Emotional bias for the perception of rising tones

Ana Tajadura-Jiménez\textsuperscript{a}, Aleksander Väljämäe\textsuperscript{b} and Daniel Vastfjall\textsuperscript{c}

\textsuperscript{a}Chalmers University of Technology, Division of Applied Acoustics, Sven Hultinsgata 8a, 41296 Gothenburg, Sweden
\textsuperscript{b}Laboratory for Synthetic Perceptive, Emotive and Cognitive Systems (SPECS) - Universitat Pompeu Fabra, Ocata 1, 08003 Barcelona, Spain
\textsuperscript{c}Chalmers University of Technology, Division of Applied Acoustics - Chalmers Room Acoustics Group, Sven Hultins gata 8a, 41296 Gothenburg, Sweden
ana.tajadura@chalmers.se
1 Introduction

Our survival directly depends on the abilities of detecting approaching objects, determining their potential threat and predicting their possible impact on our body. Hence, a perceptual priority towards approaching objects would provide with a significant survival advantage, since it would increase the time and the attentional resources available to be able to avoid these objects. Indeed, ample body of research has shown that animals have evolved to process dynamic information in a different way than static information [1, 2, 3, 4]. For instance, a brief flash presented physically aligned with a moving object appears to lag behind the moving object (i.e. the ‘flash-lag effect’ [5]). This effect is not restricted to vision since a similar effect has been found in audition with a brief tone spatially aligned with a moving sound source [6]. Furthermore, there is evidence that people are able to anticipate the time to contact of both visual and auditory approaching sources [2, 7, 8, 9, 10].

Accordingly, studies on auditory moving sources have shown the existence of a bias at perceptual level in responding to approaching (or looming) versus receding sounds. It exists an asymmetry in loudness change with higher loudness estimations for the looming versus receding sounds [2, 11, 12]. Similarly, the loudness of sounds increasing in intensity tends to be higher that the one of stationary sounds with the same level (c.f. [13]). This directional preference for looming versus receding and stationary sounds seems to exist also at neural level, as found both in animal (e.g. [14]) and human research [15, 16]. Although it has been suggested that it might exist a short-term auditory memory effect with global loudness judgments just based on the end level of sounds (the so-called “recency effect” [13, 17]), some authors have pointed out that these effects might account for a greater biological salience of approaching sounds [11, 14, 15, 18]. In the present study, we aimed at exploring whether models of attention and emotion could bring further support to the latter hypothesis.

Salient events have the capability of evoking emotional responses. These emotional responses to events often elicit an automatic attentional switch towards these events, thus modulating subsequent perceptual processes [19, 20, 21]. 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 than emotional stimuli (e.g., an angry face in [22]). We investigated the effect of looming and receding sounds on listeners by using direct measures of listeners’ emotional state and behavioural measures on subsequent attentional and perceptual processes. We hypothesized that a greater salience would evoke a greater increase in listeners’ emotional arousal which would capture and hold attention.

In particular, in Experiment 1 and 2 we investigated the effect of listening to looming or receding tones, with different intensity ranges and periods of intensity change on the reaction times to a subsequent behavioral task. In Experiment 3, physiological and self-reported emotional ratings for the different sounds were collected. In the present study we adopted a dimensional approach to emotions [23, 24] where emotions are characterized in terms of two continuous dimensions valence or pleasantness (positive versus negative) and arousal or activation (excited versus calm).

2 Experiments 1 and 2: Effects on behavior

2.1 Methods

Participants. In Experiment 1 twelve participants (mean age 26 years; age range from 22 to 34 years; 6 females) took part. In Experiment sixteen participants (mean age 25 years; age range from 21 to 43 years; 7 females) took part. In all three experiments reported here, participants had normal hearing and were naïve as to the purposes of the study. They gave their informed consent prior to the experiments and were paid for their participation. The experiments were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus and materials. The experiment was conducted in a dark sound-attenuated room one single participant was seating. Auditory stimuli were delivered via headphones. Visual stimuli were presented on a LCD-screen placed at a distance of approximately 0.5 meters from the participant. Image resolution was 1024x768 pixels and the field of view for the images was 37°×30°. A game pad was used to collect participants’ data. Presentation® software (Version 9.90, www.neurobs.com) was used to control stimuli delivery and response recordings.
In Experiment 1, auditory stimuli consisted of 1 kHz tones (44.1 kHz sampling rate) with rising or falling intensity range from 68 to 86 dB(A) (as measured at the participants' ear position). The time for the intensity change varied between three possible values (1 s, 2 s or 3 s). In Experiment 2, the time for intensity change was fixed to 2 s, and there were two possible ranges for intensity change, ‘loud’ (68-86 dB(A)) and ‘soft’ (50-68 dB(A)). Critically, ‘soft’ sounds had lower intensity than ‘loud’ sounds at any point (as in Maier and Ghazanfar’s study). All stimuli were preceded and followed by a 300 ms constant intensity tone, thus resulting in stimuli with a total duration of 1.6, 2.6 or 3.6 s (as the sounds used in Maier and Ghazanfar, 2007). An onset/offset ramp of 10 ms was applied to all the auditory stimuli.

Visual stimuli consisted of photographs from the International Affective Picture System (IAPS [25]). The IAPS is a set of normative emotional pictures rated in a 9-point scale for valence and arousal. Stimuli were chosen according to their arousal and valence values to form two groups of 15 ‘neutral’ (around 5 points in valence scale and low arousal) and 15 ‘negative’ pictures (negative valence and moderate arousal). Neutral pictures depicted mushrooms or home objects, while negative ones showed ruins, dirtiness, insects, accidents or guns. In addition, six positive stimuli were included as a contrast to negative pictures, in order to avoid a bias response towards ‘neutral’. Responses to the trials with positive pictures were not included in the analysis of the results. Finally, six extra pictures were used to instruct participants in their tasks for the experiment.

The different stimulus conditions were randomly presented, with an intertrial interval of 3500 ms. During the last 2500 ms of this intertrial 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 0.5 second, and numbers from 5 to 3 were accompanied by a 250-ms 1 kHz tones with an intensity of 86 dB(A).

**Design.** Participants were exposed to pairs of stimuli formed by a tone and a photograph. Photographs were from the “negative” or “neutral” group. The tones were rising or falling in intensity and are referred to as ‘looming’ or ‘receding’ tones later in the text, since research has shown that auditory looming perception mostly relies on the intensity change of the auditory signals [2, 7, 14], at least for distances between the moving source and listener larger than 2 m [9].

In Experiment 1, the time for the intensity change of the tones varied between three possible values (1 s, 2 s or 3 s). This resulted in twelve possible conditions with the following factorial design: 2 sound directions (looming or receding) x 3 periods of intensity change (1 s, 2 s or 3 s) x 2 picture emotional valence (negative or neutral).

In Experiment 2, ‘looming’ or ‘receding’ tones with both ‘loud’ and ‘soft’ versions were used. This resulted in eight possible conditions with the following factorial design: 2 sound directions (looming or receding) x 2 intensity range (loud or soft) x 2 picture emotional valence (negative or neutral).

**Procedure.** Participants arrived individually to the laboratory. After receiving written and verbal instructions, they sat and the headphones were positioned. Both experiments were divided into two experimental blocks, each of them with trials containing pairs of stimuli, formed by a tone followed by a photograph. A practice block, with 6 trials, was completed before the experiment to familiarize participants with the paradigm.

Participants were required to make 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). They were instructed to emphasize speed, but to refrain from anticipatory and inaccurate responses. The response was made by pressing one out of three buttons in the gamepad and reaction times (RTs) were collected. These blocks contained 15 repetitions of each experimental condition plus 6 extra trials with positive pictures (which were not considered in the subsequent analysis), making a total of 186 trials per block for Experiment 1 and 126 trials per block for Experiment 2, which took about 15 minutes to complete. Participants had a short break between blocks.

### 2.2 Results

During the experiment, reaction times (RTs) for each trial were collected. RTs exceeding ±3 standard deviations from the mean RT for each participant were recursively discarded. On average, 92.6 ± 2.6 percent of the trials from each participant if Experiment 1 (range: 88.1-97.2) and 94 ± 3.4 percent of the trials from each participant in Experiment 2 (range: 87.1-97.5) were included in the analyses of the RTs. Data was subjected to repeated-measures ANOVAs where Greenhouse-Geisser correction was used for unequal variances.

In Experiment 1, the ANOVA contained as within-participant factors ‘sound direction’ (looming or receding), ‘period of intensity change’ (1 s, 2 s or 3 s) and ‘picture emotional valence’ (negative or neutral). Results showed (see Fig. 1a) that each factor had a significant effect. When considering ‘sound direction’, participants responded faster to pictures presented after a ‘looming’ versus a ‘receding’ sound ($F_{(1,11)} = 7.975; p = 0.017$). This difference between looming and receding sounds in RTs to pictures was much more evident for negative pictures than for neutral ones. The ‘period of intensity change’ had a significant effect ($F_{(1.9,21)} = 10.36; p < 0.001$), with participants responding faster to pictures when sound was presented for 2 or 3 s than when it was presented for 1 s. Bonferroni adjusted pairwise comparison revealed a significant difference between 1 s and 2 s periods ($p = 0.003$), between 1 s and 3 s periods ($p = 0.007$), but not between 2 s and 3 s periods ($p = 1$). Finally, when considering ‘picture emotional valence’, participants responding much faster (more than 70 ms) to negative than to neutral pictures ($F_{(1,11)} = 14.4; p = 0.003$).

In Experiment 2, the ANOVA contained as within-participant factors ‘sound direction’ (looming or receding), ‘intensity range’ (loud or soft) and ‘picture emotional valence’ (negative or neutral). Results showed (see Fig. 1b) that participants responded faster (58 ms) to pictures when sound was at the ‘loud’ intensity range ($F_{(1,15)} = 6.8; p = 0.02$). When considering ‘picture emotional valence’, participants responded faster (more than 65 ms) to negative than to neutral pictures ($F_{(1,15)} = 8.6; p = 0.01$). A significant interaction between direction and picture valence was found ($F_{(1,15)} = 6.1; p = 0.026$), which revealed that the effect of direction on RTs was more important for...
negative pictures. An ANOVA performed only in the conditions with negative photographs, with within-participant factors ‘sound direction’ (looming or receding), and ‘intensity range’ (loud or soft) revealed a significant effect for both factors, with faster responses after ‘loud’ sounds \((F(1, 15) = 6.4; \ p = 0.023)\) and after ‘looming’ sounds \((F(1, 15) = 5; \ p = 0.041)\).

All electrodes were filled with electrode paste and attached on the previously cleaned skin. EMG signals were sampled at a rate of 3125 Hz, amplified and bandpass filtered from 10 to 400 Hz (e.g. Andreassi, 2001). 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 for analysis (e.g., [28]).

The digital data collection was controlled by AcqKnowledge 3.8.1 software.

**Design.** Two factorial designs were used. On one hand, sounds with ‘loud’ intensity range were submitted to a design with 2 sound directions (looming or receding) x 3 periods of intensity change (1 s, 2 s or 3 s). On the other hand, sounds with a 2 s-period of intensity change were submitted to a design with 2 sound directions (looming or receding) x 2 intensity range (loud or soft).

**Procedure.** Participants arrived individually to the laboratory. After receiving written instructions, they sat and the electrodes were attached. During a rest period of about 5 min, participants received additional verbal instructions, headphones were positioned and a short practice block was completed to familiarize participants with the paradigm and test the physiological equipment. The experiment consisted of 6 blocks with the 8 different sound conditions each. There were two types of blocks. In blocks 1, 3 and 5 after each sound, valence and arousal ratings of participants’ feelings towards the sounds were collected by using the Self-Assessment Manikin (SAM [29]), a 9-point pictorial scale. In the rest of the blocks sounds were presented one after another with a silent interstimulus interval of 3 seconds. Electrodes were removed at the end of the experiment.

**3.2 Results**

**Effects on self-reported emotional experience.** Self-reported valence and arousal were used as dependent variables for two multivariate ANOVAs where Wilks’ Lambda was used as the multivariate criterion. The first ANOVA, performed on the results for the sounds with ‘loud’ intensity range, used as within-participant factors ‘sound direction’ and ‘period of intensity change’. There was a significant effect of both factors, sound direction \((F(2, 26) = 34.4; \ p < 0.001, \ \Lambda=.274)\) and period of intensity change \((F(4, 106) = 21.1; \ p < 0.001, \ \Lambda=.31)\), and a significant interaction between them \((F(8, 106) = 3.7; \ p = 0.007, \ \Lambda=.771)\). Approaching and longer sounds were perceived as more unpleasant and arousing (see Fig. 2-left panel).
The second ANOVA, performed on the results for the 2 s-duration sounds, used as within-participant factors ‘sound direction’ and ‘intensity range’. There was a significant effect of both factors, sound direction ($F_{(2, 26)} = 20.3; p < 0.001, \Lambda = .391$) and intensity range ($F_{(2, 26)} = 91.4; p < 0.001, \Lambda = .125$), and a significant interaction between them ($F_{(5, 65)} = 3.5; p = 0.046, \Lambda = .79$). Approaching and louder sounds were perceived as more unpleasant and arousing (see Fig. 2-right panel).

Effects on physiology  EDA and facial EMG recordings were individually inspected for possible artefacts. Data from the ZM was missed for 12 over 28 participants. First, sounds with ‘loud’ intensity range were subjected to repeated-measures ANOVAs with within-participant factors time interval (6 1-s intervals) x sound direction (approaching or receding) x period of intensity change (1 s, 2 s or 3 s). Time interval was included as a within-participant factor, only for the EMG signals, because physiological responses vary across time beginning at the onset of a stimulus (Cacioppo, Tassinary, & Berntson, 2000). ‘Sound direction’ had a significant effect on the CS ($F_{(1, 24)} = 5.2; p = 0.032$) and ZM muscle activities ($F_{(1,13)} = 4.9; p = 0.046$), with approaching sounds leading to bigger activity. The interaction between ‘sound direction’ and ‘period of intensity change’ showed a marginally significant effect on EDA ($F_{(1,2,32,9)} = 3.6; p = 0.058$), with a clearer asymmetry between approaching and receding sounds for longer sounds. Second, sounds with a 2 s-period of intensity change were submitted to repeated-measures ANOVAs with within-participant factors time interval (6 1-s intervals) x sound direction (approaching or receding) x intensity range (loud or soft). ‘Sound direction’ had a significant effect on the CS activity ($F_{(1, 26)} = 4.4; p = 0.046$) and showed a close to significance trend for EDA ($F_{(1, 27)} = 3; p = 0.095$) with approaching sounds leading to bigger activity. ‘Intensity range’ had a significant effect on the ZM activity ($F_{(1,14)} = 4.6; p = 0.049$) and EDA ($F_{(1, 26)} = 9.1; p = 0.006$). The interaction between ‘sound direction’ and ‘intensity range’ showed a significant effect on EDA ($F_{(1,27)} = 4.8; p = 0.038$), with a clearer asymmetry between approaching and receding sounds for louder sounds.

5 Conclusion

The results of this study suggest that tones rising in intensity elicit stronger emotional responses on listeners than tones falling in intensity, both being perceived as more arousing and more unpleasant. The emotional power of the sounds was dependent on the intensity range of the sounds and the period of intensity change, with bigger asymmetries between rising and falling tones for louder and longer sounds. This study combined three different methodologies to measure emotional reactions, subjective, behavioural and physiological measures, and the asymmetry in the effects caused by sounds perceived as looming versus receding was reflected at all the three different levels.

These findings bring further support to the hypothesis that approaching sounds have a greater biological salience than receding ones as previously suggested by behavioural and neurophysiological studies [11, 14, 15, 18]. An increase in emotional arousal caused by approaching sounds may have an ecological explanation. Emotional events often evoke a switch in attention towards these events [19, 20, 21], and in the case of approaching objects, that shift in attention would provide with a bigger chance to avoid a possible impact of these objects on our body. Moreover, in this study the louder approaching sounds, which in a natural environment could represent bigger or closer objects, elicited the biggest emotional responses, what again seems to provide with a significant survival advantage.

Nevertheless, further studies are required to test and reveal the nature of the responses to dynamic auditory stimuli. For instance, using stimuli other than tones may alter the observed effect, as previous research showed for broadband noise [11, 14]. Future research may investigate the response to approaching natural sounds or sounds that are more common in our everyday life, since they might be more critical from an ecological perspective. Even though the physical properties of sounds play a major role on the reactions induced, other variables related to subjective interpretation and meaning should be considered.

Acknowledgments

This work was carried out as part of the PRESENCCIA project, a European Union–funded Integrated Project under the IST Program (Project Number 27731).
References


