Copyright (2010) Acoustical Society of America. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the Acoustical Society of America. The following article appeared in:

Suied, C., Susini, P., McAdams, S. & Patterson, R. (2010). Why are natural sounds detected faster than pips? Journal of the Acoustical Society of America, 127(3), EL105-EL110, DOI: <u>https://doi.org/10.1121/1.3310196</u>

and may be found at:

https://asa.scitation.org/doi/10.1121/1.3310196

Fast detection for natural animal sounds: why?

Clara Suied, Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, Downing Street, Cambridge, CB2 3EG, United Kingdom clarasuied@gmail.com

Patrick Susini, Institut de Recherche et de Coordination Acoustique / Musique and Unité Mixte de Recherche 9912, Centre National de la Recherche Scientifique, 1 place Igor Stravinsky, 75004 Paris, France Patrick.Susini@ircam.fr

Stephen McAdams, Centre for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 1E3 smc@music.mcgill.ca

Roy D. Patterson, Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, Downing Street, Cambridge, CB2 3EG, United Kingdom rdp1@cam.ac.uk

Suied, JASA-EL

Abstract

Simple Reaction Times (RT) were used to measure differences in processing time between natural animal sounds and artificial sounds. When the artificial stimuli were sequences of short tone pulses, the animal sounds were detected faster than the artificial sounds. The animal sounds were then compared with acoustically-modified versions (white noise modulated by the temporal envelope of the animal sounds). No differences in RTs were observed between the animal sounds and their modified counterparts. These results show that a fast detection may be achieved for natural sounds. In the present task, this could be explained by acoustic properties of natural sounds.

© 2009 Acoustical Society of America

Keywords

Auditory warning; Reaction times; Natural sounds.

PACS number: 43.66.Lj

Suied, JASA-EL

1. Introduction

The purpose of an auditory warning is to alert the user of a given system (car, plane, hospital equipment) to a potentially dangerous situation and/or to the arrival of information on visual displays (Patterson, 1982). Several acoustical parameters have been shown to be good candidates to modulate the perceived urgency of an auditory warning: e.g., the higher the pitch and the faster the speed (in case of a multiple-burst sound), the higher the perceived urgency (Edworthy et al., 1991). By contrast with these artificial auditory warnings, some authors have proposed the use of "everyday sounds" as warnings. For example, Graham (1999) observed shorter response times for everyday sounds (car horn, tire-skid) than to conventional warnings (tone) and argued that everyday sounds are understood more quickly and easily than abstract sounds. However, simple acoustic differences (in spectral content, for example) might be sufficient to explain the reaction-time advantage for everyday sounds.

More than an increase in the *perceived* urgency, a warning signal is efficient when it induces a faster detection and increases the probability of an appropriate reaction under urgent conditions. In a companion study (Suied et al., 2008), we have shown the advantages of an objective measurement (reaction time, RT) to assess correctly the level of urgency of a sound.

In this study, we present a pair of experiments performed to investigate whether natural sounds are detected faster than artificial sounds by human listeners. Firstly, we show that natural sounds are detected faster than artificial simple sounds (Experiment 1). Then, we demonstrate that simple acoustic considerations (mainly differences in spectral content) can explain this behavioral advantage, rather than a very early recognition of the sound (Experiment 2).

2. Experiment 1: artificial sounds vs. animal sounds

Suied, JASA-EL

2.1 Methods

Twelve volunteers (7 women; mean age 36 ± 10 years) participated in this experiment. All were naïve with respect to the purpose of the experiment. None of them reported having hearing problems. The study was carried out in accordance with the Declaration of Helsinki. All participants provided informed consent to participate in the study.

Two categories of sounds were compared: classical warning sounds and animal sounds. Four sounds were tested in each category. For the classical warning sounds, we used the same template for the stimuli as in our companion paper (Suied et al., 2008). The template for the different stimuli was an isochronous sequence of short pulses. Each pulse of the burst was a 1-kHz pure tone, 20 ms in duration, and included 5-ms linear onset and offset ramps. Stimuli varied along a single dimension, the InterOnset Interval (IOI), defined as the time elapsed between the onsets of two pulses. The four IOIs tested were 100 ms, 50 ms, 33 ms, and 25 ms (these four sounds are designed hereafter as: *IOI100, IOI50, IOI33*, and *IOI25*). The total duration of each burst was 220 ms. The natural sounds were animal sounds obtained from the SoundIdeas database (a lion sound, two different leopard sounds, and one jaguar sound, designed hereafter respectively as: *lion, leo1, leo2, jag*). They were modified to be 220 ms in duration (with a linear ramp of 10 ms at the end of the sound).

Loudness equalization was performed on the eight stimuli, to avoid any RT differences due to loudness differences (see Chocholle, 1940; Suied et al., 2008). A group of nine other listeners participated in this preliminary experiment. Loudness matches were obtained with an adjustment procedure. The listener was asked to adjust the comparison stimulus until it seemed equal in loudness to the standard stimulus. The IOI100 stimulus was used as the standard stimulus. The level of the standard stimulus was fixed at 76 dB SPL. The mean level difference at which the comparison and the standard stimuli were judged to be equal in loudness were 0.5 dB and 6 dB. The IOI50 sound was presented at 75.5 dB

4

SPL, IOI33 at 75.5 dB SPL, IOI25 at 75.2 dB SPL, lion at 73.6 dB SPL, leo1 at 75 dB SPL, leo2 at 73.7 dB SPL, and jag at 70 dB SPL.

The sound samples were presented at a 44.1-kHz sampling rate. They were amplified by a Yamaha P2075 stereo amplifier and presented binaurally over Sennheiser HD 250 linear II headphones. The experimental sessions were run using a Max/MSP interface on an Apple computer. Participants responded by using the space bar of the computer keyboard placed on a table in front of them. The responses were recorded by Max/MSP, with a temporal precision for stimulus presentation and responses of around 1 ms. The experiments took place in a double-walled Industrial Acoustics Company (IAC) sound booth.

One exemplar of each of the eight stimuli was presented in random order for each trial. Following a standard simple RT procedure, participants had to respond as soon as they detected the sound by pressing the space bar as fast as possible. They were asked to keep the finger of their dominant hand in contact with the space bar between trials. The inter-trial interval was randomly fixed between 1 and 7 s. These stimuli were presented in six separate blocks of trials. Each block consisted of 96 stimuli. The stimuli of different IOIs were randomly intermixed. The number of stimuli of different IOIs was equal in each block (12 each), thus leading to 72 repetitions for each stimulus and each participant. Participants performed practice trials until they were comfortable with the task.

Responses were first analyzed to remove error trials (RTs less than 100 ms and RTs greater than 1000 ms). Each RT value was transformed to its natural logarithm (see Ulrich and Miller, 1993; Luce, 1986), before averaging ln(RT) for each condition (see Suied et al., 2009 for similar analyses on RTs). To identify between-condition differences in mean ln(RT), a repeated-measures ANOVA was conducted with Sound as a within-subject factor (IOI100, IOI50, IOI33, IOI25, Lion, Leo1, Leo2, Jag). A Kolmogorov-Smirnov test was performed to check for the normality of the distribution of residuals of the ANOVA. For this analysis, we

pooled together the results for all conditions in order to increase the power of the statistical test. Finally, to account for violations of the sphericity assumption, *p*-values were adjusted using the Huynh-Feldt correction. p < 0.05 was considered to be statistically significant.

2.2 *Results*

There were no anticipations, only 0.2% misses and 0.2% of RTs greater than 1000 ms. These outlier data were discarded. The Kolmogorov–Smirnov test revealed that the distribution of the residuals of the ANOVA was not different from a normal distribution (d = 0.07; N = 96; p > 0.1). This result validates the log-transformation and shows that the original distribution of RTs was indeed log-normal.

The repeated-measures ANOVA of ln(RT) revealed a significant main effect of Sound $[F(7,77) = 27.25; \epsilon = 0.5; p < 0.0001]$. These data are represented in Fig. 1. We then performed four mutually orthogonal contrasts (F(4,44) = 30.09; p < 0.00001) that show that: (1) RT was significantly shorter for the animal sounds than for the IOI sounds [Lion, Leo1, Leo2, Jag compared to IOI100, IOI50, IOI33, IOI25, t(1) = 6.7; p < 0.00001]. (2) RT was significantly longer for the Lion sound than for the three other animal sounds [t(1) = 3.5; p < 0.005). (3) RT to the IOI100 sound was significantly longer than for the three other IOIs sounds (t(1) = 4.6; p < 0.005]. (4) RT tended to be shorter for IOI33 and IOI25 than for IOI50 [marginal significance: t(1) = 1.8; p = 0.09].

2.3 Discussion

Animal sounds led to a shorter RT than artificial sounds. This could be due to a very early recognition of animal sounds. We could also hypothesize that because of some fundamental acoustical characteristic, these animal sounds induced a brain stem reflex by signaling an important and urgent event (for a review, see Juslin and Vastfjall, 2008), and this might be responsible for the shorter RT. It could also simply reflect the difference in spectral content between the two categories of sounds: by statistical facilitation only, the greater the number of frequency channels activated, the shorter the detection process. Experiment 2 was designed to distinguish between these two possibilities.

For the IOI sounds, the shortest RTs were to IOI33. These data are consistent, at least qualitatively, with a multiple-look model for temporal integration (Viemeister and Wakefield, 1991). The IOI50 sound contains more pulses than the IOI100 sound (and similarly for the IOI33 and IOI50 sounds), so it may lead to more "looks," which might, in turn, induce shorter RTs. The threshold at 33 ms could, however, reflect another process: the lower limit of melodic pitch is around 30 Hz (Pressnitzer et al., 2001). Interestingly, Russo and Jones (2007) have recently found that the urgency of pulse trains is closely related to the perception of pitch: the pulse repetition rate corresponding to the transition between a pitch percept and independent pulses was judged as the most urgent, and led to very short RT.

For the animal sounds, the longer RT has been observed for the Lion sound. This "Lion effect" will be discussed together with the results from Experiment 2 (see 3.5).

3. Experiment 2: animal sounds vs. modulated noises

In this experiment, we compared animal sounds to modified versions of the same sounds (white noise modulated with the temporal envelope of the animal sounds) in order to control for differences in spectral and temporal complexity between natural and artificial sounds in Experiment 1.

3.1 Methods

Twelve new volunteers (5 women; mean age 31 ± 7 years) participated in this experiment. All were naïve with respect to the purpose of the experiment. None of them

reported having hearing problems. The study was carried out in accordance with the Declaration of Helsinki. All participants provided informed consent to participate in the study.

The four animal sounds used previously in Experiment 1 were tested again in Experiment 2. The temporal envelopes of the four natural sounds were applied to white noise to provide the "Modulated Noise" versions, denoted hereafter by the prefix "MN_". The temporal envelope was extracted using a half-wave rectifier followed by a low-pass filter (sixth order Butterworth filter, with a cut-off frequency at 5 kHz). As in Experiment 1, the eight stimuli were equalized in loudness. The MN_Lion sound (used as the reference sound) was presented at 76 dB SPL, Lion at 78 dB SPL, Leo1 at 77.9 dB SPL, Leo2 at 78 dB SPL, Jag at 74.1 dB SPL, MN_Leo1 at 76 dB SPL, MN_Leo2 at 76.2 dB SPL, and MN_Jag at 75.5 dB SPL.

In addition, at the end of the main experiment, we verified that the participants could categorize the eight sounds correctly into "animal" and "non animal" categories. They all did this task very easily.

The apparatus, procedure and statistical analyses were the same as in Experiment 1.

3.2 Results

There were no anticipations, only 0.3% misses and 0.3% of RTs greater than 1000 ms. These outlier data were discarded. A Kolmogorov–Smirnov test revealed that the distribution of the residuals of the ANOVA was not different from a normal distribution (d = 0.11; N = 96; p > 0.1). This result validates the log-transformation and shows that the original distribution of RTs was indeed log-normal.

The repeated-measures ANOVA on $\ln(RT)$ revealed a significant main effect of Sound $[F(7,77) = 6.72; \epsilon = 1; p < 0.0001]$. These data are represented in Fig. 2. Three mutually orthogonal contrasts (F(3,33) = 11.62; p < 0.00001) showed that: (1) there was no clear

8

difference between RTs for the animal sounds compared to those for the MN versions [Lion, Leo1, Leo2, Jag compared to MN_Lion, MN_Leo1, MN_Leo2, MN_Jag, t(1) = 2.1; p = 0.06]; the MN sounds tended to be detected faster than the natural sounds (see Fig. 2); (2) as in Experiment 1, RTs were significantly longer for the Lion sound than for the three other animal sounds [t(1) = 5.5; p < 0.0002]; (3) RTs were significantly longer for the MN_Lion sound than for the three other MN sounds [t(1) = 2.9; p < 0.02].

3.3 Discussion

We observed similar RTs to real animal sounds and their MN versions. If anything, there was a trend for faster RT for the MN sounds, which could be due to the higher number of channels activated for the MN sounds than for the animal sounds. This result validates the "acoustic" hypothesis, showing that results of Experiment 1 were indeed due to a difference in spectral and temporal content between natural and artificial sounds. Although we do not deny a plausible and potential specificity in the encoding and recognition of natural sounds, these findings suggest that, at least for simple detection tasks, the behavioral advantage for natural sounds can be easily explained by simple acoustic differences.

The "Lion effect" observed in Experiment 1 (that is, a longer RT for the Lion sound compared to the other animal sounds) was reproduced in Experiment 2. Interestingly, this Lion effect held for the MN sounds, which preserved only the temporal envelope of the sounds. The importance of the temporal envelope for speech recognition has already been evidenced (Shannon et al., 1995). From the current data, it also seems that the temporal envelope has an impact on the speed of detection. This requires further investigation.

4. Conclusions

In this study, we have shown that natural sounds are detected faster than artificial sounds and have established that this facilitation can be explained by simple acoustic considerations.

Acknowledgments

We would like to thank Marie Magnin and Sabine Langlois for their help. This work was partly supported by Renault SA.

References and links

Edworthy, J., Loxley, S., and Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. Hum. Factors, 33, 205–231.

Graham, R. (1999). "Use of auditory icons as emergency warnings: Evaluation within a vehicle collision avoidance application." Ergonomics, 42, 1233–1248.

Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. Behav. Brain Sci., 31, 559–575.

Patterson, R. D. (1982). "Guidelines for auditory warning systems on civil aircraft." Civil Aviation Authority Paper 82017.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (2001). "The lower limit of melodic pitch." J. Acoust. Soc. Am., 109, 2074–2084.

Russo, F. A., and Jones, J. A. (2007). "Urgency is a non-monotonic function of pulse rate." J. Acoust. Soc. Am., 122, EL185-190.

Suied, C., Susini P., and McAdams S. (2008). "Evaluating warning sound urgency with reaction times." J. Exp. Psychol., Appl.,14, 201–212.

Suied, C., Bonneel, N., and Viaud-Delmon, I. (2009). "Integration of auditory and visual information in the recognition of realistic objects." Exp. Brain Res., 194, 91–102. Shannon, R. V., Zeng, F. G., Wyngoski, J., Kamath, V., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues." Science, 270, 303–304.

Ulrich R., and Miller, J. (1993). "Information processing models generating lognormally distributed reaction times." J. Math. Psychol., 37, 513–525.

Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks. J. Acoust. Soc. Am., 90, 858–865.

Figure Captions

Fig. 1. Reaction times (RTs) of the animal sounds and IOI sounds are presented (from left to right: Lion, Leo1, Leo2, Jag, IOI100, IOI50, IOI33, and IOI25, see text for details). RTs were first transformed to a log scale and then averaged across all participants. The log scale was converted back to ms for display purposes. The error bars represent one standard error of the mean. RTs to the animal sounds were shorter than those to the IOI sounds.

Fig. 2. Reaction times (RTs) of the animal sounds and Modulated Noise (MN) sounds are presented (from left to right: Lion, Leo1, Leo2, Jag, MN_Lion, MN_Leo1, MN_Leo2, MN_Jag, see Fig. 1 for details). RTs to the animal sounds were similar to RTs for the MN sounds that preserved the temporal envelope of the sound.



