HOW LONG DOES IT TAKE TO IDENTIFY EVERYDAY SOUNDS

Anne Guillaume, Lionel Pellieux, Véronique Chastres, and Claude Blancard

Institut de Médecine Aérospatiale du Service de Santé des Armées
BP 73
91 223 Brétigny sur Orge France
aguillaume@imassa.fr

Carolyn Drake

Laboratoire de Psychologie Expérimentale
Université René Descartes
71, avenue Edouard Vaillant
92 774 Boulogne-Billancourt France
drake@idf.ext.jussieu.fr

ABSTRACT

Previous studies of alarm design have concluded that the faster a mental representation of the cause of the alarm is activated, the quicker the adapted reaction. In order to select sounds that are quick to identify, an experiment was carried out using a gated stimulus paradigm with 117 everyday sounds. Almost half of the sounds were identified in less than 150 ms, including both classical alarms and sounds from other categories of everyday sounds. Thus it should possible to identify acoustic properties of each category of alarms within an integrated alarm system in order to improve discrimination among them.

1. INTRODUCTION

Both acoustical characteristics and cognitive processes must be taken into consideration when designing auditory warning signals [1, 2, 3]. Alarms trigger an alert reaction, shifting attention from the main task to the dangerous situation. The judgment about the degree of urgency results from the mental representation created for the alarm in this particular context. The operator will decide on a course of action based on this urgency judgement. Alarm definition involves finding the most direct link between an alarm and its cause or its mental representations in order to limit the attention requirement of listeners to decode the signal; in other words to optimize the cognitive resources of the subject while informing him of a potential danger [4]. In order to rapidly inform the subject, the mental representation must be obtained as quickly as possible. We have previously [5, 6] suggested that the identification of auditory warning signals takes place very early in processing compared with other everyday sounds. We now ask the question of whether more informative sounds can also be identified quickly, whether sounds from different categories can also be identified rapidly. If this turns out to be the case, it will then be possible to attribute the acoustic properties of a different category for each alarm of an integrated alarm system. Using few cognitive processes, each alarm would be easily discernable from the others. In order to isolate rapidly identified sounds, participants completed a free identification task of six categories of everyday sounds: sounds that are produced by water, auditory warning signals (called signalling sounds by Ballas [5]), sounds produced by animals, sounds produced by people, musical instruments, sounds produced by everyday activities. Three tasks were carried out: In the first task, the sounds are presented using a gated stimulus paradigm which has been employed in studies of word perception to investigate the continuous acoustic analysis that occurs in the recognition of individual words [7]. In our experiment, it consists in presenting step by step increasing durations of the sound. The second task was then completed, consisting in the free identification of the full duration sounds. The third task involved listeners rating the sounds on four characteristics that are known to influence stimulus identifiability [5, 8]. The experiment is still in progress.

2. EXPERIMENT

2.1. Procedure

Participants completed a four-part procedure during four 1-hour sessions. The first three sessions were devoted to Task 1 and the fourth to Tasks 2 and 3.

Task 1: The listeners completed the free identification for each step of the gated stimuli. The first step lasted 50 ms. For each further presentation it was increased by 50 ms up to the end of the sound. Each time the listeners gave a confident index from 1 (not at all confident) to 7 (very confident). When three 7s had been successively coded, the next sound was presented.

Task 2: The listeners then performed a free identification of the full duration sounds and rated the ease of recognizing the sound, its typicality, its degree of familiarity and the pleasantness of the sound on scales from 1 to 7. Ease of sound recognition (1 = very difficult to recognize and 7 = very easy to recognize) refers to intuitive recognition of the sound and should be correlated with an early high rating of the confidence index.

name of the sound	duration	mean	std dev	median	name of the sound	duration	mean	std dev	median
dog bark	750	50	0	50	harmonica	950	164	38	150
blow whistle (x2)	400	50	0	50	cough	900	164	56	150
belch	550	50	0	50	pans	950	170	144	100
naughty whistling	950	50	0	50	sheep	950	175	27	175
train	950	50	0	50	camera	600	179	107	200
poured water	1000	50	0	50	broken glass 2	900	179	95	150
tongue banging	950	57	19	50	mosquito	1000	183	154	150
alarm 1	950	64	38	50	cymbal	950	193	137	150
uncorking a bottle of wine	950	64	24	50	girl shout	950	200	115	150
spattered water 1	800	64	24	50	piano	800	200	232	100
spattered water 2	1000	64	38	50	closed door	400	200	122	250
long car horn	950	64	38	50	sniffing (x2)	1000	200	71	200
banged glasses	950	64	24	50	poured champagne without	1000	213	263	100
hammer stroke	950	67	29	50	piling glasses	700	213	103	225
alarm 2	950	71	39	50	alarm 3	1000	220	130	250
alarm 5	1000	71	57	50	rumpling of paper	950	221	202	150
watch alarm	1000	71	39	50	tearing up paper	950	225	176	225
long whistle	950	71	27	50	poured soda in a glass	1000	225	177	225
tapping water	950	71	39	50	apple crunching	950	229	27	250
water drop	300	71	57	50	baby cry	850	230	249	50
roll on the drum	300	71	39	50	swarm	950	233	161	300
gutter whistling	950	71	27	50	starting up car	950	236	103	200
filling up of bottle	1000	75	27	75	house bell	950	236	146	350
long car horn	1000	75	35	75	applause with echo	1000	238	284	125
car horn (x2)	350	75	61	50	big sigh	950	250	167	250
end of gargle	950	79	27	100	sniffing	850	258	120	225
flute	950	83	41	75	donkey	950	263	165	200
frog (x2)	950	83	61	50	opening a soda	950	263	144	325
djembe	1000	86	24	100	truck	950	267	126	250
bottle drains away 1	950	93	73	50	spoon knocking a plate	950	270	236	250
small dog bark	200	93	61	50	transverse flute	1000	271	272	200
bicycle bell	950	93	45	100	poured water in a glass	1000	271	111	250
fork knocking a plate	900	100	87	50	balloon deflation	950	275	106	275
tambourine	900	100	50	100	noose blowing	950	275	29	275
sparrow hawk	950	107	45	100	alarm 4	950	300	196	325
fly	900	107	53	150	match	950	325	119	325
shaking paper	950	107	67	100	flute in bamboo	950	325	96	350
coin	950	108	97	75	cutting paper	950	325	106	325
zip (opening)	550	110	55	150	horse	950	330	67	300
rooster	950	114	85	100	applause at the beginning	1000	333	337	200
kiss	250	117	29	100	stapler	750	350	0	350
zip (closing)	750	120	45	100	bottle drains away 2	1000	350	265	250
xylophone	1000	120	67	150	crowd	1000	350	436	150
hunting horn	950	125	137	75	violin	950	363	269	325
clarinet	950	129	107	50	poured champagne with glup	1000	380	268	450
gargle	950	129	49	150	adhesive tap e	950	390	55	400
organ	950	129	168	50	pneumatic pick	950	400	212	400
police siren	1000	129	81	150	ping pong ball	950	425	254	350
broken glass 1	500	133	29	150	wind	1000	425	106	425
hen	800	142	97	125	outflow of water	450	429	180	350
beaten violin	600	143	79	150	lion	950	500	100	500
camera (x3)	950	150	71	150	handsaw	950	510	195	500
lip vibration	950	150	63	175	cows	1000	540	89	550
goat	850	150	58	150	faucet	1000	583	379	750
guitar	1000	150	95	150	wolf	950	600	141	600
monkey	950	150	0	150	cat	700	617	58	650
falling water	850	157	19	150	car alarm	1000	625	35	625
indian call	950	158	128	100	teeth brushing	950	675	155	675
bubbles in water	950	164	94	150					

Table 1: The 117 sounds are listed with their entire duration, the mean, the standard deviation and the median of the first correct identification time expressed in ms. They are ranked in increasing order of the mean of the first correct identification time.

Typicality refers to the notion that the sound fits well with the mental representation that the listener had of the sound (1 = not at all typical and 7 = very typical). Familiarity refers to how usual or common a sound is in the subject's experience (1 = highly unfamiliar and 7 = highly familiar). The role of familiarity as a property that influences sound identification has been well established [4]. Pleasantness refers to how pleasing or agreeable a sound appears to a listener (1 = very unpleasant and 7 = very pleasant). It has been shown to be an important emotional attribute of everyday sounds [4]. Emotion is part of the perception of urgency and can interfere with the reaction to warning signals. Participants were initially presented one gated sound in order to be familiarized with the tasks.

2.2. Subjects

All 9 subjects had normal hearing. Participants came from a wide variety of professional backgrounds (e.g. psychology students, technicians).

2.3. Stimuli

The stimuli were 117 sounds from several origins: 67 were recorded in an semi-anechoic chamber in IMASSA (44 kHz, 16 bits), 7 were obtained from audio CDs auvidis COMPO or sound library sonoteca (22 kHz minimum , 8 or 16 bits), the others were obtained from various internet websites: www.cofc.edu/Marcell M. (22 kHz re-sampling in 44 kHz, 16 bits), ftp.ircam.fr/private/pcm/steve/M cGillWav (44kHz, 16bits), www.getsound.com, www.skynet.be/adhocsound (44kHz, 16bits), www.sounddogs.com, www.soundwave.com, www.soundsound.com, www.soundorama.com (22 kHz at least re-sampling in 44 kHz , 8 or 16 bits). Because of low frequency noise, a high pass filter (cutoff frequency 50 Hz) was used in order to facilitate the detection of the beginning of the sounds.

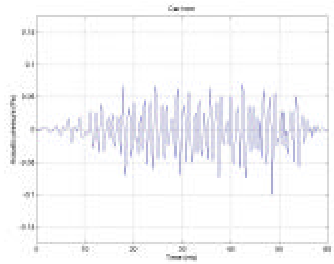
The sounds are listed in Table 1. They were presented randomly to each participant. Loudness was equalized by 5 listeners for all the original sounds, before they were truncated. Sound level was 70 dBA. The final offset ramp of each gated sound was fixed at 10 ms.

2.4. Apparatus

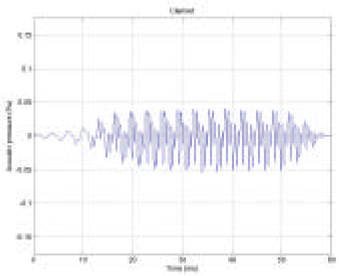
After resampling, the sounds were emitted by RP2.1 TDT processor at 48.8 kHz. The headphone amplifier was HB7 TDT. Listeners sat in a soundproof room and listened to sounds through headphones (Beyer DT990). In both tasks, the subjects gave their answers by typing on a keyboard connected to a computer.

2.5. Results

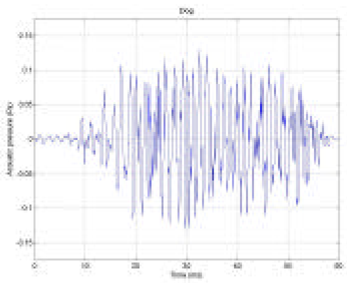
Task 1: The median and the mean of the first correct identification time are presented in Table 1. Here, we focused on the subset of sounds that were identified during the first 150 ms period. Several sounds from each category were identified after this short presentation duration. They represent 47.9% of the sounds. The fastest identified sounds (median = 100 ms, mean = 100 ms) are listed in Table 1 from "dog bark" to "tambourine". Then the rapid identified sounds (median = 150 ms, mean = 150 ms) go from "sparrow hawk" to "monkey". Figures 1abcde and f show examples from each category of the temporal waveform. We are still characterizing the difference in their acoustic properties but it appears that the temporal structure is important.



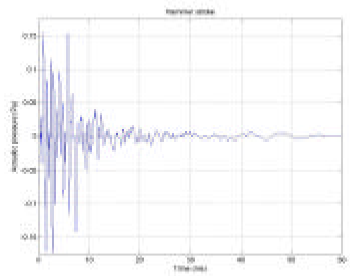
a) car horn



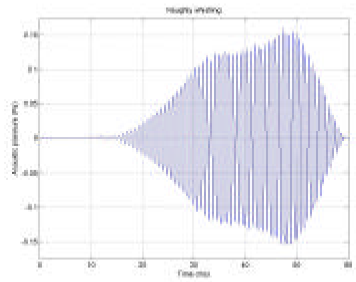
b) clarinet



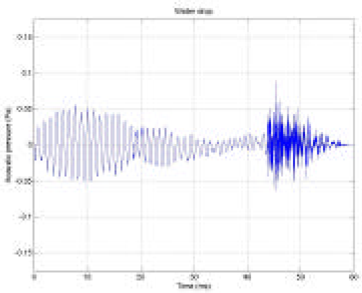
c) dog bark



d) hammer stroke



e) whistling



f) water drop

Figure 1: Time wave forms of sounds from each category during the first 50 ms.

Task 2: Correlations between the time of first correct identification and each rating were low though very significant. We obtained $r(819)=-.35$, $p<.0001$ for the easiness to recognize the sound, $r(819)=-.48$, $p<.0001$ for typicality, $r(819)=-.13$, $p=.0001$ for the degree of familiarity and $r(819)=-.26$, $p<.0001$ for the pleasantness of the sound.

3. DISCUSSION

The results argue in favor of very early identification for almost half of the presented sounds. This early identification concerns not only auditory alarms as expected but also sounds from the other categories. The choice of the categories was inspired by the literature [5, 8, 9]. Their number could be increased as long as the difference in the acoustic properties remains adequate to allow early discrimination. The best correlation was indeed obtained between typicality and the time of first correct identification. This correlation would probably improve with the number of listeners. Nevertheless further work is needed in order to define more accurately the possible categories.

4. CONCLUSION

Guillaume et al. [3] assume that the evocation of a mental representation of the cause of the alarm is essential to an adapted reaction of the operators. To design a system of auditory alarms, the acoustical properties of each has to be identified very quickly. In order to select such sounds, an experiment was carried out using the gated stimulus paradigm. The results showed that not only auditory alarms can be identified very early as expected but also sounds from different categories such as sounds that are produced by water or sounds produced by animals for instance.

5. REFERENCES

[1] Edworthy, J., Loxley, S., and Dennis, L. (1991). “Improving auditory warning design: relationship between warning sound parameters and perceived urgency”, *Human factors* 33: 205-231.

[2] Hellier, E., Edworthy, J., and Dennis, I. (1993). “Improving auditory warning design: quantifying and predicting the effects of different warning parameters on perceived urgency”. *Human Factors* 35: 693-706.

[3] Guillaume, A., Pellieux, L., Chastres, V., and Drake, C. (2003). “Judging the urgency of non-vocal auditory signals: perceptual and cognitive processes”. *Journal of Experimental Psychology: Applied*, 9: 196-212.

[4] Guillaume, A., Drake, C., Rivenez, M., Pellieux, L and Chastres, V. (2002). “Perception of urgency and alarm design”. *Proceedings of the 8th International Conference on Auditory Display*: 357-361.

[5] Ballas, J.A. (1993). “Common factors in the identification of an assortment of brief everyday sounds”. *Journal of Experimental Psychology: Human Perception and Performance* 19: 250-267.

[6] Copti, L. (2001). “Sélection d’une banque de sons de l’environnement selon la familiarité et la durée optimale de reconnaissance”. Mémoire de maîtrise. Université de Montréal. Faculté des études supérieures.

[7] Warren, P., and Marslen-Wilson W. (1987). “Continuous uptake of acoustic cues in spoken word recognition”. *Perception and psychophysics* 41: 262-275.

[8] Marcell, M.M., Borella, D., Greene, M., Kerr, E., and Rogers, S. (2000). “Confrontation naming of environmental sounds”. *Journal of Clinical and Experimental Neurophysiology* 22: 830-864.

[9] Bonebright, T.L. (2001). “Perceptual structure of everyday sounds: a multidimensional scaling approach”. *Proceedings of the 7th International Conference on Auditory Display*: 73-78.