Embodied Auditory Perception: The Emotional Impact of Approaching and Receding Sound Sources

Ana Tajadura-Jiménez

Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden

Aleksander Väljamäe

Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden;
Laboratory for Synthetic Perceptive, Emotive and Cognitive Systems Institute of Audiovisual Studies, Universitat Pompeu Fabra, Barcelona, Spain

Erkin Asutay

Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden

Daniel Västfjäll

Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden;
Decision Reserch, Eugene, OR, USA

Author note

Ana Tajadura-Jiménez is now at the Psychology Department, Royal Holloway, University of London.

This work was carried out as part of the PRESENCCIA project, a European Union–funded Integrated Project under the IST Program (Project Number 27731). We appreciate the helpful comments of two anonymous reviewers on a previous version of this article.

Corresponding concerning this article should be addressed to Ana Tajadura-Jiménez, Psychology Department, Royal Holloway, University of London. Egham, Surrey, TW20 0EX, UK. Tel. +44(0)1784 276551; Fax. +44(0)1784 434347. E-mail: ana.tajadura@rhul.ac.uk

Abstract

Research has shown the existence of perceptual and neural bias towards sounds perceived as sources approaching versus receding a listener. It was suggested that a greater biological salience of approaching auditory sources may account for these effects. Also, that these effects may only hold for those sources critical for our survival. In the present study we bring support to these hypotheses by quantifying the emotional responses to different sounds with changing intensity patterns. In two experiments, participants were exposed to artificial and natural sounds simulating approaching or receding sources. The auditory-induced emotional effect was reflected on the performance of participants in an emotion-related behavioral task, their self-reported emotional experience and their physiology (electrodermal activity and facial electromyography). The results of this study suggest that approaching unpleasant sound sources evoke more intense emotional responses in listeners than receding ones, while such effect of perceived sound motion does not exist for pleasant or neutral sound sources. The emotional significance attributed to the sound source itself, the loudness of the sound and loudness change duration seem to be relevant factors in this disparity.

Keywords: emotion, biological salience, auditory sources, approaching, looming

Embodied Auditory Perception: The Emotional Impact of Approaching and Receding Sound Sources

Plenty of research during the last decades has been trying to understand how the physical parameters of sensory information relate to our perception, interpretation and response to events. However, there are still many unknowns in this area. The current research contributes to extend knowledge by focusing on the perception of sounds with changing intensity level, and to a larger extent, approaching versus receding sound sources. Previous sound research has found asymmetries, both at perceptual and neural levels, in detecting and responding to sounds with rising versus falling intensity level. Sounds with rising intensity level tend to be judged as being louder (c.f. Susini, McAdams, & Smith, 2007; Stecker & Hafter, 2000), having a higher loudness change (as assessed by global loudness ratings; Neuhoff, 1998; Neuhoff, 2001; Susini, McAdams, & Smith, 2007) and being longer in perceived duration (DiGiovanni & Schlauch, 2007; Grassi & Darwin, 2001; Schlauch, Ries, & DiGiovanni, 2001) than those with decreasing or with constant intensity level. Similarly, a directional sensitivity for sounds with rising intensity over those with falling or constant intensity level seems to exist also at neural level, as found both in animal (Maier, Chandrasekaran, & Ghazanfar, 2008; Maier & Ghazanfar, 2007; Romei, Murray, & Thut, 2008) and human research (Bach *et al.*, 2008; Romei, Murray, & Thut, 2008; Seifritz *et al.*, 2002). Interestingly, these asymmetrical effects are not general to all sorts of sound sources: Previous research has shown the existence of an asymmetry in loudness or duration judgments between sounds with increasing versus decreasing intensity level for tones, synthetic vowel sounds (Neuhoff, 1998), words and a drum strike (DiGiovanni & Schlauch, 2007), but not for broadband noise (Neuhoff, 1998).

In view of these findings, some authors have adopted an ecological perspective and suggested that a greater biological salience of sounds with increasing intensity level might account for these effects (Guski, 1992; Maier & Ghazanfar, 2007; Neuhoff, 1998; Neuhoff, 2001; Neuhoff, Planisek, & Seifritz, 2009; Stecker & Hafter, 2000). This view considers that rising/falling-intensity level sounds are mostly perceived as approaching/receding sound sources (e.g., Maier & Ghazanfar, 2007; Neuhoff, 2001), given that auditory research has shown that intensity change seems to be the most salient cue for auditory motion perception (Lutfi & Wang, 1999; Rosenblum, Carello, & Pastore, 1987). Then, these authors propose the existence of a "perceptual priority" for detecting approaching vs. receding sound objects. Moreover, it has been suggested that the fact that the effect is limited to tones or tone-like sounds is due to these sounds being similar to natural sounds that might be critical for our survival (Neuhoff, 1998).

In fact, there is some evidence from research with natural sounds pointing to the existence of such perceptual priority for approaching sources: People are able to anticipate the time to contact of both visual and auditory approaching sources (Gordon & Rosenblum, 2005; Neuhoff, 2001; Rosenblum, Carello, & Pastore, 1987; Rosenblum, Wuestefeld, & Saldana, 1993; Schiff & Oldak, 1990). Overall, ample research on moving objects has shown that animals have evolved to process dynamic information in a different way than static information (Ghazanfar, Neuhoff, & Logothetis, 2002; Maier, Neuhoff, Logothetis, & Ghazanfar, 2004; Neuhoff, 2001; Schiff, Caviness, & Gibson, 1962). An example of this asymmetry in the processing of static and dynamic information can be found in the so-called 'flash-lag effect' in which a brief flash presented physically aligned with a moving object appears to lag behind the moving object (e.g., Mackay, 1958). This effect is not restricted to vision since a similar effect has been found on audition, with a brief tone spatially aligned with a moving sound source, and it can also occur cross-modally, as shown between audition and vision (Alais & Burr, 2003). Possible explanations for these effects include that our brain may be able to extrapolate the dynamic information of a moving object to predict its future location (e.g., Nijhawan, 1994) or that moving stimuli may evoke a larger attention shift than static ones (Baldo & Klein, 1995). A "perceptual priority" for moving objects provides a significant survival advantage that might be critical for the case of objects approaching our body, since they may potentially threat our safety (e.g., Neuhoff, 1998). In this case, a perceptual priority for detecting approaching objects would increase the time and the attentional resources available to avoid those objects.

However, to the best of our knowledge, there is still not formal evidence of a greater biological salience or perceptual priority of approaching over receding sound sources[1]. There is also a need for support for the hypothesis that the approaching/receding disparity is limited to sources critical for our survival. One possibility to bring support to these hypotheses is to quantify the emotional responses of listeners to different approaching/receding sound sources, because emotional responses are directly linked to salience, and ultimately, to survival. In its most basic form, affect (or emotion) can be seen as the human alarm system (LeDoux, 1996): Negative affect signals a potential threat and the need for taking an action. In this way, emotions serve to establish our position *vis-à-vis* our environment (Levenson, 1994) and to keep a constant margin of safety around our body (Damasio, 1999; Graziano, 2001). Any alterations in the position of nearby objects with respect to us (to our body) may activate our alarm system and this might be especially the case when the objects detected are rapidly approaching us. Some previous evidence in this line suggests that tones with rising versus falling intensity level facilitate activation of the amygdala, a brain region identified as a warning area (Bach *et al.*, 2008). Besides, the same study showed that approaching sounds, compared to receding or static sounds, enhance autonomic orienting response, as measured by heart rate and electrodermal responses. Approaching sounds were found to accelerate reaction times (RTs) to auditory (constant intensity tones), but not to visual, targets in a subsequent Go-NoGo-task. Those effects on autonomic orienting response and speed of RTs might be attributed to emotional processes, but still, a direct link with emotion needs to be established. Emotions arise in response to significant events and evoke an automatic, frequently prior to awareness, response that modulates the subsequent attentional, perceptual and behavioral processes (Barrett, Niedenthal, & Winkielman, 2005; Öhman, 1993; Phelps & LeDoux, 2005; Phelps, Ling, & Carrasco, 2006). For instance, in visual dot-probe tasks facilitation in reaction time is observed when the target (a dot-probe) appears after a short-time interval at the same location as emotional eliciting stimuli (e.g., an angry face in Bradley, Mogg, & Millar, 2000).

Hence, the specific aim of the current studies was to investigate listeners' emotional responses to sounds perceived as approaching or receding auditory sources. We adopted a perspective of embodied emotion (e.g., Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005; Niedenthal, 2007), which considers that in the interaction with the environment our body plays a major role. It acts as a reference frame (Damasio, 1999) and forms the basis for all information processing, including cognitive, perceptual, emotional and action response processing (see embodied theories, e.g. Niedenthal, *et al.*, 2005; Niedenthal, 2007). In addition, we considered a motivational approach to emotion that assumes that affect is processed by at least two specialized systems, an approach and a defense system, in charge of processing appetitive (i.e. positive) and threat-related (i.e. negative) information, respectively (Cacioppo & Gardner, 1999; Lang, Bradley, & Cuthbert, 1990). This approach allows characterizing emotions in terms of two continuous dimensions: *Valence* or *pleasantness* (positive versus negative) and *arousal* or *activation* (excited versus calm; Lang, 1995; Russell, 1980). We combined different measures of emotional reactions: subjective, physiological and behavioral measures. We included one novel behavioral measure which quantified the reaction times to emotional photographs presented after the different sound conditions, thus, directly assessing the emotional impact of listening to the sounds. In two experiments we questioned whether (1) the perceived change in spatial distance of sound sources influences the attributed significance and the emotional responses to them, and whether (2) the effects are the same for all types of sound sources or rather there is an interaction between their physical and semantic parameters. In other words, we looked at whether the disparity is dependent on listeners' categorization of auditory source as being a potential threat or not. A previous study in the visual domain found that unpleasant pictures perceived as approaching led to more intense self-reported emotional experience and to increased startle responses than unpleasant pictures perceived as static or receding (Muhlberger, Neumann, Wieser, & Pauli, 2008). No such effects of perceived source motion were observed for neutral and pleasant pictures. The authors attributed this effect of picture motion, which was only observed for unpleasant stimuli, to a "negativity bias" (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). The negativity bias can be described as an urgent survival-related need to undertake specific actions against potential threats; a need that does not exist when exposed to neutral or pleasant events. We expected similar effects for the case of moving auditory sources, especially since the auditory system is considered to act as a warning system (e.g., Juslin & Västfjäll, 2008). The auditory system has been characterized as a change detector that responds to certain sound properties indicating a rapid change (e.g., sudden noise, increase or decrease in intensity or pitch, sensory dissonance), which might convey danger, by producing an increased activation of the central nervous system (brain stem reflex; c.f. Juslin & Västfjäll, 2008).

To begin with, Experiment 1 included artificial tones, similar to those used in most of previous research (e.g. Neuhoff, 1998). We collected behavioral (Experiments 1a and 1b), subjective and physiological (Experiment 1c) responses to tones with increasing or decreasing intensity level. We experimented with tones with different ramp durations and intensity ranges. From an ecological perspective, the intensity ramp duration of a sound source is associated to the source velocity, and the intensity range, with size or proximity of the source to the listener. If the asymmetry between sounds with rising and falling intensity levels is due to rising tones being more salient or categorized as a potential threat, then we expected that:

(1) Rising intensity tones will elicit more unpleasant and arousing emotional responses than falling intensity tones, because approaching sources may represent a bigger threat to the listener than receding sources;

(2) these emotional responses will be modulated by ramp duration and intensity range of the sounds, given the ecological implications of these physical parameters (velocity, size, proximity).

Hence, with Experiment 1 we aimed to extending previous research on artificial tones and providing support to the biological salience hypothesis proposed by some authors. Subsequently, we investigated to what extent the effects observed using a simple tone could be affected by the manipulation of the nature of the auditory source: In Experiment 2 we included natural sounds differing on their conveyed emotional significance. We questioned whether the effects hold for all types of sound sources or there is a negativity bias for auditory moving sources, similar to the one observed for visual sources. Thus, our hypothesis was that:

> (3) The approaching/receding disparity is dependent on the type of auditory source, specifically, on the auditory source itself being categorized as a potential threat.

Similarly to many of the previous studies in this area, the present research adopts an analogy between rising/falling-intensity level sounds and approaching/receding auditory sources (for the sake of simplicity, these will be referred to as approaching/receding sounds in the rest of the text).

Experiment 1

In Experiment 1 we investigated the emotional responses to approaching or receding artificial sound sources. Approaching and receding sounds (i.e. sound sources) were simulated by tones rising or falling in intensity level, respectively. Our hypothesis was that approaching tones would lead to more intense emotional responses in listeners compared to receding tones. Emotional responses were assessed by a multi-level measurement approach to emotional responses that combines behavioral, subjective and physiological measures (Bradley & Lang, 2000). In particular, Experiments 1a and 1b explored the effect of sounds on the behavior on a subsequent emotion-related task. In Experiment 1c, we investigated the subjective and physiological emotional effects of listening to the different sounds. For the behavioral task, unpleasant-arousing and neutral photographs were presented after the tones for a speeded alternative forced-choice task. If auditory approaching sources are more significant than receding ones, we expected faster RTs to photographs presented after approaching versus receding sounds (*Hypothesis 1*). Moreover, if this approaching/receding disparity is due to listeners categorizing approaching sound sources as potential threats, we expected that the asymmetry in RTs to photographs presented after approaching vs. receding sounds would be larger for the unpleasant-arousing photographs, as compared to the neutral ones (the negativity bias; *Hypothesis 2*). We expected that the asymmetry between approaching/receding tones would be reflected also at subjective and physiological levels; in other words, that approaching tones would be subjectively rated as more arousing and unpleasant (*Hypothesis 3*), and elicit larger physiological changes in listeners (*Hypothesis 4*) than receding tones. These hypotheses were examined with tones increasing and decreasing in intensity level, with three different durations of intensity ramp (Experiments 1a and 1c) and with two different ranges of intensity change (Experiments 1b and 1c). We expected that the emotional responses to the different sounds would be modulated by ramp duration and intensity change (*Hypothesis 5*): In a natural context, short sounds inform of rapid changes and high-energy dissipation; thus, they may represent high forces, stiff materials, or rapidly moving objects, which may be felt as dangerous. Intensity may be related to power and size, and it is inversely proportional to the distance to the sound object; thus, loud sounds may become associated to danger (Herman & Ritter, 2004; Lenti Boero & Bottoni, 2008).

Experiment 1a

Methods

Participants. 12 participants took part (6 females; mean age 26 years; age range: 22-34 years). Participants had normal hearing and were naïve as to the purposes of the study. They were paid for their time and gave their informed consent prior to the inclusion in the study. The experiments were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, materials and data collection. The experiment was conducted in a dark sound-attenuated room where the participants seated in a chair. The auditory stimuli were delivered binaurally via headphones (Sennheiser HD-414) and consisted of 1 kHz pure tones (44.1 kHz sampling rate) rising or falling in intensity level. The range of intensity change was from 68 to 86 dBA (as measured at the participant's ear position), and the duration of intensity ramp varied between three possible values (1, 2 or 3 s). To avoid startle response all stimuli were preceded and followed by a 300 ms constant intensity tone (Maier & Ghazanfar, 2007), thus resulting in stimuli with a total duration of 1.6, 2.6 or 3.6 s (see Figure 1). A 10 ms onset/offset ramp was applied to the auditory stimuli to prevent clipping.

Visual stimuli were presented with a LCD-screen positioned in front of the participants at eye level and 0.5 meters away from them. Image resolution was 1024x768 pixels and the visual angle for the images was 37°×30°. The visual stimuli consisted of fifteen 'neutral' and fifteen 'negative, moderate emotion-arousing' photographs (referred as 'negative' or 'unpleasant' photographs later in the text) taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997). Their mean IAPS normative valence and arousal ratings (in a 9-point scale) for ‘neutral’ photographs were 4.9 ± 1.1 (± standard deviation (SD) indicated) and 2.6 ± 1.8, respectively, and for ‘negative’ photographs, 3 ± 1.6 and 5.2 ± 2.2. ‘Neutral’ photographs depicted mushrooms or home objects, while ‘negative’ photographs showed ruins, dirtiness, accidents or guns. In addition, six ‘positive’ stimuli (mean valence and arousal ratings were 7 ± 1.8 and 4.8 ± 2.2 respectively) were included as a contrast to ‘negative’ photographs, in order to avoid a response bias in participants towards ‘neutral’. Responses to the trials with ‘positive’ photographs were not included in the analysis of the results. Finally, six extra photographs were used to instruct participants in their tasks for the experiment.

A game pad held in the participant’s hands was used to collect participant’s data. Presentation® software (Version 9.90) was used to control stimuli delivery and record responses.

Design. The experimental trials consisted of a pair of stimuli formed by a tone followed by a photograph. The photograph was ‘positive’ (not analyzed), ‘neutral’ or ‘negative’. The tone was either rising or falling in intensity level, which is referred to as, respectively, ‘approaching’ or ‘receding’ sound later in the text. The duration of intensity ramp varied between three possible values (1, 2 or 3 s); this resulted in twelve possible conditions with a 2 x 3 x 2 factorial design (Sound Direction [approaching, receding] x Ramp Duration [1, 2, 3 s] x Picture Emotional Valence [Negative, Neutral]).

Procedure. Participants arrived individually to the laboratory. At the start of the experimental session, the headphones were positioned. Participants were instructed to centrally fixate the screen during each experimental trial and to perform a speeded three-alternative forced-choice (3AFC) task regarding their feelings towards the photograph (‘positive’, ‘negative’ or ‘neutral’; ‘neutral’ was defined as *“neither positive nor negative”*).RTs for each trial were collected. The speed of responding was emphasized to participants, but they were also instructed to refrain from making anticipatory responses. Responses were given by pressing one out of three buttons in the gamepad. The participants completed a practice block with 6 trials to familiarize themselves with the experimental paradigm. Next, the participants completed two experimental blocks with each experimental condition presented 15 times within the block plus 6 extra trials with positive photographs (which were not considered in the subsequent analysis). This resulted in 186 trials per block. The different stimulus conditions were presented randomly, with an inter-trial interval of 3.5 s. During the last 2.5 s of this inter-trial interval a countdown from 5 to 1 was presented, which was intended to avoid confounding effects of startle response to the auditory stimuli. In this countdown, numbers were displayed in the screen for 500 ms, with numbers from 5 to 3 accompanied by a 250-ms 1 kHz tone, 86 dBA. Figure 1 schematically represents the time course for the trials. Each block of trials lasted on average for 15 minutes. Participants had a short break between blocks.

[Insert Figure 1 about here]

Data analyses. RTs exceeding ±3 standard deviations from the mean RT for each participant were recursively discarded (e.g., Kitagawa & Spence, 2005). Repeated measures analyses of variance (ANOVA) were performed on the RT data from the various different conditions. The within-participants factors differed for each of the experiments. Alpha level was fixed at 0.05 for all statistical tests. Greenhouse-Geisser correction was used to correct for unequal variances. Significant effects by ANOVA were followed by paired t-test comparisons.

Results

On average, a total of 92.6 ± 2.6 percent (± SD indicated) of the trials from each participant were included in the data analyses of the RTs. The within-participants factors for the ANOVA were 'sound direction' (approaching or receding), 'ramp duration' (1, 2 or 3 s) and 'picture emotional valence' (negative or neutral). The results (see Fig. 2A – left panel) showed that there was a significant main effect of each of the factors. For 'sound direction', the RTs were significantly faster (15 ms) for photographs presented after approaching versus receding sounds ($F(1, 11) = 7.975$; $p = .017$). The 'ramp duration' had a significant effect on the RTs ($F(1.9, 21) = 10.36$; $p < .001$), with faster RTs for photographs presented after sounds with a ramp duration of 2 or 3 s, as compared with those presented after the 1 s-ramp sounds. Follow-up comparisons (paired *t*-test) revealed significant differences ($p < .01$) between 1 and 2 s-ramps and between 1 and 3 s-ramps, but not between 2 and 3 s-ramps ($p = 1$). Finally, for the factor 'picture emotional valence', RTs were facilitated (more than 70 ms) for negative vs. neutral photographs ($F(1, 11) = 14.4$; $p = .003$). In fact, the difference between approaching and receding sounds in RTs to photographs was more evident in the responses to negative photographs, compared to those to neutral photographs (see Fig. 2A – left panel). Follow-up comparisons revealed that the approaching/receding difference for the 3 s-ramp sounds was significant for negative but not for neutral pictures ($t(11) = -2.5$; $p = .028$). No other significant interactions between factors were observed.

Experiment 1b

Methods

Participants. 16 participants took part (7 females; mean age 25 years; age range: 21-43 years). Participants had normal hearing and were naïve as to the purposes of the study. They were paid for their time and gave their informed consent prior to the inclusion in the study. The experiments were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, materials and data collection. The experimental set-up, visual stimuli and data collection were identical to those used in Experiment 1a. The auditory stimuli differed in that in Experiment 1b, the duration of intensity ramp was fixed to 2 s, and there were two possible ranges of intensity change, 'loud' (68-86 dBA) and 'soft' (50-68 dBA). Critically, 'soft' sounds had lower intensity level than 'loud' sounds at any point (analogous to the design in Maier & Ghazanfar, 2007).

Design. In Experiment 1b, the ramp duration was fixed to 2 s, but the range of intensity change varied between 'loud' and 'soft' ranges. This resulted in eight possible conditions with a 2 x 2 x 2 factorial design (Sound Direction [approaching, receding] x Intensity Range [Loud, Soft] x Picture Emotional Valence [negative, neutral]).

Procedure and data analyses. The procedure and data analyses were identical to those described for Experiment 1a, except that the experimental design resulted in 126 trials per block.

Results

On average, a total of 94 ± 3.4percent (± SD indicated) of the trials from each participant (range: 87.1-97.5) were included in the data analyses of the RTs. The within-participants factors for the ANOVA were 'sound direction' (approaching or receding), 'intensity range' (loud or soft) and 'picture emotional valence' (negative or neutral). The results (see Fig. 2A – right panel) revealed that RTs were significantly faster (58 ms) for photographs presented after a sound with 'loud' vs. 'soft' intensity level ($F(1, 15) = 6.8$; $p = .02$). For 'picture emotional valence', RTs were faster (more than 65 ms) for negative vs. neutral photographs ($F(1, 15) = 8.6$; $p = .01$). A significant interaction between 'sound direction' and 'picture emotional valence' was observed ($F(1, 15) = 6.1$; $p = .026$)**,** which revealed that the effect of 'sound direction' on RTs was more prominent for negative photographs. To explore this effect, a second ANOVA was performed only in the conditions containing negative photographs. This ANOVA, containing as within-participants factors 'sound direction' (approaching or receding), and 'intensity range' (loud or soft), revealed a significant main effect for each of the factors, with faster RTs for photographs presented after 'loud' vs. 'soft' sounds ($F(1, 15) = 6.4$; $p = .023$) and after approaching vs. receding sounds ($F(1, 15) = 5$; $p = .041$).

[Insert Figure 2 about here]

Experiment 1c

Methods

Participants. 28 participants took part (13 females; mean age 27 years; age range: 18- 46 years; 7 had also participated in Experiment 1a and 3 in Experiment 1b). Participants had normal hearing and were naïve as to the purposes of the study. They were paid for their time and gave their informed consent prior to the inclusion in the study. The experiments were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, materials and data collection. The experimental set-up was similar to the one described above. Experiment 1c contained all auditory stimuli from Experiments 1a and 1b, but did not include visual stimuli. A keyboard was used for registering participant's self-reported emotional responses. Furthermore, in Experiment 1c, a BIOPAC *MP150* System was used to record electrodermal activity (EDA, using a *GSR100C* amplifier) and facial electromyography (EMG, using *EMG100C* amplifiers) of participants while listening to the sounds. For recording EDA, surface Ag/AgCl circular electrodes with 8 mm diameter were attached to the first and second fingers of the participants' nondominant hand. Facial EMG activity was recorded from the left *Corrugator Supercilii* (CS) and *Zygomatic Major* (ZM) muscles, using surface Ag/AgCl circular electrodes with 4 mm diameter. EDA is a sensitive and valid indicator for the lower arousal range (Boucsein, 1992), while the activity of the CS and ZM muscles is assumed to be linked to unpleasant and pleasant emotions respectively (Andreassi, 1995). All electrodes were filled with electrode paste and attached to the previously cleaned skin. Physiological signals were sampled (at a rate of 3125 Hz for EMG signals, and a rate of 390.6 Hz for EDA signals) and amplified. The digital data collection was controlled by AcqKnowledge 3.8.1 software.

Design. Experiment 1c looked at the subjective and physiological responses to the sound conditions of Experiments 1a and 1b (8 sound conditions in total, including 'approaching' or 'receding' sounds with 3 ramp durations, and with 'loud' and 'soft' versions of the 2s ramp). Therefore, the same two factorial designs used in those experiments, excluding the factor ‘picture emotional valence’ (since Experiment 1c did not include pictures), were used here (i.e. 2 Sound Direction x 3 Ramp Duration and 2 Sound Direction x 2 Intensity Range, with the ‘loud’ version of the 2s ramp sounds included in both designs).

Procedure. *A*t the start of the experimental session the electrodes were attached. During a rest period of about 5 minutes, physiological recordings were tested and participants completed a short practice block to familiarize themselves with the experimental paradigm. Next, the participants completed six experimental blocks, each of them containing all eight sound conditions. During even blocks participants were only required to listen to the sounds, presented one after another with a silent inter-stimulus interval of 3 seconds. During odd blocks, after each sound, participants were asked to rate their feelings towards that sound using the 9-point valence and arousal pictorial scales of the Self-Assessment Manikin (SAM, Lang, 1980).

Data analyses. Self-reported valence and arousal were used as dependent variables for two multivariate ANOVAs (MANOVAs) where Wilks’ Lambda was used as the multivariate criterion. For the physiological data, EDA and facial EMG recordings were individually inspected for possible artifacts. EMG signals were band-pass filtered from 10 to 400 Hz (Andreassi, 1995). Change scores were calculated separated for each EMG signal by subtracting the average response for each 1-second interval for the 6 seconds following sound onset from the mean activity during the 1 s preceding sound onset (baseline), yielding 6 time intervals per sound (e.g., Dimberg, 1990). The average of the change scores from 0 to 4 seconds following sound onset was used for the factorial analysis. For EDA signals, change scores were calculated by subtracting the maximum amplitude response for each 1-second interval for the 6 seconds following sound onset from the maximum amplitude response during the 1-second baseline. The maximum of the change scores from 3 to 6 seconds following sound onset was used for the factorial analysis. All data were individually z-scored to control for individual differences in responsiveness.

Results

Effects on self-report. Self-reported valence and arousal ratings for the emotional responses to the different sounds in Experiment 1c were used as dependent variables for two MANOVAs. The first MANOVA, performed on the results for the 'loud' sounds set, containing as within-participants factors 'sound direction' and 'ramp duration' (2x3 design). The results (see Fig. 2B; higher ratings of valence and arousal represent more pleasant and arousing experiences, respectively) revealed that there was a significant main effect of both factors, 'sound direction' ($F(2, 26) = 34.4$; $p < .001$, $\Lambda = .274$) and 'ramp duration' ($F(4, 106) = 21.1$; $p < 0001$, $\Lambda = .31$), which was significant ($p < .001$) for both valence and arousal ratings. It also existed a significant interaction between 'sound direction' and 'ramp duration' factors ($F(24$ $106) = 3.7$; $p = .007$, $\Lambda = .771$), mainly reflected in the valence ratings. Approaching and longer sounds were perceived as more unpleasant and arousing than their counterparts.

For the second MANOVA, performed on the results for the 2 s-duration sounds, the within-participants factors were 'sound direction' and 'intensity range' (2x2 design). The results (see Fig. 2B) showed that there was a significant effect of both factors, 'sound direction' ($F(2, 26) = 20.3$; $p < .001$, $\Lambda = .391$) and 'intensity range' ($F(2, 26) = 91.4$; $p < .001$, $\Lambda = .125$), which was significant ($p < .001$) for both valence and arousal ratings. It also existed a significant interaction between both factors ($F(2, 26) = 3.5$; $p = .046$, $\Lambda = .79$), mainly reflected in the valence ratings. Approaching and louder sounds were perceived as more unpleasant and arousing than their counterparts.

Interestingly, the self-reported results (Fig. 2B) showed that the feelings towards the artificial tonal sound are mostly unpleasant (valence rating < 5).

Effects on physiology. After individually inspecting the physiological recordings from Experiment 1c, ZM data of 12 participants were excluded from analysis due to the presence of artifacts, leaving 16 valid ZM data sets. The analysis of the results entailed two separate ANOVAs that examined the effects on physiology for the different sound conditions. The first ANOVA tested the effects of ‘loud’ sounds. The within-participants factors were ‘sound direction’ (approaching or receding) and ‘ramp duration’ (1, 2 or 3 s). The results (see Fig. 3) revealed that there was a significant effect of ‘sound direction’ on the CS ($F(1,27) = 5.1$; $p = .033$) and ZM muscle activities ($F(1,15) = 15.6$; $p < .001$), with approaching sounds leading to larger muscle response than receding sounds. The interaction between ‘sound direction’ and ‘ramp duration’ had a significant impact on EDA ($F(1.9,50) = 4.7$; $p = .015$) and ZM muscle activity ($F(1.9,28) = 3.6$; $p = .044$), with larger differences between approaching and receding sounds for the conditions with longer ramp. A reversed pattern was observed between the 1s-ramp sounds and the 2 and 3s-ramp sounds in the EDA elicited by approaching or receding sounds.

The second ANOVA was performed on the results for sounds with 2 s-ramp duration. The within-participants factors were ‘sound direction’ (approaching or receding) and ‘intensity range’ (‘loud’ or ‘soft’). The results (see Fig. 3) showed a trend on the effect of the factor ‘intensity range’ on the CS activity ($F(1,27) = 3.6$; $p = .068$), with larger muscle response for ‘loud’ vs.

’soft’ sounds. The interaction between ‘sound direction’ and ‘intensity range’ showed a significant effect on EDA (*F*(1,27) = 4.2; *p* < .05), with a reversed pattern observed between ‘loud’ and ‘soft’ sounds in the response to approaching and receding sounds. This pattern for ‘loud’ sounds showed a larger elicited EDA for approaching than for receding sounds.

[Insert Figure 3 about here]

Discussion

Experiment 1 investigated the emotional responses to tones rising or falling in intensity level. This experiment included several sound conditions, which differed in the duration of the intensity ramp and in the range of intensity change. Our main hypothesis was that if approaching sound sources were more salient, they would exert a more intense emotional response in listeners than receding sound sources. The sound influence on the emotional state of listeners was assessed with a behavioral emotion-related task, self-report and measuring physiology. For the behavioral task (Experiments 1a and 1b), we measured the reaction times to ‘negative’ and ‘neutral’ photographs presented immediately after each of the sounds. As expected (*Hypothesis 1*) approaching sounds resulted in faster RTs to photographs than those obtained for receding sounds. This approaching/receding disparity was especially evident in the RTs to ‘negative’ photographs (*Hypothesis 2*). These findings support the assumptions that listeners may categorize approaching sounds as being more salient, thus, requiring a quick action from the listener’s side. Moreover, approaching sounds may be specifically categorized as potential threats, since a negativity bias was observed when comparing conditions with ‘negative’ and ‘neutral’ photographs.

Furthermore, the emotional response to approaching/receding sound sources was modulated by ramp duration and intensity change (*Hypothesis 5*). Participants had faster RTs to photographs presented after 'loud' sounds and sounds with longer intensity ramp. In a natural context, loud sounds may be emitted by sources that are closer to the listener, with bigger size or with more power than those sound sources emitting soft sounds. Hence, loud sound sources may become associated to danger. On the contrary, approaching sounds with a long ramp for intensity change may resemble a sound source approaching the listener at a lower velocity than short-ramp sounds. Thus, it might be hypothesized that short-ramp sounds, i.e. approaching faster the listener, might constitute a bigger threat to listeners than long-ramp sounds. We propose two tentative explanations that might account for the faster response to longer approaching sounds observed in our results. One possible explanation might be found in the fact that sound intensity perception is dependent on the signal duration (Small, Cox, & Brandt, 1962; Zwicker & Feldtkeller, 1967). For the 1 s-ramp sounds to be perceived as loud as the others, the intensity level should be increased by roughly 15/20 dB. Therefore, the less intense emotional responses elicited by the short-ramp sounds might be the result of them being perceived as more 'quiet' than the longer-ramp sounds. An alternative explanation might be that the longer the sounds are presented, the more time listeners have to evaluate the sound and prepare for the response.

Consistent with our hypotheses, approaching artificial sounds were subjectively rated as more arousing and unpleasant (*Hypothesis 3*), and had a larger physiological effect in listeners (*Hypothesis 4*), than receding ones. This effect on physiology was reflected in the facial EMG measures; it was observed also in the EDA but only for the conditions with 'loud', 2s and 3s-ramp sounds (the explanation proposed in the previous paragraph for the short-ramp sounds may also apply here). In addition, louder sounds elicited a larger CS muscle activity, what may correspond to a more unpleasant emotional experience (Andreassi, 1995). Longer approaching sounds elicited a larger ZM muscle activity than their receding counterpart. In general, ZM muscle activity corresponds to a more pleasant emotional experience (Andreassi, 1995), although some studies showed no reliable effects of valence on activity over the ZM muscle (Larsen, Norris, & Cacioppo, 2003).

To sum up, these findings bring support to the hypothesis that approaching tonal sound sources are more biologically salient than receding ones. They exert a more intense emotional response in listeners, as measured at behavioral, subjective and physiological levels. However, it is difficult to extrapolate these results to a biological context and draw some firm conclusions regarding the greater biological salience of approaching auditory sources, due to the fact that the sounds tested in Experiment 1 were artificial simple tones. With these results, it remains unclear whether the physical properties of the sound (i.e. intensity level) or the significance attributed to it (i.e. a possible threat) stand for the approaching/receding asymmetry observed for the artificial tone. Hence, we conducted Experiment 2 in order to contrast physical and semantic effects and further explore our results in a more plausible biological context.

Experiment 2

In Experiment 2 we investigated the emotional influence of ecological (or natural) sounds perceived as approaching or receding natural sources. The stationary versions of these natural sounds differed in the emotional responses that they evoke in listeners (valence and arousal values). Experiment 2 contained both original versions of the sounds and a synthesized version of them, which was ‘meaning neutralized’, i.e. source identifiability was changed but the physical parameters across the sound were left intact. These ‘meaning neutralized’ versions were included in order to test whether emotional responses evoked by the sounds are dominated by the meaning and significance that listeners attribute to them or by their physical properties. The aim of this experiment was to investigate whether the effects observed in Experiment 1 for tones prevail for different types of natural sounds, which differ in their attributed emotional significance. It should be noted that previous research has found that the approach-receding disparity is limited to tones or tone-like sounds (Neuhoff, 1998). Hence, expanding the paradigm to include ecologically relevant sounds is an important step for the generalizing of this effect. In Experiment 1, approaching tones elicited more intense emotional responses in listeners than receding ones. If the greater biological salience of approaching sounds over receding sounds is mainly due to physical properties (i.e. intensity level), then the approaching/receding disparity should be prevalent for all natural sounds (*Hypothesis 1*). However, if there exist semantic effects, this is, the meaning that people attribute to the sound sources does matter, then the emotional responses should vary depending on the significance conveyed by the sounds *per se*. In other words, if the effects observed in Experiment 1 are due to approaching sounds being categorized as threat-related then unpleasant approaching sounds should elicit the strongest emotional responses (*Hypothesis 2*). In addition, if such physical/semantic interaction exists, then there should also be a clear difference between the effects observed between the original sounds and their ‘meaning neutralized’ versions (*Hypothesis 3*). In Experiment 2a we investigated the emotional effects of sounds by observing the behavior of participants on a subsequent emotion-related task, and in Experiment 2b we explored the effects of sounds on subjective emotional experience.

Experiment 2a

Methods

Participants. 15 participants took part (8 females; mean age 29 years; age range: 23-54 years; 8 had also participated in Experiment 1). Participants had normal hearing and were naïve as to the purposes of the study. They were paid for their time and gave their informed consent prior to the inclusion in the study. The experiments were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, materials and data collection. The apparatus and visual materials resembled those in the behavioral studies in Experiment 1a & 1b, but with different sounds. The auditory stimuli consisted of seven different sounds represented by two different versions of, rising or falling intensity level. These 7 sounds were: An artificial 'tone' (similar to that used in Experiment 1; used as a reference sound) and three natural sounds, 'laughing baby', 'growling dog' and 'footsteps', in their original and 'meaning neutralized' versions. The 'laughing baby' and the 'growling dog' sounds were chosen to have similar waveforms but were representing a pleasant and an unpleasant sound source. They were chosen according to the emotional ratings given for similar sounds in the International affective digitized sounds (IADS) database (Bradley & Lang, 1999). In this database, for the sound called 'DogGrowl' the normative valence and arousal ratings (in a 9-point scale) are 2.73 ± 1.92 and 7.77 ± 1.71, respectively, and for the sound called 'BabyLaugh' these are 7.92 ± 1.55 and 6.04 ± 2.08. The 'footsteps' sound was chosen to be a ‘neutral’ sound source (a similar sound in the IADS called ‘Walking’ has normative valence and arousal ratings 4.15 ± 1.28 and 5.43 ± 1.9, respectively). The intensity level varied approximately from 68 to 86 dBA, as measured at the participant’s ear position. All natural sounds (i.e. all sounds except the ‘tone’) were equalized in loudness. The duration of the intensity ramp was fixed to 2 s. All stimuli were preceded and followed by a 300 ms constant intensity tone (Maier & Ghazanfar, 2007), thus resulting in stimuli with a total duration of 2.6 s. A 10 ms onset/offset ramp was applied to the auditory stimuli to prevent clipping.

The ‘meaning neutralized’ versions of the natural sounds were created by a neutralization procedure that leaves the physical parameters intact across the sound but changes source identifiability. This neutralization procedure is based on spectral broadening procedure. In this procedure, the original sound is divided into windows of arbitrary length, which are transformed into frequency domain via FFT. Then the amplitudes are averaged within frequency bands of arbitrary width, and, the modified sound is converted back into time domain again by an inverse FFT. In the present research, 4096 sample-long Hanning-windows with 75% overlapping were used. The averaging in the frequency domain was done within 1/3 octave bands. As a result of this procedure, not only the temporal envelope of the original sound was kept intact, but also the loudness-time functions. However, due to the spectral broadening, the identifiability of sounds was substantially degraded.

A game pad held in the participant’s hands was used to collect participant’s behavioral data. Presentation® software (Version 9.90) was used to control stimuli delivery and record responses.

Design. The experimental trials consisted of a pair of stimuli formed by a sound followed by a photograph. The photograph was 'positive' (not analyzed), 'neutral' or 'negative'. There were 7 different sounds, in both 'approaching' and 'receding' versions. This resulted in 28 possible conditions with a 7 x 2 x 2 factorial design (Sound Type x Sound Direction [approaching, receding] x Picture Emotional Valence [negative, neutral]).

Procedure. The procedure for this experiment resembled the one followed in Experiment 1. Each experimental condition was presented 30 times, resulting in 840 trials. These trials were presented randomized in 6 experimental blocks, with 140 trials each plus 6 extra trials with positive photographs (not considered in the subsequent analysis). Each block of trials lasted on average for 15 minutes. The experiment was completed in two different experimental sessions. In each session, participants completed 3 experimental blocks, with a short break between blocks. During these blocks, RTs for each trial were collected. In addition, in the beginning and the end of each experimental session, participants were asked to listen to the sounds, starting by the 'meaning neutralized' sounds, and try to identify them (first and last identification block).

Data analyses. RTs were treated as in Experiment 1. The results from the identification blocks were inspected in order to know whether participants were able to identify the sounds, and whether this identification was dependent on the repeated exposure to the sounds.

Results

Sound identification. Results from the identification block showed that natural sounds could be effectively recognized. Both Experiment 2a and 2b are considered here since the same sounds were used in both experiments. The 'growling dog' sound was described as a dog, tiger, lion or wolf; participants used the adjectives growling, angry, mad, aggressive and close in their description. The ‘laughing baby’ sound was described as a baby or child, happy, unhappy, laughing, crying or talking. The ‘footsteps’ sound was described as footsteps or walking. The identification of the ‘meaning neutralized’ sounds varied with exposure to the sound, with few participants recognizing the real source of the sound. The ‘neutralized growling dog’ sound was described as breath, lion, air vent, something artificial and scary, dangerous dog, elevator shaft, echo, dragon, tiger, train, heavy machinery, and a steam engine. In the last identification block, about half of the participants identified the sound as a “filtered”, “manipulated” or “diverted” dog sound. The ‘neutralized laughing baby’ sound was at first described as strong wind, baby, lion, cat, baby in a box, baby monster, diverted child, monster, kindergarten or baby scream; in the last identification block many participants recognized the sound as some sort of baby sound. Finally, the ‘neutralized footsteps’ sound was the most difficult to identify for participants, who described it as a bag full of sand in movement, an old train starting, brushing of snow on a car, scrubbing, a horse, an engine, a whip, sweeping, a fire extinguisher, a helicopter, grass clipping or a monster. In the last identification block only two participants identified the sound as “computerized” footsteps.

Effects on behavior. On average, a total of 94 ± 3.2 percent (± SD indicated) of the trials from each participant (range: 87.2-98.6) were included in the data analyses of the RTs. The within-participants factors for the ANOVA were ‘sound type’, ‘sound direction’ (approaching or receding) and ‘picture emotional valence’ (negative or neutral). The results showed an observable trend for an interaction between ‘sound type’ and ‘sound direction’ ($F(2.4,34) = 2.5$; $p = .085$), which became significant when analyzing only the RTs to negative photographs

($F(2.8,40) = 3.1$; $p = .039$). Follow-up planned comparisons (paired *t*-test) with only the conditions with negative photographs (see Fig. 4A) revealed that the difference in RTs to photographs presented after approaching vs. receding sounds was significant for the 'tone' ($t(14) = 2.4$; $p = .03$) and the 'meaning neutralized growling dog' sounds ($t(14) = 2.25$; $p = .041$).

We performed a further ANOVA for within-participants factors 'neutralization' (neutralized or non-neutralized), 'sound type' ('laughing baby', 'growling dog' or 'footsteps'), 'sound direction' (approaching or receding) and 'picture emotional valence' (negative or neutral). The results showed an observable trend towards faster RTs for photographs following the 'meaning neutralized' versions of the sounds ($F(1,14) = 3.9$; $p = .068$), and a significant interaction between the factors 'neutralization', 'sound type' and 'sound direction' ($F(1.5,20.5) = 4.8$; $p = .029$). The RTs to photographs following the approaching 'laughing baby' sound were slightly faster than for those following the receding sound, but this pattern was reversed for the 'meaning neutralized' version of the sound. For the other types of sounds, the approaching 'meaning neutralized' sounds led to faster RTs than the receding sounds.

Experiment 2b

Methods

Participants. 12 participants took part (3 females; mean age 27 years; age range: 22-33 years; 4 had also participated in Experiment 1). Participants had normal hearing and were naïve as to the purposes of the study. They were paid for their time and gave their informed consent prior to the inclusion in the study. The experiments were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, materials and data collection. The experimental set-up was similar to the one described for Experiment 2a. Experiment 2b contained all auditory stimuli from Experiment 2a but did not include visual stimuli. A keyboard was used for registering participant's self-reported emotional responses.

Design. The same factorial design used in Experiment 2a, excluding the factor 'picture emotional valence', was used (i.e. 7 Sound Type x 2 Sound Direction).

Procedure. Participants were required to listen to the sounds and, after each sound, rate their feelings towards that sound using the 9-point valence and arousal pictorial scales of the Self-Assessment Manikin (SAM, Lang, 1980). Besides, for those conditions containing approaching sounds, participants were asked to *"imagine that you hear this sound in everyday life. How likely is it that this sound will be followed by a significant event?"* (Salience), and *"how likely is it that this sound will be followed by a threatening event?"* (Threat; these questions were adopted from a previous study reported in Neuhoff, 2008). An experimental block containing all sound conditions was repeated four times, with the possibility of taking a short break in half of the experiment. Participants completed identification blocks in the beginning and the end of each experimental session.

Data analyses. Self-reported valence and arousal for sounds were treated as in Experiment 1. Self-reported salience and threat for each sound were used as dependent variables for two separate ANOVAs. The results from the identification blocks were inspected in order to know whether participants were able to identify the sounds, and whether this identification was dependent on the repeated exposure to the sounds (see the results of Experiment 2a).

Results: Effects on self-report

Self-reported valence and arousal ratings for the emotional responses to the different sounds were used as dependent variables for a MANOVA containing as within-participants factors 'sound type' and 'sound direction'. The results (see Fig. 4B) revealed that there was a significant main effect of both factors, 'sound type' ($F(12, 106) = 6.3$; $p < .001$, $\Lambda = .34$) and 'sound direction' ($F(2, 8) = 4.8$; $p = .044$, $\Lambda = .46$), and a significant interaction between the factors ($F(12, 106) = 9.5$; $p < .001$, $\Lambda = .23$). A further ANOVA was performed for within-participants factors 'neutralization' (neutralized or non-neutralized), 'sound type' ('laughing baby', 'growling dog' or 'footsteps') and 'sound direction'. Apart from the significant effects observed with the previous MANOVA, the results revealed a significant interaction between the 'neutralization' factors and each of the other two factors, 'sound type' ($F(4,34) = 6.2$; $p < .001$, $\Lambda = .34$) and 'sound direction' ($F(2,8) = 18$; $p < .001$, $\Lambda = .18$), and the interaction of all three factors ($F(4,34) = 11.5$; $p < .001$, $\Lambda = .18$). Approaching sounds tended to be judged as more arousing and unpleasant than receding sounds in the case of the 'tone', the 'growling dog' sound and in the neutralized versions of the 'growling dog' and 'footsteps' sounds. On the other hand, the approaching 'laughing baby' sound was judged as more pleasant and less arousing than its receding version. For the ‘footsteps’ and neutralized ‘laughing baby’ sounds, the difference in the ratings for approaching and receding sounds was small, insignificant.

Two separate ANOVAs were performed for the ratings for the perceived salience and threat of the approaching sounds. The within-participants factor was ‘sound type’. The results revealed that there was a significant main effect of this factor both for the salience ($F(2.6,28) = 3.5$; $p = .035$) and threat ($F(2.4,26) = 9.1$; $p < .001$) ratings. The ‘growling dog’ sound was judged as the most salient and threatening sound (mean value ± *SEM* was 5.9 ± 0.6 for salience; and 6.6 ± 0.6 for threat), followed by the ‘tone’ (5 ± 0.6 for salience; and 4.5 ± 0.6 for threat), ‘footsteps’ (4 ± 0.6 for salience; and 3.2 ± 0.5 for threat), and ‘laughing baby’ (3.8 ± 0.7 for salience; and 2.3 ± 0.6 for threat) sounds. Two further ANOVAs were performed for the salience and threat ratings, with within-participants factors ‘neutralization’ (neutralized or non-neutralized) and ‘sound type’ (‘laughing baby’, ‘growling dog’ or ‘footsteps’). Apart from the significant effect of sound type observed with the previous ANOVA, the results revealed a significant effect of the factor ‘neutralization’ ($F(1,11) = 5.5$; $p = .039$ for salience; and $F(1,11) = 6$; $p = .033$ for threat), and a significant interaction between the ‘neutralization’ and ‘sound type’ factor for the threat ratings ($F(1.5,17) = 9.6$; $p = .003$). The neutralized versions of the sounds were judged as being more salient and threatening than the original non-neutralized versions, except for the threat rating for the ‘growling dog’ sound, which did not varied much with neutralization. For the neutralized sounds the mean salience and threat values (± *SEM*) were respectively: 6.1 ± 0.5 and 6.5 ± 0.6 for neutralized ‘growling dog’; 5 ± 0.5 and 4.7 ± 0.6 for neutralized ‘footsteps’; and 4 ± 0.5 and 4 ± 0.6 for neutralized ‘laughing baby’.

[Insert Figure 4 about here]

Discussion

The aim of Experiment 2 was to explore the emotional influence of approaching and receding natural sounds, and to compare these effects to those observed for artificial sounds in Experiment 1 where the same experimental methodology was applied. The natural sounds differed in the emotional significance that listeners may attribute to them. If the results from Experiment 1 reflect a survival-related adaptation to prioritize approaching sources, then we expected that those results would hold for the case of natural sound sources. Therefore, our *Hypothesis 1* was that all approaching natural sound sources would elicit more intense emotional reactions in listeners than their receding counterparts. However, if there is an interaction between the physical and the semantic properties of the sound sources, we expected that the emotional preference for approaching vs. receding sound sources would differ for the different types of natural sources. In this case, we expected the approaching/receding disparity to be greater for unpleasant, threat-related sound sources (*Hypothesis 2*). The results of Experiment 2 showed that the unpleasant approaching sound ('growling dog'), as well as the artificial approaching sound ('tone'), were subjectively considered as eliciting more unpleasant and arousing emotional feelings, compared to their receding versions. This finding was corroborated by the behavioral results which showed that faster RTs were obtained for the photographs following the approaching vs. receding version of these two sounds. On the other hand, the emotional effects differed for the pleasant ('laughing baby') and neutral ('footsteps') sounds. For the pleasant sound, the approaching version of the sound was judged as evoking emotional feelings with higher positive valence and lower arousal than its receding version, while for the neutral sound there was little difference in emotional judgments between the approaching and receding versions. Moreover, no distinct effect was observed for these pleasant and neutral sounds on the RTs to the photographs.

Our results confirmed *Hypothesis 2*, since the emotional preference for approaching sounds was only shown for the unpleasant sound, which at the same time was judged as the most threatening and salient from all presented sounds. Interestingly, the artificial tonal sound fell into the same category of unpleasant sounds, with similar responses as those induced by the ‘growling dog’ sound (note that larger effects observed for the tone may be mostly caused by the higher loudness of this sound). Hence, the results opposed to *Hypothesis 1*, since the emotional bias for approaching sounds was not observed for all natural sources, but this bias seemed to be dependent on the significance attributed to the sound source itself. This finding obtained further support from the results for the ‘meaning neutralized’ versions of the sounds (*Hypothesis 3*): When the sound could not be identified any longer, the approaching/receding disparity was dependant on the threat and salience associated to the sound source *per se*. Indeed, the data from the neutralized sounds showed a significant correlation between the RT for negative pictures with the ratings of salience and threat for these sounds (Pearson correlation coefficients were $r = -0.95$ and $r = -0.98$ respectively, $p < .05$). No such effects were evident for the original versions of the natural sounds (see Fig. 4 and mean salience and threat values in the previous section).

General discussion

The results of the present study suggest that, under certain circumstances, sound sources perceived as approaching exert more intense emotional responses in listeners than sound sources perceived as receding. The emotional significance attributed to the auditory source itself, the loudness of the sound and the perceived duration of loudness change seem to be relevant factors in the approaching/receding disparity on the emotional response to moving sound sources.

One of the critical findings to have emerged from Experiment 1 was that tones rising in intensity level had a more intense emotional effect in listeners than tones falling in intensity level. This asymmetry in the emotional effect was reflected in an emotion-related behavioral task, self-reported emotional experience and physiological state of listeners. Importantly, these results allow unifying previous results in auditory research to emotional responses. They provide with evidence supporting the previously stated hypothesis that approaching tonal sounds have a greater biological salience than receding tonal sounds. This hypothesis was formulated in relation to recent results in psychophysical studies involving loudness change estimations and loudness estimations (Neuhoff, 1998; Neuhoff, 2001; Stecker & Hafter, 2000; Susini, McAdams, & Smith, 2007) or duration estimations (DiGiovanni & Schlauch, 2007; Grassi & Darwin, 2001; Schlauch, Ries, & DiGiovanni, 2001), as well as in neuropsychological and neurophysiological studies (Bach *et al.*, 2008; Maier, Chandrasekaran, & Ghazanfar, 2008; Maier & Ghazanfar, 2007; Romei, Murray, & Thut, 2008; Seifritz *et al.*, 2002).

In the present study, the emotion-related behavioral task involved measuring the response to negatively charged or neutral photographs presented after a tonal sound of either rising or a falling intensity. Interestingly, not only this task was performed quicker in the conditions involving negative photographs, but also the asymmetry in RTs to rising and falling tones was especially evident in these ‘negative’ conditions. Taking the assumption that sounds with rising/falling intensity level are generally perceived as auditory approaching/receding sources (Maier & Ghazanfar, 2007; Neuhoff, 2001; Rosenblum, Carello, & Pastore, 1987), these results are in line with those reported previously in a visual study (Muhlberger, Neumann, Wieser, & Pauli, 2008). In that study, unpleasant pictures perceived as approaching elicited more intense emotional responses than unpleasant receding pictures, while no such effect of picture motion was found for pleasant or neutral pictures. The greater power of unpleasant events over pleasant events, the so-called “negativity bias”, has been found to be a general rule in many psychological phenomena: Similar degrees of activation (in our case a sound rising in intensity) have stronger effects on the negative compared to the positive motivational system (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Cacioppo & Gardner, 1999).

Previous research has proposed different explanations for the existent asymmetry in responses to sounds with increasing vs. decreasing intensity level. For instance, for the observed loudness bias, some authors suggested that there might be a possible short-term auditory memory effect with global loudness judgments just based on the end intensity level of sounds (the so-called “recency effect”; Susini, McAdams, & Smith, 2007; Teghtsoonian, Teghtsoonian, & Canevet, 2005). On the other hand, other authors have pointed out that these effects might account for a greater biological salience (Guski, 1992; Maier & Ghazanfar, 2007; Neuhoff, 1998; Stecker & Hafter, 2000), or a greater intrinsic warning value (Bach *et al.*, 2008), of sounds with increasing intensity level. These authors refer to examples from a natural context: first, sounds with increasing intensity level may represent approaching sources, which may constitute a potential threat; and second, in animal communication intensity fluctuation is a main parameter used for alarm vocalization (Bradbury, 1998; Owren, Rendall, & Bachorowski, 2005). The results of our study bring support to the biological salience hypothesis from an embodied emotion perspective. Salient events lead to an increase of emotional arousal, as the one that our results showed in response to tonal sounds perceived as approaching. Moreover, the strongest activation observed for the conditions combining approaching tonal sounds and negative information suggests that the perception of approaching sound sources might be linked to the activation of a defensive behavior in listeners.

Our results also indicate that an increased intensity level of sound stimuli lead to more pronounced emotional responses and to a greater observed approaching/receding disparity, than lower intensity levels. The observed dependence of the effect on the intensity change is consistent with the biological salience hypothesis, since in a natural context loud sounds may represent close, big or strong sound sources (Herman & Ritter, 2004; Lenti Boero and Bottoni, 2008). Hence, loud sound sources may become associated to danger, especially in the case of sources that are approaching the listener. This greater biological salience of loud approaching sounds has been suggested previously when explaining looming bias (Neuhoff, 1998). On the other hand, we also found that a longer rise time of the sounds lead to more intense emotional responses than sounds with shorter rise-time. In a natural context, the rise time may represent the source velocity, with short rise-time corresponding to rapidly moving objects or high forces (Herman & Ritter, 2004; Lenti Boero and Bottoni, 2008). It should be noted that in our experiment we varied the intensity rise time of the sounds while keeping constant the range of intensity change. However, sound intensity perception is dependent on the signal duration (Small, Cox, & Brandt, 1962; Zwicker & Feldtkeller, 1967). Hence, our results may convey a loudness effect as the one described above. Future research might try to investigate the effect of rise time in this paradigm by ensuring that sounds do not vary significantly in loudness. An alternative explanation for the smaller effects induced by the short rise-time sounds might be that the longer the sounds are presented, the more time listeners have to evaluate the sound and prepare for response. It might be possible that there is a particular, ecologically motivated, range of sound object velocities related to the observed affective reactions (cf. U-shaped dependence of minimum audible angle on the sound source velocity; Perrott & Saberi, 1990).

Experiment 2 extended the results from Experiment 1 to the case of natural ecological sounds. Importantly, the findings from Experiment 2 showed that the directional sensitivity for approaching over receding sounds was only observed for sounds categorized as unpleasant, arousing, threatening and salient. The artificial tonal sound used in Experiment 1, that was included also in Experiment 2 as a reference sound, fell into the same category of unpleasant and arousing sounds. Sounds categorized as neutral showed little approaching/receding bias on self-report and on the behavioral task. This lack of looming bias in the behavioral task was also observed for the sound categorized as pleasant, which was judged as more positive and less arousing in its approaching, as compared to its receding, version. This finding is particularly relevant[2] as it pertains to previous work in which it was found that tonal sounds expose a greater looming bias than noise. Some authors (e.g., Neuhoff, 2001) speculated that this approach bias observed for tones (but not for noise) might be due to the fact that tonal sounds can act as markers of animate beings. In a natural environment noise is generally produced by inanimate sources (e.g., wind, running water), whereas most tonal sounds are produced by biological organisms. Neuhoff (2001) proposed that the auditory system is tuned to detect the more critical “biological” tonal sounds, what results in a greater approach bias for these sounds. However, the present results show that it is the emotional valence attributed to sounds what determines the approach bias: The artificial tonal sounds used in this and previous studies are quite unpleasant to listeners, and unpleasant sounds have a greater approach bias than pleasant sounds.

In conclusion, our results seem to indicate that approaching unpleasant sounds elicit more intense emotional responses in listeners than receding unpleasant sounds, while such effect of perceived auditory movement is not found for pleasant or neutral sound sources. Although it is still early to draw firm conclusions, given that this study contained a few examples of sounds and only one sound example of each class, our findings seem consistent with those previously described for the visual modality (Muhlberger, Neumann, Wieser, & Pauli, 2008). This might indicate that a more general, amodal mechanism is involved in the observed looming bias. The results may be also in line with the embodiment and motivational approaches to emotion (Cacioppo & Gardner, 1999; Lang, Bradley, & Cuthbert, 1990; Niedenthal, *et al.*, 2005; Niedenthal, 2007). They might be the outcome of an existing survival mechanism: Sound sources that are rapidly approaching one, invading the margin of safety surrounding the body, need to be identified quickly in order to determine their threat value. Therefore, sound sources identified as threatening or unidentified may activate a defensive response.

Future research on auditory motion perception should extend these results by overcoming some limitations of the present study. First, in this study auditory motion was simulated by increasing/decreasing the intensity level of the sounds. It might be interesting to test our hypotheses with virtual sound sources including other motion cues (e.g. the Doppler shift or binaural cues). Second, combining auditory stimuli with moving visual objects also can provide a fruitful test-bed for refining our results. Such multisensory stimuli should disambiguate parameters of speed and distance, and serve as a tool for controlling the surrounding context and representation of looming objects. Third, it might be worth to investigate the effect of the velocity of the sound source by fully controlling other parameters, such as loudness of the sound (e.g. using other auditory cues like the Doppler shift, or, based on audio-visual simulations). Finally, and most importantly, future studies should include other auditory objects covering a full range of emotion eliciting capabilities, in order to get a better understanding of the dependence of the emotional response to approaching vs. receding sounds on their source nature.

References

Alais, D., & Burr, D. (2003). The "flash-lag" effect occurs in audition and cross-modally. *Current Biology, 13*(1), 59-63.

Andreassi, J. L. (1995). *Psychophysiology: Human behavior and physiological response* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bach, D. R., Neuhoff, J. G., Perrig, W., Seifritz, E. [published online ahead of print July 15, 2009]. Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology*. doi:10.1016/j.ijpsycho.2009.06.004.

Bach, D. R., Schachinger, H., Neuhoff, J. G., Esposito, F., Di Salle, F., Lehmann, C., *et al.* (2008). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex, 18*(1), 145-150.

Baldo, M. V. C., & Klein, S. A. (1995). Extrapolation or attention shift. *Nature, 378*(6557), 565-566.

Barrett, L. F., Niedenthal, P. M., & Winkielman, P. (2005). *Emotion and consciousness*. New York: Guilford Press.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370.

Boucsein, W. (1992). *Electrodermal activity*. New York: Plenum press.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication*. Sunderlan, MA: Sinauer.

Bradley, B. P., Mogg, K., & Millar, N. H. (2000). Covert and overt orienting of attention to emotional faces in anxiety. *Cognition & Emotion, 14*(6), 789 - 808.

Bradley, M. M., & Lang, P. J. (1999). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings* (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling and physiology. In R. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 242–276). New York: Oxford University Press.

Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. *Annu Rev Psychol, 50*, 191-214.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness* (1st ed.). New York: Harcourt Brace.

DiGiovanni, J. J., & Schlauch, R. S. (2007). Mechanisms responsible for differences in perceived duration for rising-intensity and falling-intensity sounds. *Ecological Psychology, 19*(3), 239 - 264.

Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology, 27*(5), 481-494.

Ghazanfar, A. A., Neuhoff, J. G., & Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. In *Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15755-15757.

Gordon, M. S., & Rosenblum, L. D. (2005). Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. *Perception & Psychophysics, 67*(4), 580-594.

Grassi, M., & Darwin, C. J. (2001). Perception of the duration of ramped and damped sounds with raised cosine ramps. In *Proceedings of the Sixteenth Annual Meeting of the International Society for Psychophysics*, Strasbourg, France.

Graziano, M. S. A. (2001). A system of multimodal areas in the primate brain. *Neuron, 29*(1), 4-6.

Guski, R. (1992). Acoustic Tau: An easy analogue to visual Tau? *Ecological Psychology, 4*(3), 189 - 197.

Hermann, T., & Ritter, H. (2004). Sound and meaning in auditory data display. In *Proceedings of the IEEE, 92*(4), 730-741.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behavioral & Brain Sciences, 31*(5), 559-575; discussion 575-621.

Kitagawa, N., & Spence, C. (2005). Investigating the effect of a transparent barrier on the crossmodal congruency effect. *Experimental Brain Research, 161*(1), 62-71.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, Johnson, J.H., Williams, T.A. (Ed.), *Technology in mental health care delivery systems* (pp. 119-137). Norwood, NJ.: Ablex Publishing.

Lang, P. J. (1995). The emotion probe. Studies of motivation and attention. *American Psychologist, 50*(5), 372-385.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*(3), 377-395.

Lang, P. J., Bradley, M.M., Cuthbert, B.N. (1997). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Tech. Rep. No. A-8). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology, 40*(5), 776-785.

LeDoux, J. (1996). Emotional networks and motor control: A fearful view. *Progress in Brain Research, 107*, 437-446.

Lenti Boero, D., & Bottoni, L. (2008). Why we experience musical emotions: Intrinsic musicality in an evolutionary perspective. *Behavioral and Brain Sciences, 31*(05), 585-586.

Levenson, R. W. (1994). Human emotions: A functional view. In P. Ekman, Davidson, R.J. (Ed.), *The nature of emotion: Fundamental questions* (pp. 123-126). New York: Oxford University Press.

Lutfi, R. A., & Wang, W. (1999). Correlated analysis of acoustic cues for the discrimination of auditory motion. *Journal of Acoustic Society of America*, *106*(2), 919–928.

Mackay, D. M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature, 181*(4607), 507-508.

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology, 18*(13), 963-968.

Maier, J. X., & Ghazanfar, A. A. (2007). Looming biases in monkey auditory cortex. *Journal of Neuroscience, 27*(15), 4093-4100.

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by Rhesus monkeys. *Neuron, 43*(2), 177-181.

Muhlberger, A., Neumann, R., Wieser, M. J., & Pauli, P. (2008). The impact of changes in spatial distance on emotional responses. *Emotion, 8*(2), 192-198.

Neuhoff, J. (2008). *'Behavioral Relevance' as a guiding principle in investigating the perception of sound source properties.* Oral presentation at Acoustics'08, Paris, France.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature, 395*(6698), 123-124.

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13*(2), 87 - 110.

Neuhoff, J. G., Planisek, R., & Seifritz, E. (2009). Adaptive sex differences in auditory motion perception: Looming sounds are special. *Journal of Experimental Psychology: Human Perception and Performance*. *35*, 225–234.

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*(5827), 1002-1005.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*(3), 184-211.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*(6487), 256-257.

Öhman, A. (1993). Fear and anxiety as emotional phenomena: Clinical phenomenology evolutionary perspectives, and information-processing mechanisms. In M. Lewis, Haviland, J.M. (Ed.), *Handbook of emotions* (pp. 511-536). New York: Guilford.

Owren, M. J., Rendall, D., & Bachorowski, J.-A. (2005). Conscious and unconscious emotion in nonlinguistic vocal communication. In P. N. L. F. Barrett, & P. Winkielman (Ed.), *Emotion and Consciousness* (pp. 185-204). New York: Guilford Publications.

Perrott, D. R., & Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *Journal of Acoustic Society of America, 87*(4), 1728-1731.

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.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science, 17*(4), 292-299.

Romei, V., Murray, M. M., & Thut, G. (2008). Looming sounds selectively and pre-consciously enhance visual cortex excitability. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 1*(3), 247.

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative Effectiveness of 3 Stimulus Variables for Locating a Moving Sound Source. *Perception, 16*(2), 175-186.

Rosenblum, L. D., Wuestefeld, A. P., & Saldana, H. M. (1993). Auditory Looming Perception - Influences on Anticipatory Judgments. *Perception, 22*(12), 1467-1482.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178.

Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". *Science, 136*, 982-983.

Schiff, W., & Oldak, R. (1990). Accuracy of Judging Time to Arrival - Effects of Modality, Trajectory, and Gender. *Journal of Experimental Psychology-Human Perception and Performance, 16*(2), 303-316.

Schlauch, R. S., Ries, D. T., & DiGiovanni, J. J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. *The Journal of the Acoustical Society of America, 109*(6), 2880-2887.

Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schachinger, H., *et al.* (2002). Neural processing of auditory looming in the human brain. *Current Biology, 12*(24), 2147-2151.

Small, A. M., Cox, P. G., & Brandt, J. F. (1962). Loudness as a Function of Signal Duration. *Journal of the Acoustical Society of America, 34*(4), 513-514.

Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *Journal of the Acoustical Society of America, 107*(6), 3358-3368.

Susini, P., McAdams, S., & Smith, B. K. (2007). Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica United with Acustica, 93*(4), 623-631.

Teghtsoonian, R., Teghtsoonian, M., & Canevet, G. (2005). Sweep-induced acceleration in loudness change and the "bias for rising intensities". *Perception & Psychophysics, 67*(4), 699-712.

Zwicker, E., & Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger* (2., neubearb. Aufl. Mit 217 Bildern. ed.). Stuttgart: Hirzel.

Footnotes

[1] Upon conducting these studies we were unaware of any evidence for emotional effects of approaching sounds. However, work that was done parallel in time with our studies show similar results (Bach, Neuhoff, Perrig, & Seifritz, in press).

[2] We thank an anonymous reviewer for drawing more attention to this fact.

Figure 1. Time course of the trials. Visual (upper) and auditory (lower) stimuli presented. In each trial, experimental stimuli were preceded by a countdown from 5 to 1. Auditory stimuli consisted of 1 kHz tones rising (approaching sound) or falling (receding sound) in intensity, preceded and followed by a 300 ms constant-intensity period. In Experiment 1, auditory stimuli were from the "loud" intensity range, and T varied between 1, 2 or 3 s. In Experiment 2, auditory stimuli were from both "loud" and "soft" intensity ranges, and T was fixed to 2 s. Visual stimuli used in the three-alternative forced-choice task were taken from the International Affective Picture System (IAPS) and fell into one of three categories: "negative", "neutral" or "positive".

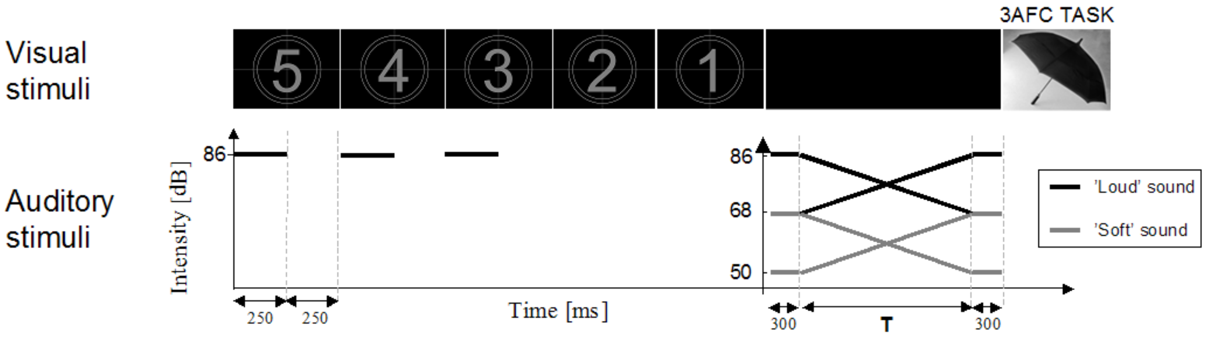

Figure 2. Results of Experiment 1. (A) Mean reaction times (in ms) ± *SEM* for the conditions in Experiment 1a (left panel), in which the sounds were ‘loud’ and differed in their intensity ramp duration, and in Experiment 1b (right panel), in which the sounds had 2s-ramp duration and differed in intensity range. (B) Mean valence and arousal ratings (in a 9-point scale) ± *SEM* for all sound conditions in Experiment 1c.

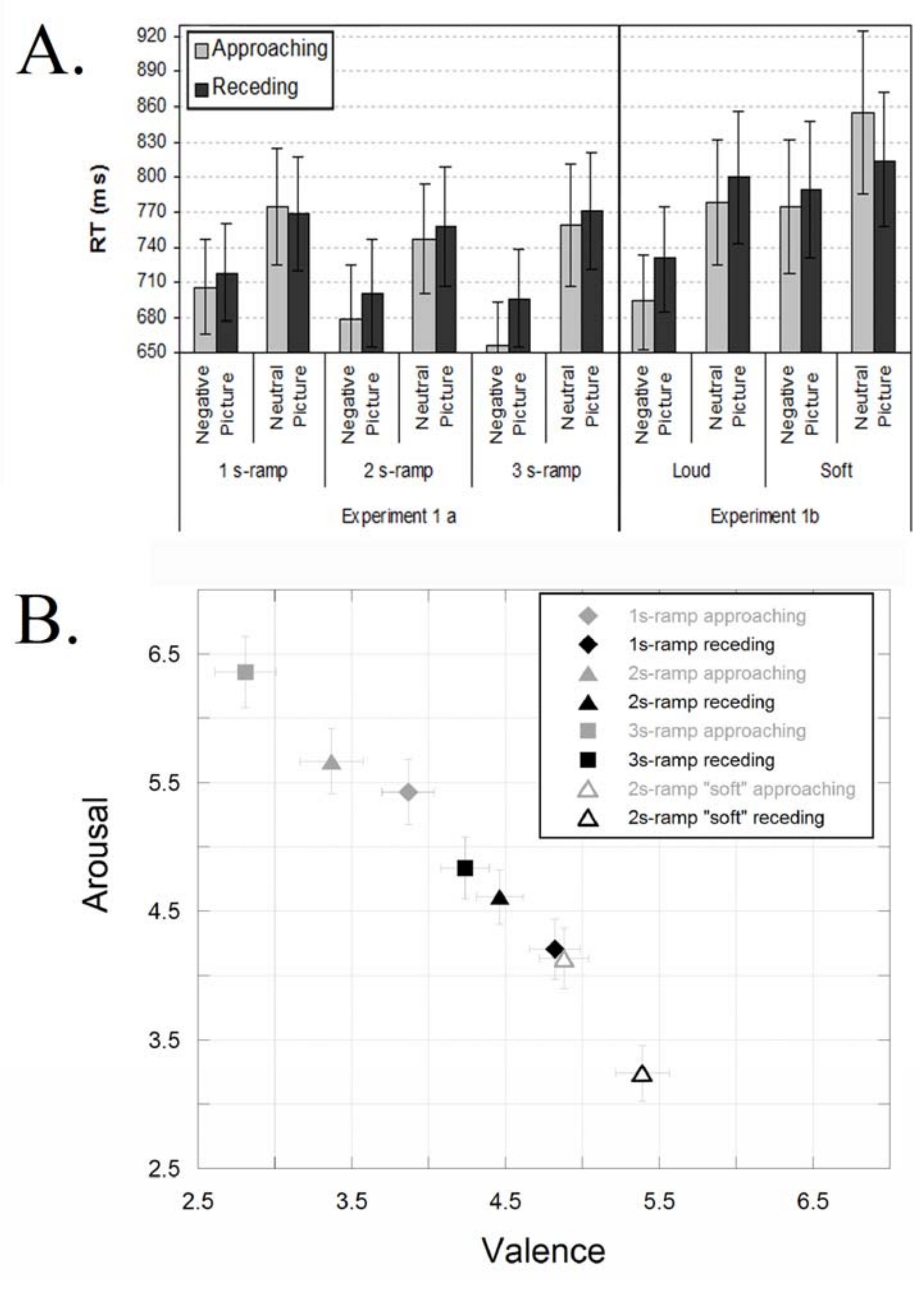

Figure 3. Results of Experiment 1. Physiological response (Facial electromyography (corrugator supercilii - CS, and zygomatic major - ZM) and electrodermal activity (EDA), mean ± *SEM*, for each 1-second interval for the 6 seconds following sound onset in Experiment 1c. CS and ZM change z-scores were calculated by subtracting the average response for each 1-second interval from the mean activity during the 1 s preceding sound onset (baseline). EDA change z-scores were calculated by subtracting the maximum response for each 1-second interval from the maximum response during the baseline.

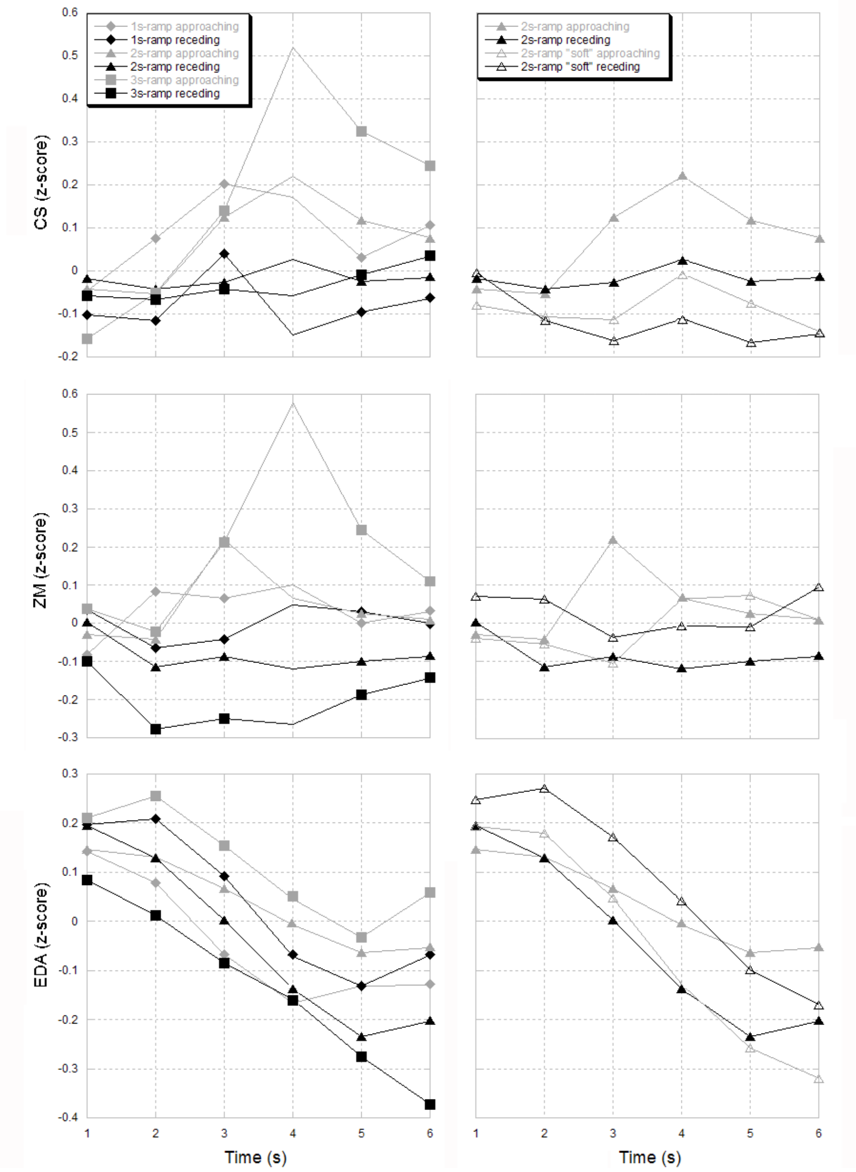

Figure 4. Results of Experiment 2. (A) Mean reaction times (in ms) ± *SEM* for the all the conditions in Experiment 2a which contained negative pictures. (B) Mean valence and arousal ratings (in a 9-point scale) ± *SEM* for all sound conditions in Experiment 2b.