IDENTIFICATION OF ENVIRONMENTAL SOUNDS: ROLE OF RHYTHMIC PROPERTIES

Anne Guillaume, Marie Rivenez, Véronique Chastres, Claude Blancard, Lionel Pellieux

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

ABSTRACT

Studies that have dealt with the effect of sound spectral and temporal properties on environmental sound identification have focused on a narrow range of sounds [1, 2, 3]. The purpose of the research was to evaluate the effect of sound temporal characteristics on the identification of 72 different environmental sounds, 29 of them having a rhythmic structure. We used a gating paradigm, involving a successive presentation of increasing increment of gates stimulus [4] that listeners had to identify. The minimum amount of time presentation (uniqueness point) for which an environmental sound was correctly recognized was recorded. We found that rhythmic sounds were identified earlier (tR = 160 ms) than non rhythmic (tNR = 239 ms). Furthermore, for rhythmic sounds, we observed a significant correlation between the uniqueness point and the duration of the first inter onset interval ($r = 0.65$). Our results suggest that sound rhythmic structure is an informative parameter in the identification process.

1. INTRODUCTION

Pollack [5] claimed that temporal pattern is a crucial feature of communication signals, so that human being as animals should be particularly sensitive to rhythmic structures. Experimental evidences of this claim have been given by various studies [2, 3] which have focused on the influence of spectral and temporal properties on the identification of environmental sounds. Especially Warren and Verbrugge [2] have shown the importance of the rhythmic structure for distinguishing an object rebounding and an object breaking in several pieces. They constructed cases of bouncing and breaking, eliminating average spectral differences between the two and even removing the naturally occurring burst in the breaking cases. They found that multiple quasi-periodic pulse trains provide sufficient information to correctly categorize breaking. This has been confirmed by Repp [3] using hand clapping sounds. However, previous studies have concentrated on a narrow range of environmental sounds (breaking sounds, hand clapping). The purpose of our research was to generalize these finding to a wider set of environmental sounds. We evaluated the influence of the sound rhythmic structure on the identification comparing rhythmic and non rhythmic sounds. In order to have a sensitive enough paradigm, we used a gating paradigm involving a successive presentation of increasing increment of gates stimulus [4] that listeners had to identify. The gating paradigm has been traditionally used in psycholinguistic to study the effect of factors such as word frequency, number of word neighbours, semantic context etc. on identification performances (see for instance [5]). It allowed to measure the uniqueness point of the sound (minimum amount of information necessary to convey the nature of a word or a sound). We compared the uniqueness point for rhythmic and non rhythmic sounds. We assumed that the rhythmic sounds should be identified earlier than non rhythmic ones since they carry a high level of temporal information (their uniqueness point should be shorter for rhythmic than non rhythmic sounds).

2. EXPERIMENT

2.1. Method

2.1.1. *Material and Procedure*

117 environmental sounds representing different categories of sources [6, 7] (sounds that are produced by water, auditory warning signals (called signaling sounds by Ballas [8]), sounds produced by animals, sounds produced by people, musical instruments, sounds produced by everyday activities) were used for this experiment (see the detail of the method in [9]). The duration of each gate was 50 ms. Sound presentation order was randomized but for one environmental sound, the gates were always presented successively in the same order (from the shortest to the longest).

Participant had to perform a free identification task for each gate. They were instructed to write down the source they thought the sound came from.

2.1.2. Participants

18 naïve listeners of IMASSA participated to this experiment. They all reported normal hearing. They were familiar with experiments in auditory perception but none of them had previously heard the stimuli of the present study.

2.2. Results

2.2.1. Sound selection

Sound selection was performed in two steps: (a) – we first selected sounds regarding to the psychometric curve obtained with the percentage of correct identification for each gate (b) – we assigned the sounds to a rhythmic and a non rhythmic categories.

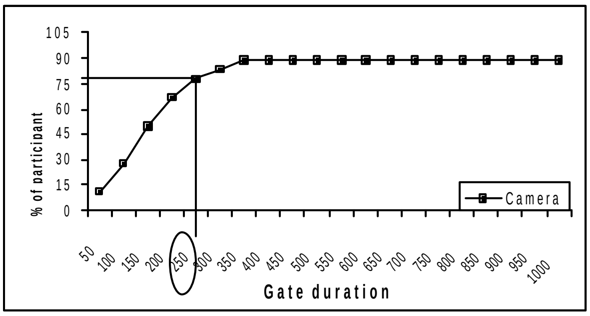

Figure 1: Psychometric curve representing the percentage of participant having correctly identified the sound of a "camera" for each gate duration (expressed in ms).

(a) – The psychometric curves related the gate durations and the percentage of participants having correctly identified the sound. In order to estimate the uniqueness point, we choosed to consider the gate duration which was identified by at least 78% of the participants. Figure 1 gives an example of psychometric curve. On this curve, the uniqueness point of the sound produced by a "camera" was estimated at 250 ms. Using this method, we selected 72 of the sounds used in our experiment. Sounds identified by less than 78% of the listeners were discarded from the analysis.

(b) – The 72 selected sounds were then categorised into rhythmic or non rhythmic sounds. A sound was considered as rhythmic if it was possible to calculate an interval between at least two bursts (IOI for Inter-Onset-Interval). We considered a burst occurrence when the sound level crossed a baseline fixed at 3dB above the minimum level of the signal in order to avoid interferences with the background noise. Figure 2 (graph below) represents the temporal envelope of a "table tennis ball rebounding" sound superimposed to its spectrogram. In this example, 10 rebounds were taking into account while others were neglected. The inter onset interval (IOI) was measured in milliseconds. It corresponded to the duration between the onsets of the successive bursts and is inversely related to the musical concept of tempo (beats per minute). For each rhythmic sound, the 1st IOI and the 1st burst durations were measured.

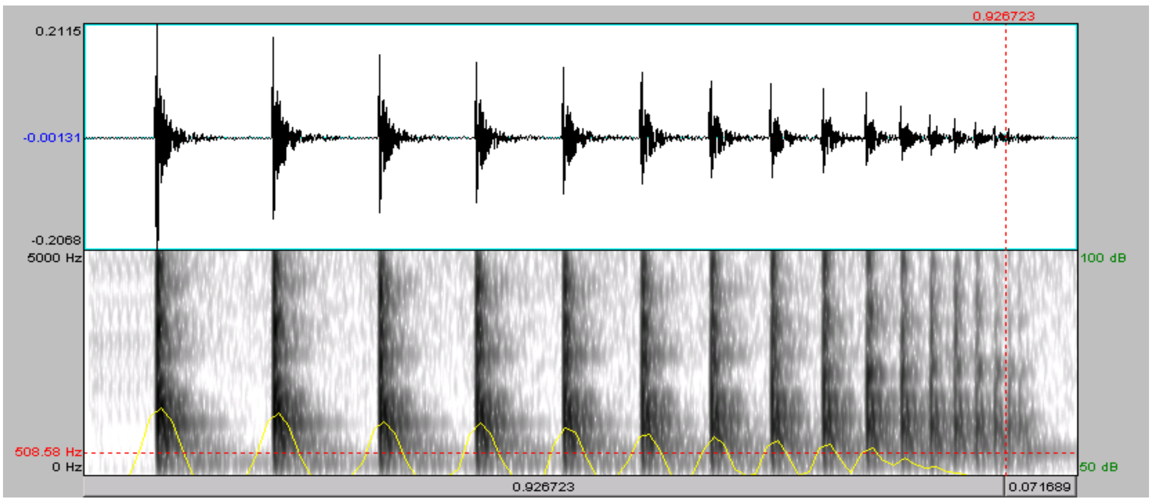

Figure 2. Above: temporal spectrum of the "table tennis ball rebounding" sound; below: sound temporal envelope superimposed on its spectrogram. The red line indicates the retained intensity level for the measures of temporal parameters.

Rhythmic sounds (n=29)	Non rhythmic sounds (n=43)
Alarm 1	Alarm 3
Alarm 2	Match
Alarm 3	applause at the beginning
watch alarm	spoon knocking a plate
camera	uncorking a bottle of wine
camera (3 bursts)	dog barking
tennis table ball	clarinet
filling up of bottle	rooster
bottle drains away 1	horn
bottle drains away 2	long car horn
lip vibration	car horn (x2)
bubbles in water	girl shout
poured champagne 1	apple crunching
poured champagne 2	cymbal
goat	starting up car
tongue banging	djembe
tapping water	falling water
faucet	sparrow-hawk
spattered water 1	zip (closing)
spattered water 2	zip (opening)
water drop	flute
frog (x2)	transverse flute
shaking paper	end of gargle
rumpling of paper	long whistle
coin	guitar
broken glass 2	Harmonica
poured soda in a glass	Indian call
poured water	car horn
poured water in a glass	fly
	noose blowing
	sheep
	small dog bark
	belch
	roll on the drum
	whistling
	naughty whistling
	police siren
	house bell
	bicycle bell
	Cough
	banged glasses
	piling glasses
	beaten violin

Table 1. Detail of the sound categorised as rhythmic and non rhythmic.

2.2.2. *Rhythmic and non rhythmic sounds*

The rhythmic sounds represented 40.3% of the selected environmental sounds (n = 29) while the others were considered as non rhythmic (n = 43). Details of the sounds categorized as rhythmic and non rhythmic are presented in Table 1.

The effect of the temporal structure of the sound on the identification of environmental sounds was evaluated measuring the following cues: 1- for both rhythmic and non rhythmic sounds, the uniqueness point which was estimated by the gate duration for which 78% of the participants correctly identified the sound and 2 – only for rhythmic sounds, the 1^{st} IOI and the 1^{st} burst durations.

When considering the average temporal estimation of the uniqueness points for the two groups of sounds, we found that the rhythmic sounds were identified earlier (160 ms) than non rhythmic ones (239 ms) (Figure 3). A t-test contrasting rhythmic and non rhythmic sounds revealed that this difference was marginally significant ($t(70) = -1.81$; $p < 0.10$).

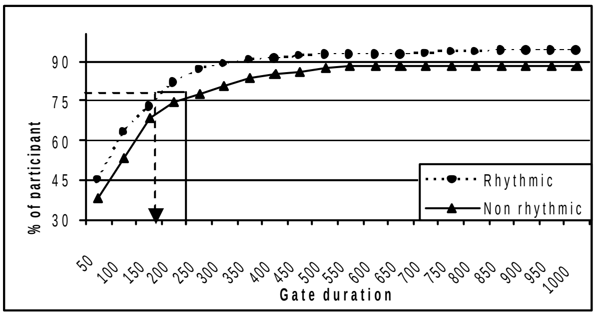

Figure 3: Psychometric curves for rhythmic (circles) and non rhythmic (triangles) sounds.

The average 1^{st} IOI duration estimated for rhythmic sounds was 81 ms and the average 1^{st} burst duration was 53 ms. These values were shorter than the average uniqueness point (160 ms).

In order to determine which temporal parameter affected the uniqueness point, we carried out a multiple regression of the duration of the 1^{st} IOI (IOI) and the 1^{st} burst (BURST) on the temporal estimation of the uniqueness point (UNI). We found that the regression of these two parameters explained 81% (R^2) of the variance of the uniqueness points ($F(2, 5) = 10.41$; $p < 0.05$). The equation obtained from the regression was:

$$UNI = 6.8\ IOI - 4.3\ BURST - 102.4$$

This equation predicted that when the uniqueness point increased by 1, the 1^{st} IOI duration increased by 6.8 and the 1^{st} burst duration decreased by 4.3. When the 1^{st} IOI and 1^{st} burst durations were 0, the uniqueness point was estimated to 102.4 ms.

The parameter giving the best prediction was the 1st IOI ($t(5) = 4.27$; $p < 0.01$). There was a significant correlation between this parameter and the uniqueness point ($r = 0.65$, $p < 0.05$): the shorter the IOI was, the shorter the temporal estimation of the uniqueness point was. The predictive power of the 1st burst duration was marginally significant ($t(5) = -2.43$; $p = 0.06$). This was confirmed by the absence of significant correlation between this parameter and the uniqueness point.

3. DISCUSSION

Among the 72 environmental sounds selected, the rhythmic sounds were identified earlier than the non rhythmic ones. The link between the uniqueness point and the first IOI indicated that the identification of environmental sounds is influenced by the temporal structure of the sound. The absence of correlation between the uniqueness point and the first burst duration suggested that the spectral information contained in the burst seems less informative than the silence between the bursts as measured with the 1st IOI.

These results are consistent with those of Warren and Verbrugge [2] who showed that removing the initial burst from breaking sounds did not reduce the performances of identification. Since in our experiment, the spectral properties of the sounds were not manipulated as they were in Warren and Verbrugge [2], it could be interesting to alter the spectral properties of the tested rhythmic environmental sounds in order to look at the influence of these changes on the uniqueness point.

4. CONCLUSION

This study was carried out using the gating paradigm in order to test the importance of the rhythmic structure of environmental sounds on their uniqueness point. The results showed that rhythmic sounds were identified earlier than non rhythmic ones. The uniqueness point was correlated to the first IOI but not to the first burst duration indicating a stronger influence of the temporal information than of the spectral ones.

Acknowledgments: This work was supported by DGA (PEA010801 Operation n°01CO014).

5. REFERENCES

[1] Lakatos, S, McAdams, S., & Caussé, R. (1997). "The representation of auditory source characteristics: simple geometric form". Perception and Psychophysics, 59, 1180-1190.

[2] Warren, W. H., & Verbrugge, R.. R. (1984). "Auditory perception of breaking and bouncing events: A case study in ecological acoustics". Journal of Experimental Psychology: Human Perception and Performance, 10: 704-712.

[3] Repp, B. (1987). "The sound of two hands clapping: An exploratory study." Journal of Acoustical Society of America, 81: 1100-1109.

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

[5] Grosjean, F. (1980). "Spoken word recognition processes and the gating paradigm". Perception and Psychophysics, 28, 267-283.

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

[7] Pollack, G. S. (2001). "Analysis of temporal patterns of communication signals". Current Opinion in Neurobiology, 11: 734-738.

[8] 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.

[9] Guillaume, A., Pellieux L., Chastres V., Blancard C., & Drake, C. (2004). "How long does it take to identify everyday sounds?" Proceedings of the 2004 International Conference on Auditory Display, Australia.