

The sound of arousal in music is context-dependent

CrossMark
click for updatesDaniel T. Blumstein^{1,*}, Gregory A. Bryant^{2,†}
and Peter Kaye³¹Department of Ecology and Evolutionary Biology, University of California, 621 Young Drive South, Los Angeles, CA 90095-1606, USA²Department of Communication Studies, Center for Behavior, Evolution, and Culture, University of California, Los Angeles, CA 90095-1563, USA³School of Music, Kingston University, Kingston Hill, Kingston upon Thames, Surrey KT2 7LB, UK

*Author for correspondence (marmots@ucla.edu).

†These authors contributed equally to this study.

Humans, and many non-human animals, produce and respond to harsh, unpredictable, nonlinear sounds when alarmed, possibly because these are produced when acoustic production systems (vocal cords and syrinxes) are overblown in stressful, dangerous situations. Humans can simulate nonlinearities in music and soundtracks through the use of technological manipulations. Recent work found that film soundtracks from different genres differentially contain such sounds. We designed two experiments to determine specifically how simulated nonlinearities in soundtracks influence perceptions of arousal and valence. Subjects were presented with emotionally neutral musical exemplars that had neither noise nor abrupt frequency transitions, or versions of these musical exemplars that had noise or abrupt frequency upshifts or downshifts experimentally added. In a second experiment, these acoustic exemplars were paired with benign videos. Judgements of both arousal and valence were altered by the addition of these simulated nonlinearities in the first, music-only, experiment. In the second, multi-modal, experiment, valence (but not arousal) decreased with the addition of noise or frequency downshifts. Thus, the presence of a video image suppressed the ability of simulated nonlinearities to modify arousal. This is the first study examining how nonlinear simulations in music affect emotional judgements. These results demonstrate that the perception of potentially fearful or arousing sounds is influenced by the perceptual context and that the addition of a visual modality can antagonistically suppress the response to an acoustic stimulus.

Keywords: nonlinearities; arousal; valence; noise; frequency modulation; fear

1. INTRODUCTION

Many vertebrates, including humans, produce and respond to harsh, nonlinear vocalizations when alarmed, possibly because these are produced when acoustic production systems (vocal cords and syrinxes) are

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsbl.2012.0374> or via <http://rsbl.royalsocietypublishing.org>.

Received 20 April 2012
Accepted 22 May 2012

overblown in stressful, dangerous situations [1,2]. These nonlinear sounds are somewhat unpredictable and this lack of predictability is hypothesized to make them difficult to habituate to [2]. Indeed, marmots (*Marmota flaviventris*), meerkats (*Suricata suricata*) and tree shrews (*Tupaia belangeri*) respond in more evocative ways to calls containing nonlinearities [3–5]. Recent work has also shown that meerkats do not habituate as easily to calls containing nonlinearities (S. Townsend & M. Manser 2011, personal communication). Humans respond more to infant cries containing nonlinearities than those not containing nonlinearities [6].

Humans compose music that may contain acoustic attributes that would normally be produced by animals in highly affective situations. For instance, rapid amplitude fluctuations, subharmonics (additional spectral components between harmonics), biphonation (sidebands adjacent to harmonics) and noise can all be seen on spectrograms of music, as well as in improvised vocal music performances [7]. Recent research suggests that abdominal muscle activity involved in vocal production is correlated with perceived noisy ‘growl-like’ characteristics in aggressive singing [8]. There is little work, however, examining how these features affect emotional judgements. Humans evolved to respond to an array of stimuli from different modalities, and it is important to understand how emotions are evoked in more complex situations. For instance, films contain narrative, acoustic, and visual components—studying sounds divorced from their more natural contexts is, arguably, somewhat artificial.

Previous work [9] has shown that film soundtracks from different genres differentially contain simulated nonlinear sounds and suggests that the soundtracks do so in order to evoke emotions: scary scenes from horror films contain noisy elements, while dramatic films suppress noise and contain more frequency transitions than would be expected by chance. We aimed to better understand the underlying emotional salience of simulated nonlinearities by presenting human subjects with music composed to include either specific nonlinear sounds or no nonlinear sounds, and then by presenting subjects with video clips containing the identical music. If nonlinearities are generally arousing, we expected subjects to rate music or videos with nonlinear elements as more arousing than those without them.

2. MATERIAL AND METHODS

We conducted two experiments. The first varied the presence or absence of specific nonlinear attributes in music. Using Avid Pro-TOOLS 9 and Native Instruments KONTAKT 4, we (PK, GB) composed 12 original, 10 s control music exemplars devoid of noise or abrupt frequency transitions that were designed to be emotionally neutral (not in definite major or minor modes, of medium tempi, and average dynamic amplitude range). Treatments modified these 12 compositions by adding either noise (through instrumentation changes, simulated electronic distortion using Native Instruments GUITAR RIG 4, or both) or abrupt frequency upshifts or abrupt frequency downshifts at the 5 s mark in the 10 s compositions. Figure 1 illustrates a single composition in both the control condition and the noise condition, with simulated chaotic noise added at the 5 s mark. Thus, we created a total of 36 versions of essentially the same 10 s compositions. In addition, we composed six music ‘fillers’ that served as additional controls but were not formally analysed. The 36 compositions and six fillers were divided into three counter-balanced lists. These sounds were then presented, in a random order using SUPERLAB v. 4.0 software (www.superlab.com), to 42 undergraduate students (23 females, 19 males; average age = 20.8 years) at UCLA who received course credit for participation. After hearing a practice stimulus, the experiment began. For each exemplar,

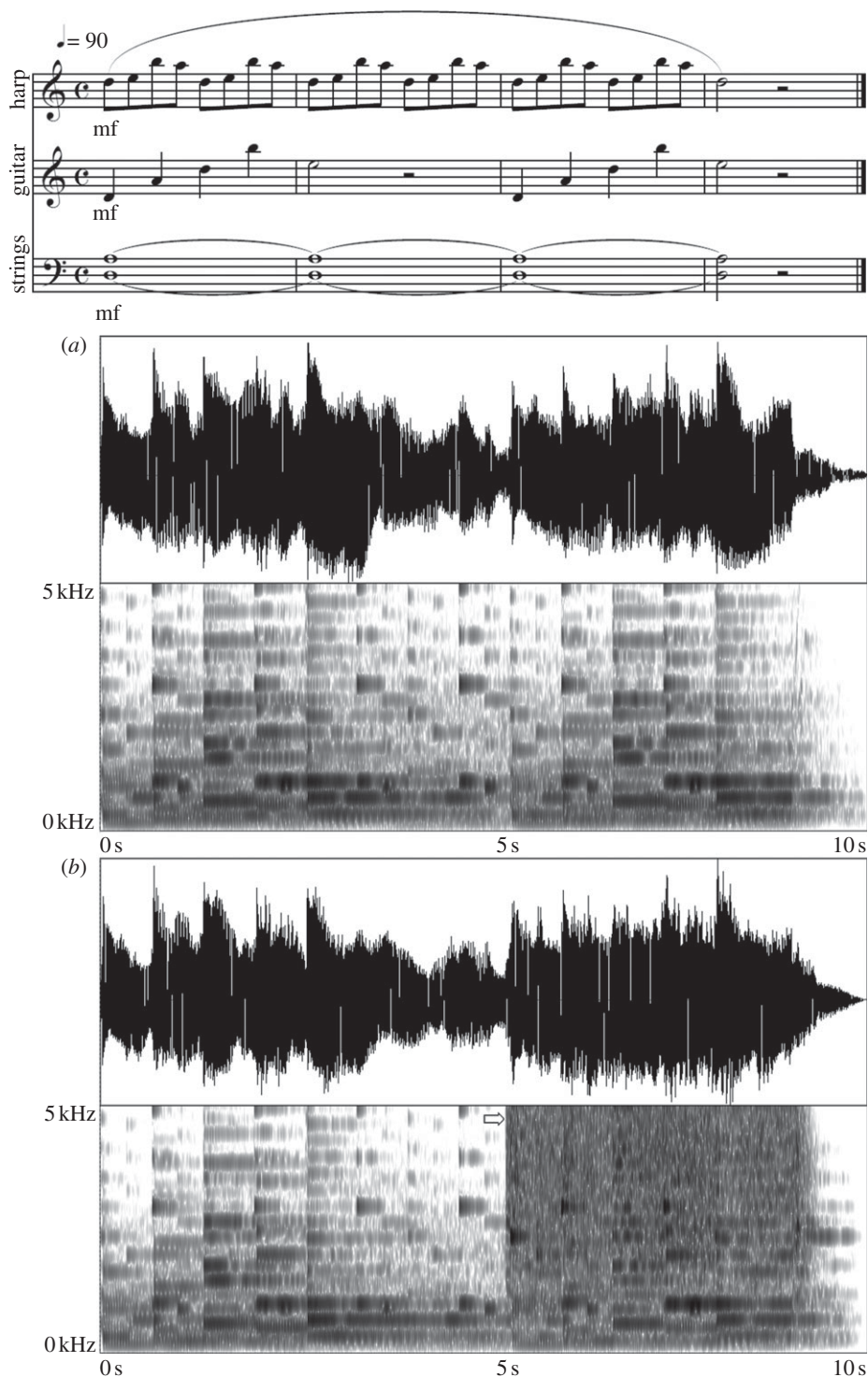


Figure 1. Sheet music, waveform and spectrogram of a sample music composition stimulus: (a) spectrogram of the control condition and (b) spectrogram of the noise treatment. Arrow indicates the onset of simulated nonlinear chaotic noise.

students were given unlimited time to rate the stimuli on two dimensions—arousal and valence—on -5 to $+5$ point scale. Arousal was defined as being emotionally stimulating or active, and valence was explained as being positive or negative with the examples of ‘happy’ and ‘sad’ provided as illustrations. Responses were analysed in SPSS 18 (www.ibm.com) using repeated measures ANOVA with planned pairwise comparisons (Fisher’s least significant difference) to compare treatments. Treatment was included as a within-subjects factor, and subject sex as a between-subjects factor.

The second experiment was identical to the first, except that we added 10 s videos to the identical musical compositions. These 10 s videos were designed to exhibit benign activities that changed at 5 s at the salient narrative point. For instance, a person sat and drank coffee at 5 s, read and turned a book page at 5 s and walked and made a right-hand turn at 5 s (see the electronic supplementary

material, for descriptions of all stimuli). By doing so, we created a multi-modal set of exemplars that were presented to 42 different UCLA students (27 females, 15 males; average age = 19.1 yrs) who were asked to rate these multi-modal stimuli on the same dimensions. Additionally, before seeing the next exemplar, students were asked to write a brief statement about what would happen next on the video; these results are not discussed here. Rating results were again analysed with repeated measures ANOVA.

3. RESULTS

Figure 2 illustrates the rating results of both experiments. In the first unimodal (music only) experiment,

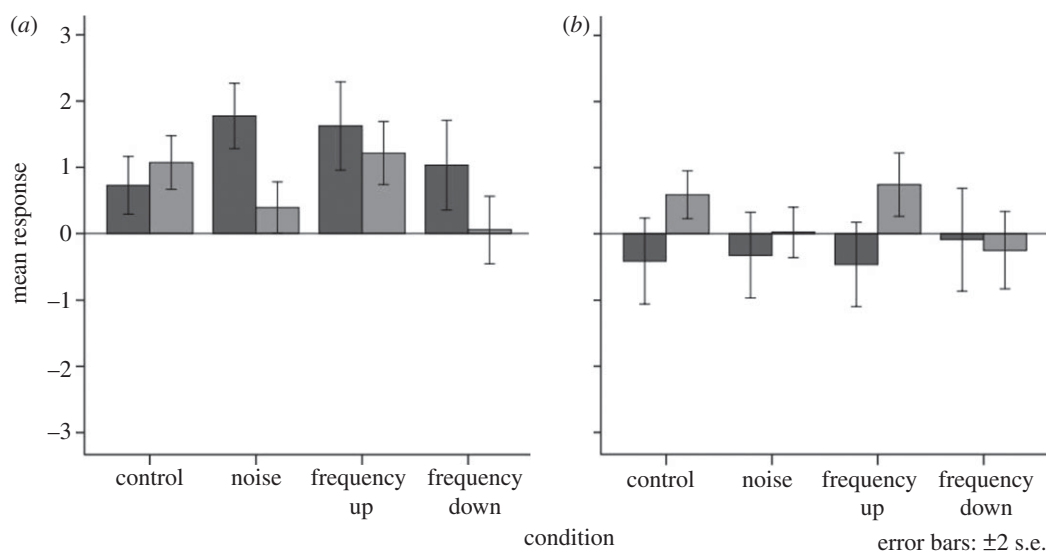


Figure 2. Mean (± 2 s.e.m.) responses to the presentation of (a) music only or (b) music + video stimuli. Dark grey bars, arousal; light grey bars, valence.

we confirmed sphericity (Mauchly's test) for arousal ($p = 0.262$) and valence ($p = 0.390$). Variation in arousal was explained by treatment, $F_{3,120} = 8.083$, $p < 0.0001$, partial $\eta^2 = 0.168$. Compared with controls, subjects reported increased arousal after hearing music that contained either noise ($p < 0.0001$) or upward pitch shifts ($p = 0.029$), though not downward pitch shifts ($p = 0.343$; figure 2). Variation in valence was significantly influenced by treatment, $F_{3,120} = 3.574$, $p = 0.016$, partial $\eta^2 = 0.082$. Compared with controls, valence decreased with the addition of either noise ($p = 0.005$) or downward pitch shifts ($p < 0.0001$), but not with upward pitch shifts ($p = 0.604$). Additionally, valence judgements differed by subjects' sex, with men giving more positive judgements overall than women, $F_{1,40} = 6.092$, $p = 0.018$, partial $\eta^2 = 0.132$, but there was no treatment by sex interaction, $F_{3,120} = 0.582$, $p = 0.628$. There were no sex differences in arousal judgements.

In the second multi-modal (music + video) experiment, we did not confirm sphericity (Mauchly's test) for arousal ($p = 0.086$) or valence ($p = 0.010$), so we report corrected tests (Greenhouse-Geisser) for both measures. No significant variation in arousal was explained by treatment, $F_{2,549,101.959} = 0.398$, $p = 0.755$, partial $\eta^2 = 0.010$, and no within-subject contrasts using estimated marginal means were significant (all F -values < 1). But significant variation in valence was explained by treatment, $F_{2,525,100.981} = 3.675$, $p = 0.014$, partial $\eta^2 = 0.084$. Compared with the controls, valence decreased with the addition of either noise ($p = 0.014$) or downward pitch shifts ($p = 0.010$). There were no observed sex differences in either measure (all F -values < 1).

4. DISCUSSION

Taken together, our results suggest that certain types of music with simulated nonlinearities can alone be evocative and that the addition of relatively nonevocative video images to music with evocative sound characteristics suppresses some affective responses to sound. To

our knowledge, this is the first demonstration that simulated nonlinearities have the predictable emotional effects of increased arousal (i.e. perceived emotional stimulation) and negative valence (i.e. perceived degree of negativity or sadness). When the video images were introduced, arousal was muted, but valence judgements were less influenced. This result implies that perceptual context influences multi-modal perception.

We expected that the introduction of distortion noise to a musical composition would increase arousal based on the role of similar nonlinearities in human and animal vocalizations. Similarly, the finding that human listeners perceive noise as negatively valenced is also consistent with this idea. But it is not clear why the introduction of a benign video would reduce these effects. It could reflect the different sources of potentially important information conveyed by the sounds and the images and the apparent visual dominance seen in this experiment. By design, the images were created to be minimally evocative (see the electronic supplementary material) while the non-control sounds were created to be affectively charged. Thus, the different types of potential information contained in these stimuli could account for the antagonistic perceptual response [10] that we discovered. Alternatively, it is possible that more stimuli increased the cognitive demands or that the visual stimuli distracted [11] participants in some way thus reducing the effectiveness of the simulated nonlinearities to evoke emotional responses. Research on multisensory integration suggests that different sources of information are often combined in domain-specific ways that provide the best available estimate of the relevant properties of a multi-modal stimulus [12]. In the current study, the benign visual information likely mediated the impact of the nonlinear sounds on arousal, and consequently, potentially reduced readiness for an appropriate behavioural response to danger given the apparent lack of danger as evidenced by the visual information.

We studied simulated nonlinearities to gain insights into the biological basis of how people respond to music and other sounds. One potential biologically inspired insight is seen in the difference between

rapid upshifts compared with rapid downshifts. We found that subjects reported increased arousal upon hearing upshifts. It is likely that this is because upshifts would naturally be associated with a sudden increase in vocal cord tension; something that might happen when a mammal is suddenly scared.

Future work should examine the physiological effects of these simulated nonlinearities and explore the connections between particular physiological measures and emotional percepts. Moreover, music cognition researchers should explore ecologically based acoustic phenomena beyond traditionally studied dimensions such as melody, timbre, and rhythm. Most of the perceptual and cognitive machinery underlying music processing likely evolved for a variety of reasons not related to music—many related to emotional vocal communication [13]. This research is one example of how acoustic properties inspired from studies of non-human animal vocalizations can be relevant in the study of human music and emotions.

We thank Ava Dobrzynska for assistance creating the visual stimuli; Shervin Aazami, Natalie Arps-Bumbera, Katie Behzadi, Mesa Dobek, Sean Lo, Alex McFarr, Ben Suddath and Jesse Vindiola for help running the experiments; two anonymous reviewers for astute and very constructive comments; and NSF-IDBR-0754247 (to D.T.B.) for partial support. Research was conducted under UCLA IRB no.10-000728.

- 1 Wilden, I., Herzog, H., Peters, G. & Tembrock, G. 1998 Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics* **9**, 171–196. (doi:10.1080/09524622.1998.9753394)
- 2 Fitch, W. T., Neubauer, J. & Herzog, H. 2002 Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Anim. Behav.* **63**, 407–418. (doi:10.1006/anbe.2001.1912)
- 3 Blumstein, D. T. & Récapet, C. 2009 The sound of arousal: the addition of novel nonlinearities increases responsiveness in marmot alarm calls. *Ethology* **115**, 1074–1081. (doi:10.1111/j.1439-0310.2009.01691.x)
- 4 Townsend, S. W. & Manser, M. B. 2011 The function of nonlinear phenomena in meerkat alarm calls. *Biol. Lett.* **7**, 47–49. (doi:10.1098/rsbl.2010.0537)
- 5 Schehka, S. & Zimmermann, E. 2012 Affect intensity in voice recognized by tree shrews (*Tupaia belangeri*). *Emotion*. (doi:10.1037/a0026893)
- 6 Mende, W., Herzog, H. & Wermke, K. 1990 Bifurcations and chaos in newborn infant cries. *Phys. Lett. A* **145**, 418–424. (doi:10.1016/0375-9601(90)90305-8)
- 7 Neubauer, J., Edgerton, M. & Herzog, H. 2004 Nonlinear phenomena in contemporary vocal music. *J. Voice*. **18**, 1–12. (doi:10.1016/S0892-1997(03)00073-0)
- 8 Tsai, C. G., Wang, L. C., Wang, S. F., Shau, Y. W., Hsiao, T. Y. & Auhagen, W. 2010 Aggressiveness of the growl-like timbre: acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Percep.* **27**, 209–221. (doi:10.1525/mp.2010.27.3.209)
- 9 Blumstein, D., Davitian, R. & Kaye, P. D. 2010 Do film soundtracks contain nonlinear analogues to influence emotion? *Biol. Lett.* **6**, 751–754. (doi:10.1098/rsbl.2010.0333)
- 10 Munoz, N. E. & Blumstein, D. T. 2012 Multisensory perception in uncertain environments. *Behav. Ecol.* **23**, 457–462. (doi:10.1093/beheco/arr220)
- 11 Chan, A. A. Y.-H. & Blumstein, D. T. 2011 Attention, noise, and implications for wildlife conservation and management. *Appl. Anim. Behav. Sci.* **131**, 1–7. (doi:10.1016/j.applanim.2011.01.007)
- 12 Driver, J. & Spence, C. 2000 Multisensory perception: beyond modularity and convergence. *Curr. Biol.* **10**, R731–R735. (doi:10.1016/S0960-9822(00)00740-5)
- 13 Juslin, P. N. & Västfjäll, D. 2008 Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* **31**, 559–575. (doi:10.1017/S0140525X08005529)