295X.84.2.127

Shinn, J. B. (2003). Temporal processing: The basics. *The Hearing Journal*, *56*. doi:10.1097/01.HJ.0000292557.52409.67

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. doi:10.1016/j.tics.2008.02.003

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. doi:10.1037/0033-2909.86.2.420

Smith, N. K., Cacioppo, J. T., Larsen, J. T., & Chartrand, T. L. (2003). May I have your attention, please: Electrocortical responses to positive and negative stimuli. *Neuropsychologia*, *41*(2), 171–183. doi:10.1016/S0028-3932(02)00147-1

Smith, N. K., Larsen, J. T., Chartrand, T. L., Cacioppo, J. T., Katafiasz, H. A., & Moran, K. E. (2006). Being bad isn't always good: Affective context moderates the attention bias toward negative information. *Journal of Personality and Social Psychology*, *90*(2), 210–220. doi:10.1037/0022-3514.90.2.210

Smith, S. D., Most, S. B., Newsome, L. A., & Zald, D. H. (2006). An emotion-induced attentional blink elicited by aversively conditioned stimuli. *Emotion*, *6*(3), 523–527. doi:10.1037/1528-3542.6.3.523

Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00015

Sollberger, B., Rebe, R., & Eckstein, D. (2003). Musical chords as affective priming context in a word-evaluation task. *Music Perception: An Interdisciplinary Journal*, *20*(3), 263–282. doi:10.1525/mp.2003.20.3.263

Spence, C., & Parise, C. (2010). Prior-entry: A review. *Consciousness and Cognition*, *19*(1), 364–379. doi:10.1016/j.concog.2009.12.001

Spruyt, A., De Houwer, J., Everaert, T., & Hermans, D. (2012). Unconscious semantic activation depends on feature-specific attention allocation. *Cognition*, *122*(1), 91–95. doi:10.1016/j.cognition.2011.08.017

Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007). Affective priming of nonaffective semantic categorization responses. *Experimental Psychology*, *54*(1), 44–53. doi:10.1027/1618-3169.54.1.44

Spruyt, A., De Houwer, J., & Hermans, D. (2009). Modulation of automatic semantic priming by feature-specific attention allocation. *Journal of Memory and Language*, *61*(1), 37–54. doi:10.1016/j.jml.2009.03.004

Spruyt, A., Klauer, K. C., Gast, A., Schryver, M. D., & De Houwer, J. (2014). Feature-specific attention allocation modulates the generalization of recently acquired likes and dislikes. *Experimental Psychology*. *61*(2), 85-98. doi: 10.1027/1618-3169/a000228

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience*, *23*(3), 604–621. doi:10.1162/jocn.2009.21383

Stein, T., Zwickel, J., Ritter, J., Kitzmantel, M., & Schneider, W. X. (2009). The effect of fearful faces on the attentional blink is task dependent. *Psychonomic Bulletin & Review*, *16*(1), 104–109. doi:10.3758/PBR.16.1.104

Stolarova, M., Keil, A., & Moratti, S. (2006). Modulation of the C1 Visual Event-related Component by Conditioned Stimuli: Evidence for Sensory Plasticity in Early Affective Perception. *Cerebral Cortex*, *16*(6), 876–887. doi:10.1093/cercor/bhj031

Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. doi:10.1207/s15327957pspr0803_1

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1–2), 22–29. doi:10.1016/j.heares.2009.12.021

Suzuki, Y., & Takeshima, H. (2004). Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, *116*(2), 918–933. doi:10.1121/1.1763601

Taatgen, N. A., Juvina, I., Schipper, M., Borst, J. P., & Martens, S. (2009). Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology*, *59*(1), 1–29. doi:10.1016/j.cogpsych.2008.12.002

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, *29*(7), 382–390. doi:10.1016/j.tins.2006.06.003

Thierry, G., & Roberts, M. V. (2007). Event-related potential study of attention capture by affective sounds. *NeuroReport*, *18*, 245–248. doi:10.1097/WNR.0b013e328011dc95

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention*. Macmillan.

Tolnai, S., Dolležal, L.-V., & Klump, G. M. (2015). Binaural cues provide for a release from informational masking. *Behavioral Neuroscience*, *129*(5), 589–598. doi:10.1037/bne0000091

Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Perception & Psychophysics*, *67*(2), 195–208. doi:10.3758/BF03206484

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Vachon, F., & Tremblay, S. (2005). Auditory Attentional Blink: Masking the Second Target is Necessary, Delayed Masking is Sufficient. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *59*(4), 279–286. doi:10.1037/h0087480

Vachon, F., Tremblay, S., Hughes, R. W., & Jones, D. M. (2010). Capturing and Unmasking the Mask in the Auditory Attentional Blink. *Experimental Psychology*, *57*(5), 346–353. doi:10.1027/1618-3169/a000041

VandenBos, G. R. (2009). *APA concise dictionary of psychology*. Washington, DC: American Psychological Association.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720–752. doi:10.1080/02699930701503567

Viemeister, N. (1996). Auditory Temporal Integration: What Is Being Accumulated? *Current Directions in Psychological Science*, *5*(1), 28–32.

Visser, T. A. W., Bischof, W. F., & Di Lollo, V. (1999). Attentional switching in spatial and nonspatial domains: Evidence from the attentional blink. *Psychological Bulletin*, *125*(4), 458. doi:10.1037/0033-2909.125.4.458

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*(4), 739–743. doi:10.3758/BF03196329

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1656–1674. doi:10.1037/0096-1523.24.6.1656

Voß, A., Rothermund, K., & Wentura, D. (2003). Estimating the valence of single stimuli: A new variant of the affective Simon task. *Experimental Psychology*, *50*, 86–96. doi:10.1026//1618-3169.50.2.86

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594. doi:10.1016/j.tics.2005.10.011

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of Attention and Emotion on Face Processing in the Human Brain. *Neuron*, *30*(3), 829–841. doi:10.1016/S0896-6273(01)00328-2

Vuilleumier, P., & Driver, J. (2007). Modulation of visual processing by attention and emotion: windows on causal interactions between human brain regions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 837–855. doi:10.1098/rstb.2007.2092

Vuilleumier, P., & Huang, Y.-M. (2009). Emotional Attention Uncovering the Mechanisms of Affective Biases in Perception. *Current Directions in Psychological Science*, *18*(3), 148–152. doi:10.1111/j.1467-8721.2009.01626.x

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*(11), 1271–1278. doi:10.1038/nn1341

Weinberg, A., & Hajcak, G. (2010). Beyond good and evil: The time-course of neural activity elicited by specific picture content. *Emotion*, *10*, 767–782.

Weise, A., Schröger, E., Fehér, B., Folyi, T., & Horváth, J. (2012). Auditory event-related potentials reflect dedicated change detection activity for higher-order acoustic transitions. *Biological Psychology*, *91*(1), 142–149. doi:10.1016/j.biopsycho.2012.06.001

Weninger, F., Eyben, F., Schuller, B. W., Mortillaro, M., & Scherer, K. R. (2013). On the Acoustics of Emotion in Audio: What Speech, Music, and Sound have in Common. *Frontiers in Psychology, Emotion Science*, *4*, 1–12. doi:10.3389/fpsyg.2013.00292

Wentura, D., Müller, P., & Rothermund, K. (2014). Attentional capture by evaluative stimuli: Gain- and loss-connoting colors boost the additional-singleton effect. *Psychonomic Bulletin & Review*, *21*(3), 701–707. doi:10.3758/s13423-013-0531-z

Wentura, D., Voss, A., & Rothermund, K. (2009). Playing TETRIS for science counter-regulatory affective processing in a motivationally “hot” context. *Acta Psychologica*, *131*(3), 171–177. doi:10.1016/j.actpsy.2009.05.008

Wiegand, K., & Gutschalk, A. (2012). Correlates of perceptual awareness in human primary auditory cortex revealed by an informational masking experiment. *NeuroImage*, *61*(1), 62–69. doi:10.1016/j.neuroimage.2012.02.067

Wiens, S., & Syrjänen, E. (2013). Directed attention reduces processing of emotional distracters irrespective of valence and arousal level. *Biological Psychology*, *94*(1), 44–54. doi:10.1016/j.biopsycho.2013.05.001

Williams, M. A., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2005). Differential amygdala responses to happy and fearful facial expressions depend on selective attention. *NeuroImage*, *24*(2), 417–425. doi:10.1016/j.neuroimage.2004.08.017

Williams, M. A., Visser, T. A. W., Cunnington, R., & Mattingley, J. B. (2008). Attenuation of Neural Responses in Primary Visual Cortex during the Attentional Blink. *The Journal of Neuroscience*, *28*(39), 9890–9894. doi:10.1523/JNEUROSCI.3057-08.2008

Winer, B. J. (1962). *Statistical Principles in Experimental Design*. New York, NY: McGraw Hill.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. doi:10.1016/j.tics.2009.09.003

Winkler, I., & Schröger, E. (2015). Auditory perceptual objects as generative models: Setting the stage for communication by sound. *Brain and Language*, *148*, 1–22. doi:10.1016/j.bandl.2015.05.003

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*(18), 8722–8726.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, *79*(3), 170–191. doi:10.1016/0013-4694(91)90136-R

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255–260. doi:10.1037/0278-7393.21.1.255

Woods, D. L. (1995). The component structure of the N 1 wave of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology-Supplements Only*, (44), 102–109.

Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and Generalization of Auditory Temporal–Interval Discrimination in Humans. *The Journal of Neuroscience*, *17*(10), 3956–3963.

Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M. M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature*, *387*(6629), 176–178. doi:10.1038/387176a0

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, *24*(1), 3–47. doi:10.1080/02699930903205698

Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews*, *41*(1), 88–123. doi:10.1016/S0165-0173(02)00248-5

Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, *11*(10), 946–953. doi:10.1093/cercor/11.10.946

Zhang, C., Lu, L., Wu, X., & Li, L. (2014). Attentional modulation of the early cortical representation of speech signals in informational or energetic masking. *Brain and Language*, *135*, 85–95. doi:10.1016/j.bandl.2014.06.002

Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., … Luo, Y. (2014). Three stages of emotional word processing: an ERP study with rapid serial visual presentation. *Social Cognitive and Affective Neuroscience*, *9*, 1897–903. doi:10.1093/scan/nst188

Appendices

Appendix A: Note on Applying Group Factor in Experiments 3-5 and Experiment 7

In Experiments 3-5 and in Experiment 7 we induced valence experimentally by associating affective meaning to tone-frequencies in a balanced design. We introduced counterbalancing group as a between-participants factor into the main analysis in these experiments for two reasons following the method suggested by Pollatsek and Well (1995). (1) First, we expected that physical differences of the four tones have a distinct impact on the “exogenous” auditory N1 (see, e.g., Burkard et al., 2007) and on tone-detection performance. Although employing a Latin square design allows evaluating the main effect of valence without being confounded by the tone differences (see, e.g., Winer, 1962), we have to consider the following: When applying a standard analysis, i.e., one-factorial repeated measures ANOVA (or MANOVA), the power of our analysis would be extenuated by the main effect of the orthogonal factor (here: tone-frequency). Thus, the first reason to apply counterbalancing group factor was to use the correct error term. In this analysis, the valence × counterbalancing group interaction accounts for the part of the variance introduced by the main effect of tone-frequency and thereby improves the power for the test for valence (Pollatsek & Well, 1995). It is important to note that this interaction does not mean what is suggested by denomination: It does not mean that the valence main effect is different for the random samples but typically due to a main effect of the second within-participants factor of the Latin square design – here: tone-frequency – which is not in the focus of interest in our experiments. Furthermore, note that in the method of Pollatsek and Well (1995) the reduction of the error sum of squares by including the grouping factor is adequately weighted by a reduced number of denominator *df*s. (2) Second, by introducing the counterbalancing group as additional factor into the main analysis, we accounted for the slightly different sample sizes of the final counterbalancing groups of Experiment 3 and Experiment 7.

Appendix B: Counterbalancing Schemes used in Experiments 3-5 and Experiment 7

Table A1. Assignment of tone-frequency to valence conditions according to a Latin square design in Experiment 3. No-go tones were presented only during the valence induction phase.

	Counterbalancing group			
Tone-frequency	**1**	**2**	**3**	**4**
Low	Neutral	Positive	No-Go	Negative
Middle Low	Positive	No-Go	Negative	Neutral
Middle High	No-Go	Negative	Neutral	Positive
High	Negative	Neutral	Positive	No-Go

Table A2. Assignment of tone-frequency to valence conditions according to a Latin square design in Experiments 4-5 and Experiment 7. In Experiment 5 and Experiment 7, no-go tones were presented with high tone-frequency in each balancing group during the valence induction phase. No-go tones were not applied in Experiment 4.

	Counterbalancing group		
Tone-frequency	**1**	**2**	**3**
Low	Neutral	Positive	Negative
Middle Low	Positive	Negative	Neutral
Middle High	Negative	Neutral	Positive