Evaluating warning sound urgency with reaction times

Clara Suied[1], Patrick Susini[1] and Stephen McAdams[2]

[1] IRCAM – CNRS, 1 place Igor Stravinsky 75004 Paris, France. +33 1 44 78 48 85. Clara.Suied@ircam.fr

[2] CIRMMT, Schulich School of Music, McGill University 555 Sherbrooke St. W. Montreal, QC, Canada H3A 1E3

Abstract

Acoustic parameters have been shown to affect the perceived urgency of alarms. Previous studies have focused primarily on the subjective relationship between characteristics of the alarms and the perception of urgency. In this study, we investigate an objective measurement, reaction time (RT), to test the effectiveness of the acoustic parameters. Three experiments were performed using a speeded RT paradigm with two different tracking tasks simulating emergency conditions. The three experiments show that RT decreases as tempo increases and that, in some cases, RT decreases with the temporal irregularity of sequences. Comparisons between RT measurements and subjective judgments of urgency given by participants support the validity of our approach. Experiments 1 and 2 highlight sensory coding differences, whereas Experiment 3 emphasizes a higher-level process. This RT paradigm leads to concrete recommendations and provides a useful tool to clarify some of the factors involved in alarm processing.

Keywords

Auditory warnings; Urgency; speeded reaction time; Tracking task.

1. Introduction

The use of sounds to present information is now relatively common in applications such as hospital equipment or high performance aircraft (Stanton & Edworthy, 1999). More recently, new in-car technologies have led to an increasing number of sound interfaces, ranging from information related to satellite navigation to warning signals used to alert drivers to potential danger. Among all potentially dangerous events that may be encountered while driving, some of them could now be avoided by using new technologies such as Adaptive Cruise Control (ACC). The distance to other vehicles and the relative speed could be calculated by in-car computers. When this distance is too small, for example, a warning signal informs the driver that he or she has to react as soon as possible. The question then becomes one of alerting drivers to potential dangers and inciting them to act as quickly as possible. The main goal of the study presented in this paper is thus to examine the relevance of reaction time measurements to the evaluation of warning sound urgency.

Until a few years ago, warning signals were badly designed, and Patterson first reported some of the problems typically associated with auditory displays from aviation or medical environments (Patterson, Edworthy, Shailer, Lower & Wheeler, 1986; Patterson, 1990). One of the main problems he highlighted is that warning signals can become a distraction during times of high workload, instead of just attracting listeners' attention and providing relevant information about a potential problem or sudden urgency.

On the basis of these established facts, Patterson (1990) proposed a pragmatic methodology in four steps to design warning sounds properly in an industrial and high workload environment. These four steps include determining the appropriate level of loudness, designing a small pulse of sound, incorporating the pulse into a longer burst of sounds, and forming a complete auditory warning signal using bursts of sounds followed by short periods of silence. His work is mainly focused on the first step, i.e. taking into account the background noise in which the sounds are to be heard. The three final steps concerned the interpretation of the warning signal (design of pulse, burst and warning). Patterson proposed that once the ’structure’ of the auditory warning is designed, the perceived urgency may be altered by adjusting the pitch, the intensity and the speed of the burst.

In order to give precise recommendations on "urgency matching", i.e. to associate the appropriate warning sound with the appropriate urgent situation, knowledge of the effect of sound parameters on perceived urgency is required. Edworthy and her colleagues (Edworthy, Loxley, & Dennis, 1991) thus conducted an extensive series of experiments showing that some pulse and burst parameters had clear and consistent effects on the perceived urgency of a warning sound. They were able to define a series of pulse parameters such as fundamental frequency, harmonic series, amplitude envelope shape, delayed harmonics, as well as temporal and melodic parameters such as speed, rhythm, pitch range, and melodic structure to describe precisely what a warning sound would have to be. The subjective judgments indicated that the faster the rate, the higher the pitch, and the more randomly irregular the frequencies of the harmonics, the greater the perceived urgency.

These results have been confirmed in different studies (e.g. Hellier, Edworthy & Dennis, 1993; Hellier & Edworthy, 1999). Hellier et al.’s (1993) results revealed the power function relation between acoustic parameters (speed, fundamental frequency, repetition number and harmonic content) and perceived urgency. The authors showed that some parameters contribute more to perceived urgency than do others, with speed being the most efficient parameter to communicate urgency.

Edworthy and her colleagues (1991) also studied the effect of temporally unpredictable events on perceived urgency. They found that an irregular rhythm is perceived as slightly less urgent than a regular one. However, it should be noted here that the irregular rhythm was a syncopated rhythm, i.e. a regular rhythm with events occurring off the main beat period (see Fraisse, 1982).

In summary, Edworthy et al.'s (1991) alarm-design principles appear to be relatively robust. For instance, their validity was tested again under the same conditions and verified by Hellier et al. (1993) and Guillaume and her colleagues (Guillaume, Pellieux, Chastres & Drake, 2003). On the basis of Patterson's four steps and these studies, urgency matching is thus possible: the most urgent warning sound can be associated with the most urgent situation. Finally, similarly to these warning studies, Walker (2002) showed that the psychophysical method used by Edworthy et al. (1991) and Hellier et al. (1993) could also be useful for developing effective data sonification, where sonification is defined as a "particular type of auditory display, in which relationships in a data set are translated into, or represented by, sounds for the purpose of understanding or discovering patterns in the data set" (Walker, 2007).

1 What is "urgency"?

The psychophysical approach adopted by Edworthy and colleagues (1991) has been further extended (Guillaume et al., 2003), highlighting the role of cognitive factors in perceived urgency. These latter authors showed that although Edworthy et al.'s (1991) results were replicated with synthesized alarms, they were not entirely valid when they applied the same methodology to real alarms, recorded from alarms used in military aircraft in France. Some real alarms did not follow the predicted pattern, because of a possible highly learned association between an alarm and its meaning. The authors concluded that the design of alarms should take into consideration the acquisition of a "mental representation".

Thus, whereas the "perceived urgency" approach has provided several useful experimental results, it is not clearly established what urgency is and what we really need to study. According to common sense, urgency is what requires immediate action or attention. In realistic conditions, an urgent situation often occurs under high workload. Auditory warnings thus signal potentially dangerous conditions or equipment malfunctions. As a result, the warning attracts attention, and the listener has to react immediately. Therefore, a warning signal is efficient when it increases the probability of an appropriate reaction under urgent conditions (Guillaume et al., 2003), which is usually associated with decreased reaction times (Bliss, Gilson & Deaton, 1995). Although an immediate reaction is not always required in applications such as high performance aircraft or medical environments (Patterson, 1990), we need it in some automobile applications (e.g., Automatic Cruise Control technology).

Consequently, the most relevant questions here are: do we react more rapidly to some sounds than to others? If we do, which are the parameters that most improve our reaction time (in terms of a decrease in reaction time)? An answer to these questions should lead to a more appropriate understanding of at least one part of the processes involved in urgent situations and to more useful recommendations of how to design powerful warning sounds.

Instead of using perceptual assessment in which a listener is asked to make some judgment about the sounds that are played (e.g., which of the two sounds is the more urgent), we need an objective measurement, such as reaction time (RT). Since the work of Donders (1969), chronometric analyses of mental processes have been performed using reaction-time paradigms (for a complete review, see Luce, 1986).

We are not the first to propose such a measurement in the context of warning sounds. To determine the degree to which a participant's response is due to these different processes, Burt et al. (1995) added physiological measures and reaction times to subjective assessment. One of the assumptions tested by the authors was that faster reaction times would occur in response to the most urgent warning. However, reaction times did not differ significantly in response to the urgency levels tested (parameters tested were fundamental frequency and harmonic series, as recommended by Edworthy et al., 1991). With these same parameters, Edworthy et al. (1991) did find differences in urgency judgments. Consequently, predictions about a possible link between higher urgency alarms and improvement in alarm reaction time were not really confirmed.

At the same time, Haas and Casali (1995) conducted a study investigating the effect of pulse format, pulse level and inter-pulse interval on *subjective* perceived urgency and *objective* reaction time. They found that only pulse format (sequential pure tones, simultaneous pure tones, or frequency-modulated tones) and pulse level significantly influenced RT. However, each pulse of the total signal had a duration of 350 ms and they found a mean reaction time of approximately 450 ms. So inter-pulse interval could not have had an effect on reaction time. The authors did not comment on this lack of effect of the inter-pulse interval on reaction time.

Another purpose of Haas and Casali's study (1995) was to investigate the relationship between perceived urgency and reaction times. The authors found a correlation between the two measures: as perceived urgency increases, response time to the signal decreases. To explain this correlation, they assumed that higher-urgency alarms would appear more important to the listener, thus leading to faster reaction times. This was obviously not the case, at least for one of the parameters tested (inter-pulse interval). Moreover, there is no obvious reason that can explain why perceived urgency judgments should produce a decrease in an objective measurement such as simple reaction time.

In industrial environments, listeners are often engrossed in alternative tasks that require attention. For this reason, a dual task paradigm is often used to determine the effect of the alarm and to evaluate attention and mental workload (Sorkin, Kantowitz & Kantowitz, 1988; Bliss et al., 1995). It is not sufficient to show that performance on reaction time should improve when the alarm is well designed. It is also important to study the effect of different alarms in a realistic context, because real alarms are embedded within systems. Two studies (Burt et al., 1995; Haas & Casali, 1995) used a cognitive loading task that imposes attentional demands, as in military or industrial environments. Burt et al. (1995) studied attentional engagement by manipulating two conditions of a tracking task, manual and automated. The authors show, in line with classical findings on dual-task performance (Pashler & Johnston, 1998), that participants produce slower reaction times during the manual tracking condition. More importantly, they also established that the Edworthy alarm-design principles were still reliable under high cognitive load.

2 The Current Experiments

In the current study, we examined the influence of acoustic parameters on reaction time under high workload conditions.

In Experiment 1, we addressed the issue introduced by Haas and Casali (1995) and investigated the influence of InterOnset Interval (IOI) on reaction time. As highlighted before, the authors did not find an effect of IOI on reaction time. However, we hypothesized that this lack of effect was due to the pulse duration (350 ms); thus, the rhythm could not be heard before the listener reacted. In the present experiment, listeners were presented with a burst of pulses that changed in IOI, and were asked to press a button as soon as they heard the sound. At the same time, listeners were asked to perform simultaneously a primary tracking or monitoring task. The tracking condition of the primary task was designed to maintain a high level of attention at all times. Experiment 2 was partly suggested by results obtained in Experiment 1. Stimuli of Experiment 1 were equalized in loudness in order to examine the effect of IOI removing any possible effect of loudness on reaction time. In Experiment 3, we continued the examination of the influence of temporal parameters on reaction time and introduced temporal irregularity. As noticed previously, temporally unpredictable events may be more attention-getting than predictable ones, but this assumption still needs to be demonstrated empirically. At the end of Experiments 1 and 3, participants also rated the two different warnings (two IOIs and two temporal regularities) on an urgency rating scale.

The first issue addressed in this research concerns how IOI and temporal differences affect participants' reaction times under high workload conditions. The second issue concerns a comparison between subjective and objective measurements.

1 Experiment 1

3 Method

3.1 Participants

Thirteen participants (four women, mean age ± standard deviation = 27.8 ± 2.4 years) were recruited for this experiment. They were compensated for their participation. None of them reported having hearing problems. All were right-handed and reported normal or corrected-to-normal vision.

3.2 Stimuli

The template for the two 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 only varied along a single dimension, the InterOnset Interval (IOI). IOI is the time elapsed from the onset of one pulse to the onset of the next. The two IOIs were 100 ms and 300 ms. The total duration of each burst was 920 ms. It should be noted here that an IOI variation leads to a variation in the number of pulses in a burst. As the two stimuli have the same total duration, we cannot study IOI and number of pulses as independent factors. We accepted this inevitable fact, and will henceforth only use the IOI terminology.

3.3 Apparatus

The sound samples were generated with a 44.1-kHz sampling rate under the control of a PC, using Matlab v.7 software. The sound samples were amplified by a Yamaha P2075 stereo amplifier and presented binaurally over a Sennheiser HD 250 linear II headphone set. Stimuli were presented at 76.5 dB SPL. 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 around 1 ms. The primary task was created using Jitter, the graphical part of Max/MSP. Performance on the tracking condition was recorded every 10 ms by calculating the distance between the target and the pointer. All data were collected in the computer memory for further off-line analysis. The experiments took place in a double-walled IAC sound booth.

3.4 Procedure

The primary task consisted of two attentional conditions, called “tracking” and “monitoring”. During the tracking condition, participants were required to track a circular target manually, in order to keep it within a circular boundary that moved at a constant speed and in a random trajectory on the screen. Figure 1 shows the circular pointer (in white; orange in the real interface) and the target (concentric circles in grey). Participants had to perform the task with their non-dominant hand. This made the tracking condition continuously demanding for each participant. It was necessary for the tracking condition to be challenging, so that the effort expended by the participants to perform it (and respond to the alarms) would approximate the effort experienced in many complex task situations (de Waard, 1996). Participants were instructed to optimize response speed and accuracy of the tracking condition. During the monitoring condition, they were required to monitor computer tracking of the circular target and to perform only the reaction time task. Although there was no measure of what participants really did during the monitoring condition, they were explicitly instructed to monitor the target visually.

Throughout the primary task, one replication for each of the two auditory warnings was presented in random order. Following a standard 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 3 and 8 seconds.

Prior to data collection, a short practice period of approximately 3 minutes was provided. The two conditions (tracking and monitoring) were presented three times each in counterbalanced order. Each participant thus participated in six experimental sessions, each session consisting of 50 stimuli. The stimuli of different IOI were randomly intermixed. The number of stimuli of different IOI was equal (25 each). RT scores were calculated for each IOI value in each block, attentional condition and repetition. We thus obtained 300 RT scores from each participant.

During the first part of the experiment, no reference was made at any time to the concept of urgency or to alarms. After the first part of the experiment, participants were informed as to the goal of the study and were asked to perform the subjective task. They had to provide subjective urgency ratings by estimating the urgency of each warning on a continuous (rating) scale labelled ’not urgent’ and ’very urgent’ at the extremes. On each trial, the two warning sounds presented were recorded with the coded value of the cursor (0 at the far left for ‘not urgent’ and 1 at the far right for ‘very urgent’). The entire experimental session lasted about 1 hour.

3.5 Statistical approach

Reaction time was defined as the time interval between the onset of a stimulus and the onset of a response. For each participant, reaction time distributions were recorded. Responses were first analyzed to detect errors and outlying points. Errors included anticipations (RTs less than 100 ms) and misses (failures to respond). Errors were discarded from the RT data and were replaced by the mean of the corresponding condition for each participant. Outliers are defined as reaction times generated by processes that are not the one being studied. These processes can be due to participants’ inattention, or guesses based on participants’ failure to reach a decision. Solutions to the problems of outliers rely on removing observations. Criteria for removing are, however, problematic because real data are inevitably rejected along with spurious data (e.g. Ratcliff, 1993; Ulrich & Miller, 1994). Despite this difficulty, outliers cannot be ignored, especially in paradigms where frequent distraction is known to occur (Osborne & Overbay, 2004). There has been a lot of controversy over whether to remove outliers or not, and over the methods used to remove outlying points (Ulrich & Miller, 1994; Ratcliff, 1979). We thus adopted two different approaches to analyze the RT data. First, we considered data without any form of outlier exclusion. Second, we adopted the most common procedure in RT analysis, i.e. removing observations greater than two standard deviations from the mean (Miller, 1991; see also Cardinal & Aitken, 2006). Missing data (i.e. excluded outliers) were replaced by the mean of the corresponding condition for each participant. The comparison of RT analysis with and without outlier exclusion would lead to a better understanding of the underlying processes involved during the current experiment. The same approach in analysing RT data was adopted for the three experiments presented in this paper.

The RT data –both with and without outlier removals - were averaged across the repetition factor and a 2 (IOI) × 2 (attentional condition) × 3 (block) repeated-measures analysis of variance (ANOVA) was then performed. Participants were treated as a random-effect variable. The remaining variables were treated as fixed-effect variables. To account for violations of the sphericity assumption, *p*-values were corrected using the Huynh-Feldt method. $p<0.05$ was considered to be statistically significant.

In a supplementary step, we analyzed a possible heightened effect of the tracking condition over the three blocks. For each 10-ms period, the positions of the target and pointer were recorded and the distance between them was computed. For each block, the mean and the standard deviation of all these distances were calculated for each participant. We performed an ANOVA to investigate the effect of a possible improvement over blocks.

4 Results

4.1 Reaction times

No anticipations were found and only two misses were observed, one in each of the primary task conditions (tracking and monitoring); they were thus discarded before completing the following analyses.

4.1.1 Analysis on RT data including all observations

Distributions of these data resembled normal distributions (Kolmogorov-Smirnov test, $p>0.05$). The repeated-measures ANOVA comparing reaction times revealed a significant main effect of IOI, $F(1,12)=5.02$, $\varepsilon=1$, $p<0.05$, and a significant main effect of the attentional condition, $F(1,12)=99$, $\varepsilon=1$, $p<0.001$. None of the other effects, i.e. the main effect of Block and all the interactions, were significant ($p>0.1$).

Mean RT decreases when IOI decreases. For both tracking and monitoring condition, the decrease is 10 ms on average (tracking condition: for IOI=100ms, M=396 ms, SD=63 ms; for IOI=300 ms, M=406 ms, SD=74 ms. monitoring condition: for IOI=100 ms, M=297 ms, SD=39 ms; for IOI=300 ms, M=307 ms, SD=48 ms). The data also show that participants had significantly longer reaction times during the tracking condition (M=401 ms; SD=69 ms) than during the monitoring condition (M=302 ms; SD=43 ms). This finding is in line with previous studies (Burt et al., 1995; Bliss et al., 1995).

4.1.2 Analysis on RT data with outlier exclusion

Overall, outliers accounted for 4% of the total number of trials (157 measurements). For the tracking condition, 38 were due to the fast-rate sound and 40 to the slow-rate sound; for the monitoring condition, 33 were due to the fast-rate sound and 46 to the slow-rate sound.

We thus analysed RT data after removing these outliers. Distributions of the RT data resembled normal distributions (Kolmogorov-Smirnov test, $p>0.05$). The repeated-measures ANOVA comparing reaction times revealed a significant main effect of IOI, $F(1,12)=14.74$, $\varepsilon=1$, $p<0.005$, a significant main effect of the attentional condition, $F(1,12)=107.90$, $\varepsilon=1$, $p<0.001$, and a significant interaction between IOI and attentional condition, $F(1,12)=5.8$, $\varepsilon=1$, $p<0.05$. None of the other effects were significant ($p>0.1$).

As in the previous analysis, the data show that mean RT decreases when IOI decreases. However, in the current analysis (i.e. with outlier exclusion), the decrease is larger in the tracking condition (14 ms) than in the monitoring condition (6 ms). This result is consistent with the larger number of outliers observed in response to the slow-rate sound compared to the fast-rate sound during the tracking condition. These observations are confirmed by a significant interaction between IOI and attentional condition (tracking condition: for IOI=100ms, M=376 ms, SD=53 ms; for IOI=300 ms, M=390 ms, SD=58 ms. monitoring condition: for IOI=100 ms, M=292 ms, SD=36 ms; for IOI=300 ms, M=298 ms, SD=38 ms). Figure 2 depicts trimmed mean RT as a function of IOI in the tracking and monitoring conditions. Each bar represents the mean of 975 RT trials (75 per participant).

4.2 Performance on the tracking condition

The data show a significant effect of Block on the mean distance between target and pointer, $F(2, 24)=9.06$, $\varepsilon=0.9$, $p<0.002$, and a significant effect of Block on the standard deviation, $F(2,24)=6.02$, $\varepsilon=1$, $p<0.01$. The mean distance as the standard deviation indeed decreased throughout the blocks. ***[SMc: This last sentence is incomplete. I don't understand it.]*** These effects can be seen as learning effects: performance on the tracking condition became more accurate and increasingly stable over the blocks.

4.3 Urgency judgment

Participants rated signals with an IOI of 300 ms (M = 0.2, SD = 0.1) as sounding less urgent than the one with an IOI of 100 ms (M = 0.7; SD = 0.2), $t(26) = -8.6$, $p \leq 0.0001$. This result was as expected. Previous findings (Edworthy et al., 1991; Hellier et al., 1993) are supported once again.

5 Discussion

One of the issues of this study concerned the comparison between subjective and objective measurements. In their study, Haas and Casali (1995) examined the relationship between perceived urgency and reaction time to warning signals. Their results suggested that there was a correlation between the two measurements: as perceived urgency increased, reaction time decreased. This conclusion was not verified for the inter-pulse interval factor (similar to the IOI factor used in our experiment). The outcome of Experiment 1 thus extends these previous results to the IOI factor. The fact that the urgency judgment and the RT measure both lead to the same conclusion (i.e. the fast-rate sound is the most urgent) does not imply that they share common perceptual and cognitive mechanisms. However, the present experiment does not allow us to conclude more precisely on potentially different mechanisms underlying subjective and objective urgency.

The second aim of this experiment concerned how IOI would affect participants' reaction times under high workload conditions. The outcome of Experiment 1 first showed that there is a straightforward relation between attention and RT: specifically, RTs in the tracking condition are longer than RTs in the monitoring condition. This result provides support for previous findings (Burt et al., 1995; Bliss et al., 1995; for a more theoretical point of view, see Pashler & Johnston, 1998). As already mentioned, it strongly suggests the existence of capacity limits in perceptual analysis. It means that a portion of the capacity is allocated for tracking performance: the more difficult the primary task, the greater the mean reaction time.

More interestingly, a significant interaction between task and IOI was observed. This interaction appeared significant only in the analysis with outlier exclusion, which is coherent with the distribution of outliers as a function of IOI. Indeed, during the monitoring task, more outliers were observed in response to the slow sound, whereas outliers were equally distributed as a function of IOI during the tracking condition. Two points can be highlighted here. Firstly, if we assume that outliers are generated by participants' inattention, we can thus conclude that the slow-rate sound elicited more outliers than did the fast-rate sound, especially in a non-attentional task. Secondly, we observed that IOI has a greater influence on RT when participants are under high attentional demand (simulating a driving situation, for instance). This finding is crucial from an applied point of view. One could argue that the IOI effect observed in this study is too small for any practical purpose. However, this interaction suggests that under a real urgent situation, IOI would have a greater effect, because of the interaction with other cognitive and motor processes that could be in play during a driving situation.

Overall, results obtained in Experiment 1 highlight a main effect of IOI on RT. The IOI effect apparently reflects a general process that can occur under different attentional conditions. RT is lower for the higher IOI value. Loudness could explain this result. Stimuli were designed to vary in terms of IOI and were presented at a level of 76.5 dB SPL. Based on recent studies concerning the loudness of modulated sounds, the two different time-varying sounds presented during the experiment at the same maximum level may have different loudnesses (for reviews, data and models on the loudness of amplitude-modulated sinusoidal carriers, see Moore, Vickers, Baer & Launer, 1999; Glasberg & Moore, 2002). Figure 3 shows main results of the predictions of Glasberg and Moore's (2002) time-varying model. The short-term loudness for the fast-rate sound (IOI = 100 ms) is slightly higher in response to the second pulse (69.4 phons) than in response to the first (68.3 phons): a form of temporal integration occurs. This effect is negligible for the slow rate (IOI = 300 ms), for which the loudness was 68.3 phons for the first and the second pulse. There is a very weak loudness difference between the two sounds. The second pulse of the slow-rate sound and the fourth pulse of the fast-rate sound both occurred at t = 320 ms. As can be seen in Figure 3, short-term loudness at t = 320 ms for the fast-rate sound is higher (69.5 phons) than for the slow-rate sound (68.3 phons). The same values are observed at t = 620 ms and t = 920 ms.

Previous experiments (Chocholle, 1940; Kohfeld et al., 1981; Arieh & Marks, 2003; Wagner, Florentine, Buus & McCormack, 2004) evaluated the relation between loudness and reaction time. Their results showed that RT is closely related to loudness: simple reaction times decreased monotonically when sound intensity increased, and equally loud stimuli produced equal RTs regardless of stimulus frequency.

Thus, the results of Experiment 1 could be explained by a loudness effect more than a specific IOI effect, although the loudness difference between the two sounds is very small. Experiment 2 was performed in order to investigate this possibility.

2 Experiment 2

6 Method

6.1 Participants

Thirteen new participants (five women; mean age ± standard deviation = 26.5 ± 1.7 years) were recruited for this experiment. They were compensated for their participation. None of the participants reported having hearing problems. All of them were right-handed and reported normal or corrected-to-normal vision.

6.2 Stimuli

Loudness equalization was performed on the two stimuli described in Experiment 1. A group of 14 listeners participated in this preliminary experiment. Loudness matches between the fast- and slow-rate sounds were obtained with an adjustment procedure (Buus, Greenbaum and Scharf, 1982). The listener was asked to adjust the comparison stimulus until it seemed equal in loudness to the standard stimulus. The fast-rate sequence was used as the comparison and the slow-rate sequence as the standard. The level of the slow-rate stimulus was fixed at 76.5 dB, as in Experiment 1. The mean level difference at which the slow-rate and fast-rate sequences were judged to be equal in loudness was 5.4 dB SPL.

6.3 Apparatus, procedure and statistical approach

These were the same as in Experiment 1, except that we did not include the urgency judgment. The slow-rate and fast-rate stimuli were presented at 76.5 dB SPL and 71.1 dB SPL, respectively.

7 Results

7.1 Reaction times

No anticipations were found and only three misses were observed during the monitoring condition. They were discarded before conducting the following analyses.

7.1.1 Analysis on RT data including all observations

Distributions of these data resembled normal distributions (Kolmogorov-Smirnov test, $p>0.05$). The repeated-measures ANOVA comparing reaction times revealed a significant main effect of IOI, $F(1,12)=8.99$, $\varepsilon=1$, $p<0.05$, and a significant main effect of the attentional condition, $F(1,12)=41.84$, $\varepsilon=1$, $p<0.001$ (tracking condition: for IOI=100ms, *M*=333 ms, *SD*=42 ms; for IOI=300 ms, *M*=342 ms, *SD*=47 ms. monitoring condition: for IOI=100 ms, *M*=270 ms, *SD*=34 ms; for IOI=300 ms, *M*=276 ms, *SD*=38 ms). None of the other effects were significant ($p>0.3$). These results are similar to those observed in Experiment 1.

7.1.2 Analysis on RT data with outlier exclusion

Outliers accounted for 4.5% of the error-free trials (174 measurements). For the tracking condition, 47 were due to the fast-rate sound and 42 to the slow-rate sound; for the monitoring condition, 37 were due to the fast-rate sound and 48 to the slow-rate sound.

We thus analyzed RT data after removing these outliers. Distributions of the RT data resembled normal distributions (Kolmogorov-Smirnov test, $p>0.05$). The repeated-measures ANOVA comparing reaction times revealed a significant main effect of IOI, $F(1,12)=8.38$, $\varepsilon=1$, $p<0.05$, a significant main effect of the attentional condition, $F(1,12)=43.82$, $\varepsilon=1$, $p<0.001$, and a significant main effect of the interaction between IOI and attentional condition, $F(1,12)=5.04$, $\varepsilon=1$, $p<0.05$ (tracking condition: for IOI=100ms, $M=320$ ms, $SD=61$ ms; for IOI=300 ms, $M=331$ ms, $SD=70$ ms. monitoring condition: for IOI=100 ms, $M=267$ ms, $SD=56$ ms; for IOI=300 ms, $M=269$ ms, $SD=57$ ms). None of the other effects were significant ($p>0.5$). These results are similar to those observed in Experiment 1. Figure 4 depicts mean RT as a function of IOI in the tracking and monitoring conditions. Each bar represents the mean of 975 RT trials (75 per participant).

7.2 Performance on the tracking condition

As explained in Experiment 1, we performed an ANOVA to investigate the effect of learning over blocks. The data show a significant effect of Block on the mean distance, $F(2,24)=4.03$, $\varepsilon=1$, $p<0.05$, and a non-significant effect on the standard deviation between target and pointer, $F(2,24)=1.49$, $p=0.24$. Participants' performance was more accurate across blocks.

7.3 Comparison of Experiments 1 and 2

In order to compare results of Experiment 1 and 2, we performed an additional mixed ANOVA on data with outlier exclusion, with Experiment as between-subjects factor and IOI, attentional condition and block as within-subjects factors. As expected from the separate analysis of Experiments 1 and 2, the mixed ANOVA revealed significant main effects of IOI, $F(1,24)=23.09$, $\varepsilon=1$, $p<0.001$, and attentional condition, $F(1,24)=143.52$, $\varepsilon=1$, $p<0.001$, and a significant interaction between them, $F(1,24)=10.84$, $\varepsilon=1$, $p<0.005$. The ANOVA also revealed a significant main effect of Experiment, $F(1,24)=7.90$, $p<0.01$, and a significant interaction between Task and Experiment, $F(1,24)=6.07$, $\varepsilon=1$, $p<0.05$.

8 Discussion

The major purpose of this experiment was to examine whether we could still observe RT differences in response to stimuli with different IOIs that were equalized in loudness. As noticed before, RT experiments have been suggested to provide an indirect estimate of loudness (Wagner et al., 2004; Arieh & Marks, 2003). If loudness was the only factor responsible for the RT effect observed in Experiment 1, no RT differences should be observed in the current experiment where stimuli are equalized in loudness.

Nevertheless, the most significant feature of this set of results is that they showed similar RT patterns to those of Experiment 1: RTs are smaller for the fast-rate sound than for the slow-rate sound. This is confirmed by a mixed ANOVA, which exhibited the same significant effects as the two separate analyses of Experiments 1 and 2 (the two additional effects related to the Experiment factor could be easily explained by individual differences, which are commonly observed in RT paradigms; see Luce, 1986). Overall, this experiment provides evidence that reaction times are sensitive to an IOI effect with or without small differences in loudness.

This finding indicates that RT is not simply a function of the amount of stimulus energy. A difference between stimulus intensity and stimulus pulse number (equivalent to IOI, here) is at the basis of the “multiple-looks” model of temporal integration in the auditory system (Viemeister & Wakefield, 1991). This model assumes that the nervous system combined multiple “looks” or “samples” of the stimulus to achieve its detection. When a listener is trying to detect a sound, short-term integration is done at a high rate (a few milliseconds): the auditory system takes multiple “looks” at the ongoing signal. A look constitutes the minimum integration time of the auditory system. Information from each individual look is stored in memory and combined later to compute a decision statistic: this is the long-term integration of the process.

This model seems to explain our results, at least qualitatively. The number of sampling opportunities is higher for the fast-rate sound than for the slow-rate one. As the number of samples increases, the probability of detection increases: listeners thus reacted faster to the fast-rate stimulus.

As noticed before, the data of Experiment 2 may also be explained in terms of the number of stimulus pulses. Within this point of view, it might be possible that listeners responded when a sufficient number of pulses occurred within a particular time window. However, as we had to control for the stimulus durations, we are not able to distinguish between these two possibilities of IOI and number of pulses.

In conclusion, the first temporal parameter studied in Experiments 1 and 2 showed a clear and consistent effect on RT, and extended previous results of the urgency literature (Edworthy et al., 1991; Hellier et al., 1993).

3 Experiment 3

In Experiment 3, we repeated the same paradigm as before with sounds varying along a temporal irregularity dimension. Unpredictable event sequences, such as those with rhythmic irregularity, have indeed been suggested to be more attention-getting than predictable regular sequences. However, a clear experimental design and additional results are still needed to support this hypothesis.

9 *Method*

9.1 Participants

Thirteen new participants (six women, mean age ± standard deviation = 27 ± 2.1 years) were recruited for this experiment. They were compensated for their participation. None of them reported having hearing problems. All participants were right-handed and reported normal or corrected-to-normal vision.

9.2 Stimuli

The template for the two different stimuli was a sequence of five short pulses, with a total duration of 540 ms. 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 only varied along a single dimension: temporal regularity. Two temporal patterns were chosen, a regular and an irregular one. The regular one had an IOI of 130 ms. The irregular one was derived from this basic sequence, by displacing the second, third and fourth pulses by 40 ms, which resulted in a lengthening or shortening of the intervals in the sequence. The IOI sequence was: {90; 210; 50; 170} ms.

9.3 Apparatus, procedure and statistical approach

These were the same as in Experiment 1, replacing the IOI factor with the temporal irregularity factor.

10 Results

10.1 Reaction times

No anticipations were found and only eight misses were observed, three in the monitoring condition and five in the tracking condition. They were thus discarded before conducting the following analyses.

10.1.1 Analysis on RT data including all observations

Distributions of these data resembled normal distributions (Kolmogorov-Smirnov test, $p>0.05$). The repeated-measures ANOVA comparing reaction times revealed significant main effects of temporal differences, $F(1,12)=8.98$, $\varepsilon=1$, $p<0.05$, and attentional condition, $F(1,12)=45.79$, $\varepsilon=1$, $p<0.001$ (tracking condition: for the regular sound, M=411 ms, SD=147 ms; for the irregular sound, M=401 ms, SD=132 ms; monitoring condition: for the regular sound, M=296 ms, SD=87 ms; for the irregular sound, M=291 ms, SD=81 ms). It should be noted here that the standard deviations are larger than those in the two previous experiments, and more generally, larger than what could be accepted in a RT paradigm (Luce, 1986). None of the other effects were significant ($p>0.4$).

Figure 5 depicts mean RT as a function of temporal difference in the tracking (panel a) and monitoring (panel b) conditions. Each bar represents the mean of 975 RT trials (75 per participant).

10.1.2 Analysis on RT data with outlier exclusion

Outliers accounted for 5% of the error-free trials (198 measurements). For the tracking condition, 64 were due to the regular sound and 37 to the irregular sound; for the monitoring condition, 54 were due to the regular sound and 43 to the irregular sound.

We analyzed the data after removing these outliers. Distributions of these data resembled normal distributions (Kolmogorov-Smirnov test, $p>0.05$). The repeated-measures ANOVA comparing reaction times revealed a significant main effect of attentional condition, $F(1,12)=51.23$, $\varepsilon=1$, $p<0.001$ (tracking condition: for the regular sound, $M=386$ ms, $SD=78$ ms; for the irregular sound, $M=386$ ms, $SD=77$ ms; monitoring condition: for the regular sound, $M=287$ ms, $SD=49$ ms; for the irregular sound, $M=284$ ms, $SD=47$ ms). None of the other factors (including the temporal differences factor) and interactions were significant (p>0.1).

Figure 5 depicts mean RT as a function of temporal differences in the tracking (panel c) and monitoring (panel d) conditions. Each bar represents the mean of 975 RT trials (75 per participant).

10.2 Performance on the tracking condition

The data show a significant effect of Block on the mean distance between target and pointer, $F(2,24)=8.75$, $\varepsilon=0.8$, $p<0.01$, and on the standard deviation, $F(2,24)=5.29$, $\varepsilon=0.9$, $p < 0.05$. There was still a small effect of learning.

10.3 Urgency judgment

Participants rated regular sounds ($M = 0.49$, $SD = 0.2$) as sounding as urgent as irregular ones ($M = 0.47$, $SD = 0.2$), $t(24) = 0.17$, $p = 0.8$.

11 Discussion

The first outcome of the third experiment was the study of RT in response to temporal irregularity-based stimuli. We asked whether a significant increase or decrease in RT would be observed when the temporal irregularity was manipulated.

The analysis of data with outlier exclusion leads to the conclusion that temporal differences do not elicit any significant improvement on RT. Thus, temporal differences could not be used to modulate urgency of warning sounds. However, the analysis of data including all observations highlights a significant effect of the temporal differences factor. Because outliers (included in this analysis) are the result, in the most common case, of participants' inattention, this result suggests that temporal irregularity captures listener's attention and tends to trigger responses more rapidly. In other words outliers occurred less often in response to irregular than to regular stimuli. Thus, outliers showed a kind of "arousal effect" of temporal irregularity: participants' attention is "captured" by temporally perturbed events.

The Dynamic Attending Theory (Jones, 1976; Jones & Boltz, 1989) proposes a theoretical framework of how listeners respond to and use tempo and time hierarchies. Jones (2004) suggests that the degree of rhythmic regularity of an event time structure affects how effectively we attend to it. This theory is based on the principle of synchrony. The author assumes that people perceive a rhythm in relation to the activity of a small system of internal oscillations. Attending is thus described as the synchronization of these internal attending periodicities with the external rhythm of an auditory sequence. The relationship between this system of internal oscillations and the external rhythm controls listeners' perception of temporal intervals. Jones (2004) considers two aspects of attending: anticipatory attending and reactive attending. First, the listener can anticipate when the next time interval will begin: there is a shift of attending in order to coincide with expected sound onsets. Such anticipations are called temporal expectancies. The second aspect involves expectancy violations. Any deviation observed by the listener from an expected timing is an expectancy violation. Reactive attending involves rapid attentional shifts associated with unexpected sound onsets.

Our results seem consistent with Jones' suggestion that a temporally perturbed event may be more salient than temporally expected ones in a very regular context. For a listener attending to a rhythmic sound pattern, a temporally unexpected sound is considered as an expectancy violation. This "surprise" contributes to a tuning of a listener's attention: the listener is engrossed in reactive attending.

Another finding of the present study that requires comment in light of the literature on perceived urgency is the finding of a non-significant effect of the temporal differences on perceived urgency. Edworthy et al. (1991) suggested that there may be a weak relationship between a syncopated rhythm and perceived urgency. With a non-metric rhythm, we did not observe any difference on perceived urgency between the two sounds.

4 General discussion

Until now, the most dominant approach in urgency studies (e.g. Edworthy et al. 1991; Hellier et al. 1993) has been to consider only the perceived urgency of alarm sounds related to their acoustical properties, such as rate, pitch or timbre. Showing that the ranking of alarms on an urgency scale could be modified by "high-level" parameters, Guillaume et al. (2003) extended these studies by adding an "activation of a mental representation" stage. Indeed, perceived urgency judgments could be modified by cognitive factors like learning effects, cultural influences, participants' strategies and the set of stimuli under study. We thus propose to consider the urgency issue in a more objective way. We designed a new urgency paradigm in order to measure reaction times rather than urgency plus "peripheral" factors (such as learning, cultural influences, etc.). We claimed that, although a correlation could be observed between RT and perceived urgency, we cannot assume that the processes underlying changes in perceived urgency are the same as those underlying changes in RT.

Several models have been proposed to account for auditory processing (McAdams & Bigand, 1994). The present study first focused on more "low-level" stages of the urgency process (Experiments 1 and 2). The observed behavioral performance apparently reflects processes occurring before the computation of perceptual properties or the ‘mental representation’ stage. Experiment 3 completed this study showing that attentional processes may be affected by perceived regularities in the stimulus structure. The current findings thus highlight both stimulus-driven and expectancy-based representations of alarm sounds.

Results of Experiment 1 and 2 show a clear and consistent effect of IOI on RT. It has been proposed that the two necessary components in the RT response are signal detection and response initiation (Green & Luce, 1971). It means that the overall RT measures contain a motor component that hides the relation between the stimulus parameter under study -- IOI -- and the detection ("sensory") component (Kohfeld, 1971). Because the motor-execution stage has been proved to be relatively invariant across changes in perceptual and cognitive task requirements (Miller & Low, 2001), we assume that RT differences are attributable to differences in the durations of sensory and/or cognitive processes rather than to differences in motor time.

Experiment 1 stimuli varied (weakly) along the loudness dimension. Because RTs have already been used as an indirect measure of loudness for tones or noises (e.g. Chocholle, 1940; Wagner et al., 2004), we concluded that this IOI effect observed in Experiment 1 was partially or perhaps entirely due to the loudness difference between the two sounds.

In Experiment 2, we asked whether the IOI effect would be observed in a design in which stimuli were equalized in loudness. Results of Experiment 2 still support the idea of a clear effect of IOI on RT. Stimuli with lower IOI had faster RT. The "multiple-look" model (Viemeister & Wakefield, 1991) seems to explain at least qualitatively our data. "Samples" at the output of initial processing stages in the auditory system are used and selected in order to maximize the goal in that task. If we assume that an improvement in performance of the auditory system leads to faster RTs, results can then be interpreted within this framework.

In Experiment 3, when manipulating temporal differences, we found no RT differences. However, a careful study of outlier data showed that outliers were much more common in response to the regular than to the irregular sound. This result could be interpreted as a capture of a listener’s attention by temporal irregularity. Indeed, rhythmic expectancies generated by hearing regular sequences within the same session might lead to expectancy violations that can capture attending (Jones, 2004).

Another noteworthy result highlighted in the three experiments is the comparison of RT between the tracking and the monitoring conditions. Mean RTs were slower under high attentional demands (tracking condition) than under a simple monitoring condition. This result is consistent with the literature on dual-task paradigms (Pashler & Johnston, 1998). When a task is performed as a switch trial, there is a sizable decrement in performance (increased reaction time). This decrement is called the shift cost (for a review, see Hsieh, 2002).

Interestingly also, the size of the IOI effect is larger in the tracking condition than in the monitoring condition. This plainly demonstrates the potential benefit of using a RT paradigm under attentional conditions simulating real ones, such as a driving situation, for instance. Indeed, RT differences elicited by different IOIs in a real urgent situation could thus be even larger than what we observed under laboratory conditions.

We point out here that a simple reaction time task may simulate only the first part of a more global warning process. In a real situation, even if a fast reaction is required, an analysis of which type of reaction has to occur to respond appropriately is still necessary. This first simple RT model of a real-world warning situation could also explain why the differences between two warning signals are small (around 10 ms). A more complex simulation could lead to greater benefits of one warning sound over another, because of the involvement of more perceptual and cognitive processes. A next step in designing new urgency experiments should then be to use a choice reaction time task, reflecting more appropriately the final application (e.g. “should I brake or accelerate?”).

Whereas the warning sounds presented here have to be studied under more complex simulations of the real application (such as in a car simulator, for example), this simple RT study is a first step in understanding more precisely the warning process. Efficient alarms are those that are, first and foremost, detected and processed more rapidly. Future warnings could then be designed on an IOI scale, with the faster IOI associated with the most urgent warning. Moreover, temporal irregularity could be used as a “wake-up” warning in potential situations of driver inattention.

5 Conclusion

We carried out three experiments in order to better understand and to extend existing recommendations on auditory warning design. We have demonstrated the importance of temporal factors (IOI and temporal differences) by the use of an objective measurement. We have shown that efficient alarms should be designed according to an IOI scale: the shorter the IOI, the more urgent. Moreover, we broaden the possible applications of temporal factors in warning sound design with the temporal irregularity parameter showing an "arousal" effect on participants. Interestingly also, we showed that the size of the IOI effect is larger under high attentional demands, highlighting the benefit of studying alarms under conditions simulating driving situations.

This RT paradigm not only leads to concrete recommendations concerning the design of powerful alarms, but also provides a better understanding of the processes involved. A decrease in IOI leads to a decrease in RT probably because of a temporal integration process in the auditory system. In addition to this “low-level” process, the arousal effect of temporal irregularity could be an illustration of a more “high-level” framework, explaining how listeners respond to and use time hierarchies.

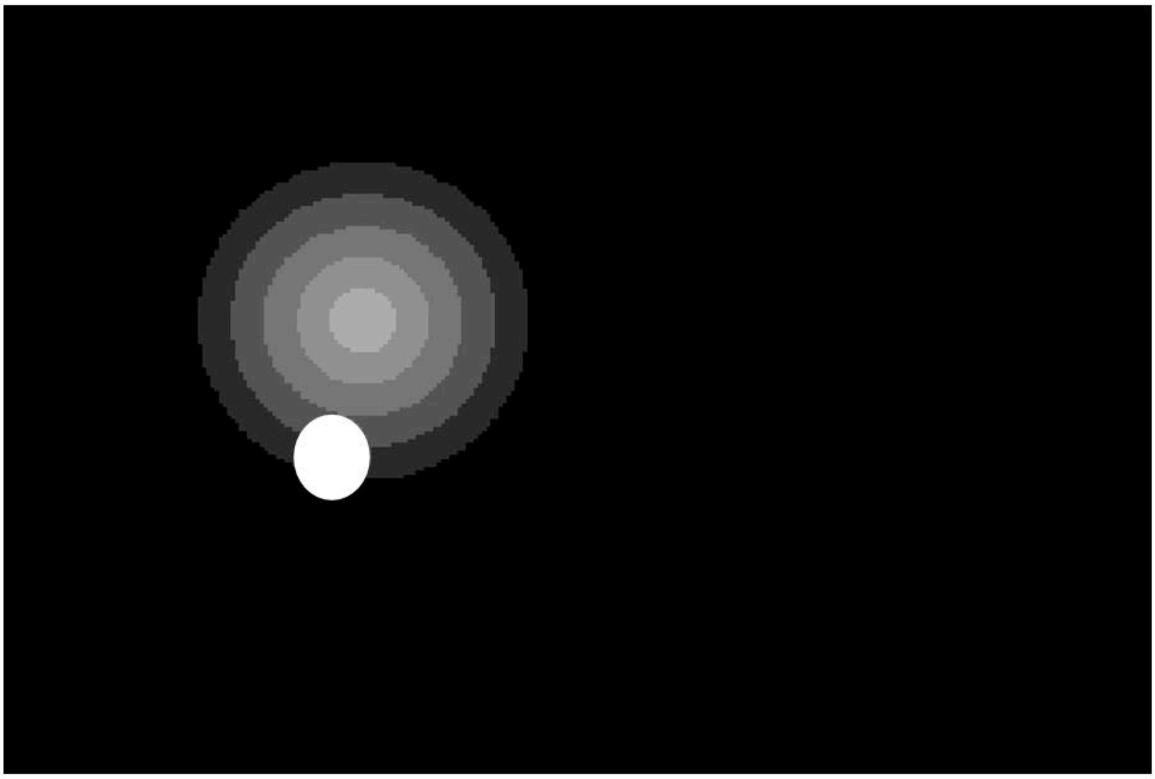

Figure 1. Screenshot of the tracking condition of the primary task used in Experiments 1, 2 and 3. The target is represented by different concentric circles in grey, and the pointer with a white circle (orange in the original experiment). Participants had to track with the mouse (represented by the white circle) the target moving at a constant speed along a random trajectory on the screen.

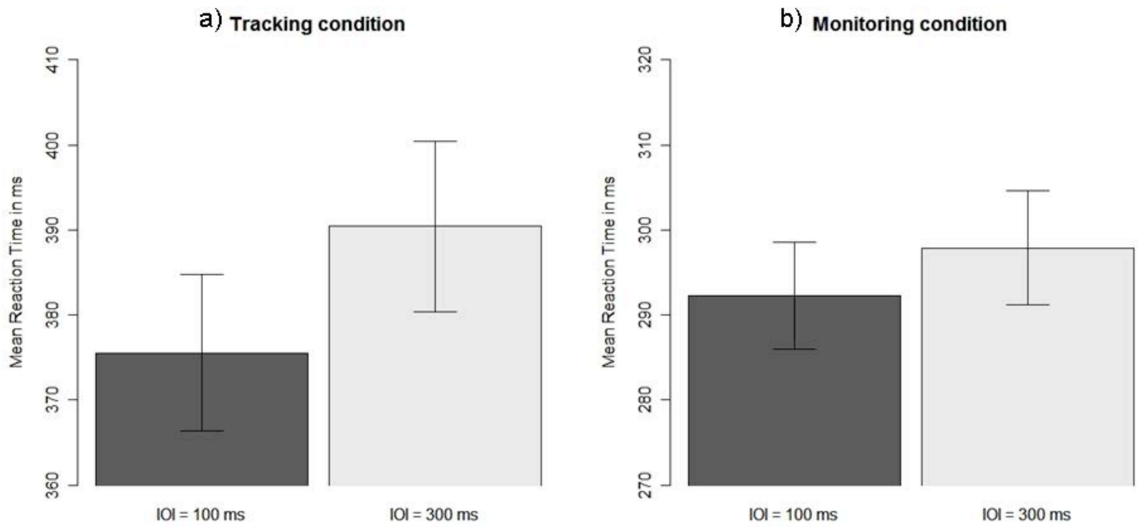

Figure 2. Mean RT as a function of IOI in the tracking (panel a) and monitoring (panel b) conditions (Experiment 1, analysis with outlier exclusion). The vertical bars represent the standard error of the mean.

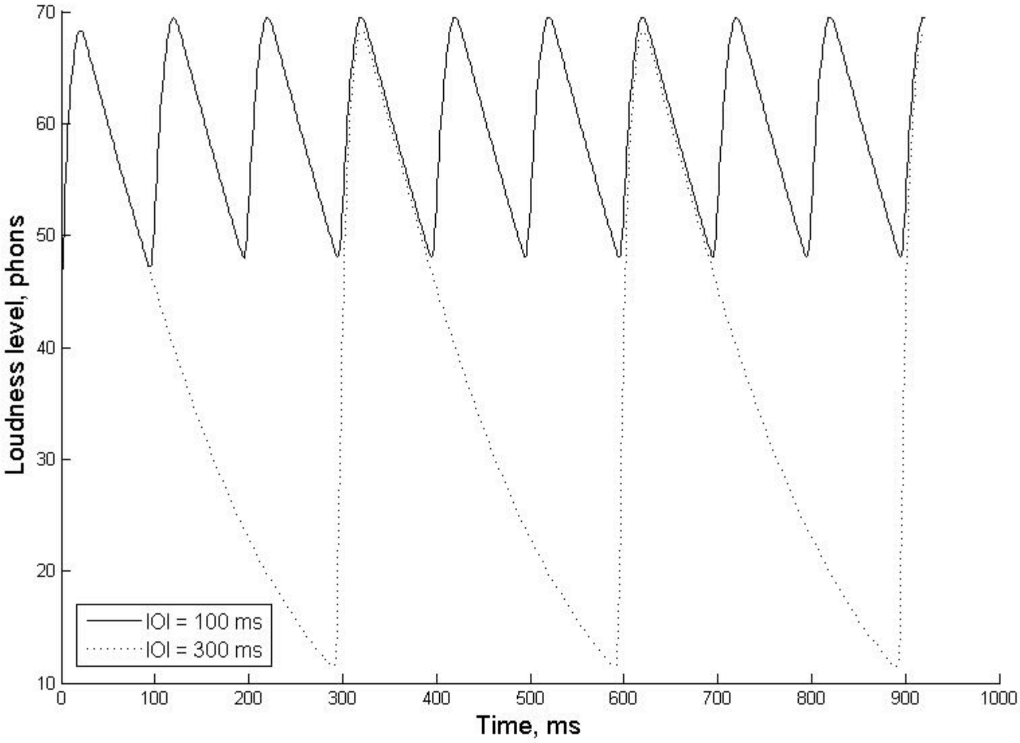

Figure 3. Time-varying model (Glasberg & Moore, 2002) showing short-term loudness as a function of time in response to the two stimuli.

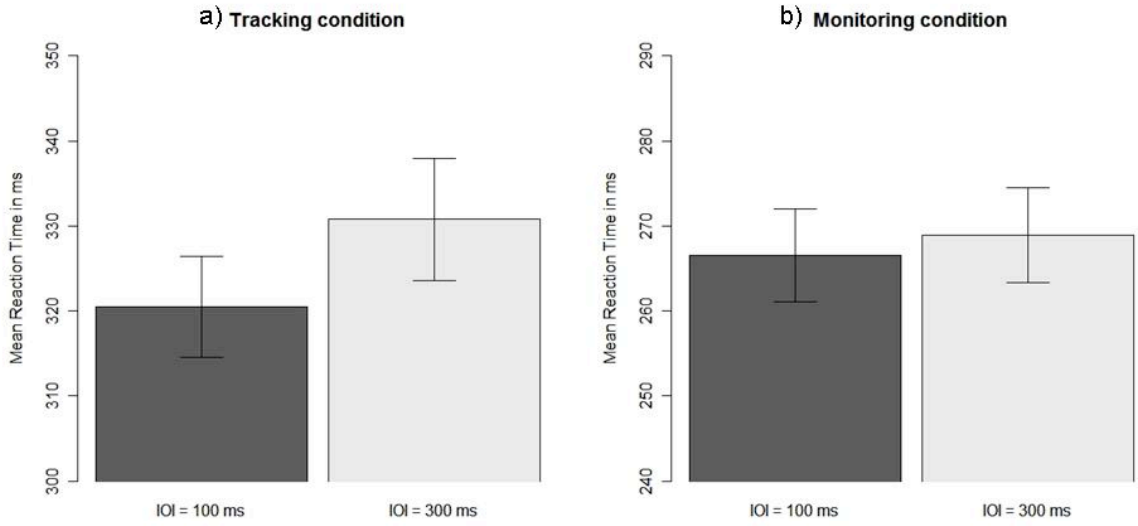

Figure 4. Mean RT as a function of IOI in the tracking (panel a) and monitoring (panel b) conditions (Experiment 2, stimuli equalized in loudness, analysis with outlier exclusion). The vertical bars represent the standard error of the mean.

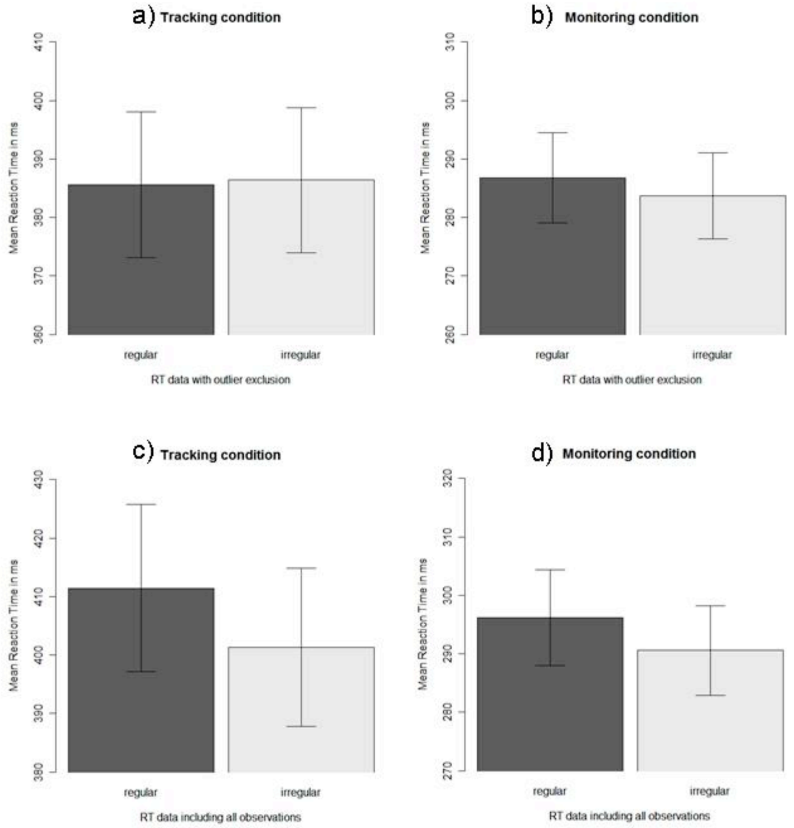

Figure 5. Mean RT as a function of temporal differences (Experiment 3) for the analysis with outlier exclusion (tracking condition, panel a; monitoring condition, panel b) and the analysis including all observations (tracking condition, panel c; monitoring condition, panel d).The vertical bars represent the standard error of the mean.

Acknowledgments

We would like to thank Laurent Pruvost for showing the way with his pilot study performed as an undergraduate research project at IRCAM. We also wish to thank Nicolas Misdariis and Sabine Langlois for their help and Brian Moore for providing the code for the loudness model. We are grateful to Judy Edworthy and two anonymous reviewers for helpful comments on an earlier version of this manuscript.

This work was supported by Renault SA.