# Influence of sound-directed attentional focus on overall loudness ratings

PACS : 43.66.Cb

Patrick Susini, Stephen McAdams
Ircam-CNRS,
1 place Igor Stravinsky
F-75004, Paris, France.

Correspondance should be adressed to: susini@ircam.fr

ABSTRACT : In a "realistic" situation, subjects listen more or less actively to the variations of sounds while they are simultaneously performing other tasks. In this study, the overall loudness ratings on nonstationary 1-kHz pure tones was compared under three listening conditions in which the degree of attention to the sounds was varied. In condition 1, the subjects had to continuously estimate loudness variations. In condition 2, they were not constrained to be continuously attentive. In condition 3, they performed an interfering visuo-motor task during sound presentation. The results show that the overall loudness ratings at the end of the stimuli significantly increased when sound-directed attention decreased.

## INTRODUCTION

Global evaluation of natural sounds of long duration corresponding to sound sequences extracted from urban environments or musical phrases raises the question of the subject's attention in a "realistic" situation. Among various factors involved, the evaluation would depend on the subject's activity during the sound sequence and thus on his or her attention (Jones and Yee, 1994).
In order to differentiate loudness, noisiness, and annoyance ratings, Berglund et al. (1975) introduced three definitions of a listening situation before stimulus presentation. For example, concerning annoyance, subjects were asked how annoyed they would feel when exposed to a given aircraft noise in an imaginary situation phrased as "After a hard day's work, you have just comfortably seated yourself in your chair and intend to read your newspaper".
In the present study, three different listening situations were compared in a loudness estimation task of sounds without meaning. The aim was to examine the influence of attentional activity on overall loudness ratings independently of the meaning of the sound. To this purpose, the degree of attention to the sounds was varied (condition 1, 2 & 3) and nonstationary 1-kHz pure tones were used.

## EXPERIMENT

Procedure
The experiment took place in three stages corresponding to the three task conditions to be compared.

Condition 1. The subjects listened attentively to the sound sequences and continuously estimated the temporal evolution of their instantaneous loudness, associating at each moment a rating equivalent to the perceived level using a cross-modal matching method (CMM) in which a force-feedback lever was adjusted so that the force felt was equivalent to the heard loudness (Susini & McAdams, 2002; Susini et al, 2002). Then, once the sequence was finished, they performed a rating of the global loudness over its whole duration using

the same method. Stimuli were presented once each in random order.

Condition 2. The subjects simply listened to the sound sequences and rated their global impression of loudness over the entire duration at the end of the sequence, as for Condition 1.

Condition 3. In this third condition, a specific experimental protocol was conceived. The subjects performed a visuo-motor task during the presentation of the sound sequences. The instructions given to them insisted strongly that they concentrate on the visual task in order to make as many judgments with as few errors as possible. Three buttons numbered 1, 2, and 3 were presented on the computer screen. They were lit up consecutively for 1-s each in a combination of three numbers (e.g. 2, 1, 3). The same button could light up three times in a row (e.g. 3, 3, 3). Any of the 27 possible combinations could thus be presented at a given moment during the trial. The subject was required to reproduce the button order by performing three key presses on keys 1, 2, and 3 on the computer keyboard as quickly as possible. Immediately following the third keypress, a message appeared indicating the correctness of the response and 300 ms after the last response a new series was presented. The number of series of three numbers reproduced by the subject during the presentation of the sound stimulus depended on the response speed of each subject. This task is not difficult, but requires nonetheless sufficient attentional and memory effort such that the subject is not simply listening to the sound. Regardless of what was occurring for the visuo-motor task, it was abruptly interrupted at the end of the sound sequence. A message then asked the subject to enter the global rating of the loudness of the sequence. The result for each number series as well as the total number of series performed were recorded for each subject.

The third condition corresponds to a more "realistic" situation by introducing an interfering visuo-motor task. In fact, the subjects listen more or less actively to the sounds in the first two tasks, which may not generally be the case when they are in a natural context, simultaneously performing other tasks.

Stimuli

The stimuli presented to the subjects were nonstationary 1-kHz pure tones of variable duration and time-varying level (Figure 1). The variation over time was controlled by synthesis with the ISPW sound card and the Max software package. Three groups of temporal profiles were used. In the first group, the signals increased linearly in level on a decibel scale from 60 to 80 dB SPL over durations of 2, 5, 10 or 20 s. This class of contours will be labeled RAMP, with individual contours notated R2, R5, R10 and R20, respectively, for the four

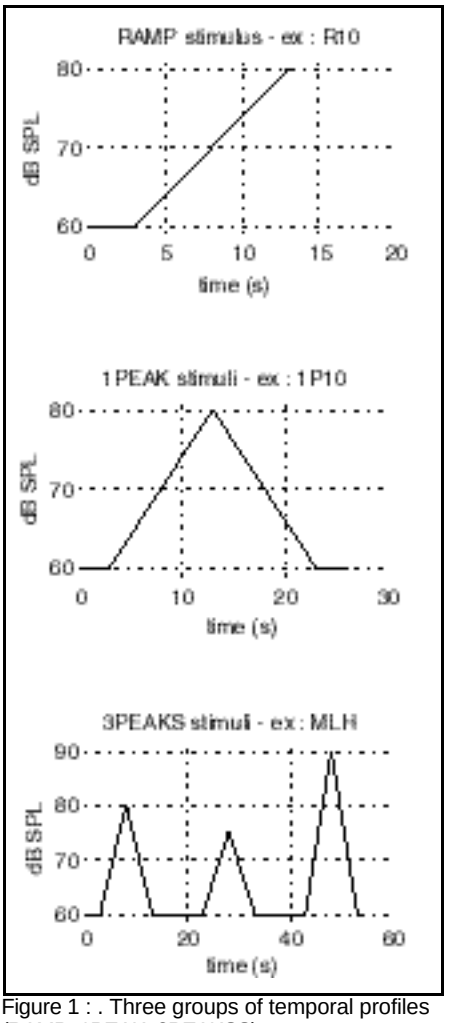

Figure 1 : . Three groups of temporal profiles (RAMP, 1PEAK, 3PEAKSS)

ramp durations. In the second group, the contours were composed of increasing (60 to 80 dB) and then decreasing (80 to 60 dB) ramps of identical duration, similarly to the single ramps of the first group, but of oppositely signed slopes. The duration of increasing and decreasing ramps were 2, 5, 10 or 20 s. This class of contours is labeled 1PEAK with individual contours denoted 1P2, 1P5, 1P10, and 1P20, respectively. The contours of the third group correspond to six combinations of three peaks, the maximum levels of which were 75, 80, and 90 dB SPL, and which are denoted, L (Low), M (Medium), and H (High), respectively. The increasing and decreasing ramps forming each peak were 5 s

in duration. The plateaus between peaks had a duration of 10 s and a constant level of 60 dB SPL. The six combinations correspond to the different permutations of the three peaks: HML, HLM, MHL, LHM, MLH et LMH. This class of contours is labeled 3PEAKSS. All classes of contours start with a 3-s plateau at 60 dB SPL. The 1PEAK and 3PEAKSS contours also end with a 3-s plateau at 60 dB SPL.

Subjects

The same group of 18 subjects participated in conditions 1 and 2. However, only 17 completed the task for Condition 2. Condition 1 was always presented before Condition 2 for subjects performing both. Eleven additional subjects performed Condition 3.

In each condition, the subjects performed six training trials. For each subject, two stimuli from each of the three contour types were chosen at random. Within each condition the sounds corresponding to the three contour types with different durations (RAMP and 1PEAK) or configurations (3PEAKSS) were presented in random order.

## RESULTS

Figures 2-4 present the average of the global ratings obtained for RAMP, 1PEAK, and 3PEAKSS stimuli, respectively. In order to simplify the presentation, we have adopted the following notation: GR1, GR2, and GR3 are the global ratings for Conditions 1, 2, and 3, respectively; Mean(CR) is the mean of the entire continuous rating profile and Max(CR) is the maximum value of the profile obtained in Condition 1. The results of Conditions 1 and 2 are detailed in Susini et al (2002).

Condition 1: Continuous evaluation plus global rating

Sustained attentive listening was required to perform the continuous loudness rating task in this condition. For RAMP stimuli, the mean and maximum values of the continuous rating profiles and the global rating increased as a function of the duration of the ramp. The effect of ramp duration is highly significant ($F(3,51)=11.7$, $p<0.0001$). GR1 is globally less than Max(CR) ($F(1,17)=34.4$, $p<0.0001$), but is equivalent to Mean(CR) ($F(1,17)=1.8$, NS).

The results obtained for 1PEAK stimuli are globally similar in nature to those for RAMP stimuli. The same pattern is found for the Max(CR)/GR1 comparison but is significantly lower than Max(CR) ($F(1,17)=30.0$, $p<0.0001$).

For the 3PEAKSS stimuli, the Mean(CR) and Max(CR) values remain fairly constant for all six configurations. An overall significant difference among configurations for global ratings following continuous estimation was found. Planned contrasts on the effect of configuration with GR1 as dependent variable revealed significant differences between configurations with the high peaks in first and second position ($F(1,75) = 9.5$, $p<0.01$) and in second and third positions ($F(1,75) = 5.7$, $p<0.05$).

Condition 2: Global rating alone

In Condition 2, the subject only gave a global rating after listening to the entire sequence, not being constrained to pay particular attention to the variations in the signal. In general, the ratings obtained are slightly higher than or equal to those for Condition 1. Repeated measures ANOVAs comparing GR1 to GR2 were performed for all three stimulus classes. For RAMP stimuli, GR1 was lower than GR2 ($F(1,16)=6.5$, $p<0.05$). For 1PEAK stimuli, GR1 was equivalent to GR2 ($F(1,16)=1.9$, NS). For 3PEAKS stimuli, the data seem quite different in form, but the differences are not reliable statistically ($F(1,16)<1$, NS). In no case was there a significant interaction with duration or configuration. In summary, the results for Condition 2 without continuous evaluation are roughly equivalent to those found for Condition 1, although the differences are somewhat less contrasted and the inter-subject variability is somewhat larger.

Condition 3: Visuo-motor interference task plus global rating

In Condition 3, subjects had to perform the number series reproduction task simultaneously with the presentation of the sound sequence and make a global rating of its loudness at the end. Globally, all subjects performed the task as instructed. Between 101 and 132 number sequences (124 on average) were completed and the error rate was lower than 4% with the exception of one subject whose rate was around 16%. We can conclude that subjects were attentively involved with this distraction task.

The global rating values in this condition are higher than in the other two conditions. Mixed ANOVAs with between-subjects factor Condition (1 versus 3 or 2 versus 3) and within-subjects factors Duration (for RAMP and 1PEAK stimuli) or Configuration (for 3PEAKS stimuli) were performed on the global ratings. For RAMP stimuli, the Condition factor was significant for both comparisons (1 versus 3: $F(1,27)=13.1$, $p<0.005$; 2 versus 3:

F(1,26)=6.7, p<0.05). For 1PEAK stimuli, Condition 3 was higher than Condition 1 (F(1,27)=9.3, p<0.005) and Condition 2 (F(1,26)=4.4, p<0.05). Finally, for 3PEAKS stimuli Condition 3 was higher than Condition 1 (F(1,27)=8.9, p<0.01) but not significantly different from Condition 2 (F(1,26)=1.6, NS), in spite of the 2-6 dB difference in five of the six configurations. This latter nonsignificant result is due to the large inter-subject variability. In no case did the Condition factor interact significantly with the Duration or Configuration factors.

## DISCUSSION

A summary of the results is presented in Table 1. These results suggest that when listeners attend to the visuo-motor task, they tend to make higher ratings of the perceived level of the stimuli compared to the other conditions. For RAMP and 1PEAK stimuli, GR3 is much closer to Max(CR) than to Mean(CR). For 3PEAKS stimuli, however, ratings in Condition 3 were situated between Mean(CR) and Max(CR) as with Condition 2.

| **Class of stimuli** | **Comparison** |
|---|---|
| RAMP | Mean(CR) < GR1 < GR2 < Max(CR) < GR3 |
| 1PEAK | Mean(CR) ≈ GR1 < GR2 << GR3 Max(CR) |
| 3PEAKS | Mean(CR) GR1 GR2 GR3 << Max(CR) |

Table 1 : Comparisons among conditions.

## CONCLUSION

The less the experimental task requires the listener to focus their attention on the sound sequence, the higher are the global loudness ratings produced. Indeed, the task in Condition 1 constrained the subject to follow the sound in real time and thus to be attentive to the instantaneous variations. In Condition 2, the subject had no constraints and attention could wander, whereas in Condition 3, attention was necessarily focussed on the visuo-motor task and could not be dedicated to tracking the sound sequence. On average, the global estimates increased across these Conditions.

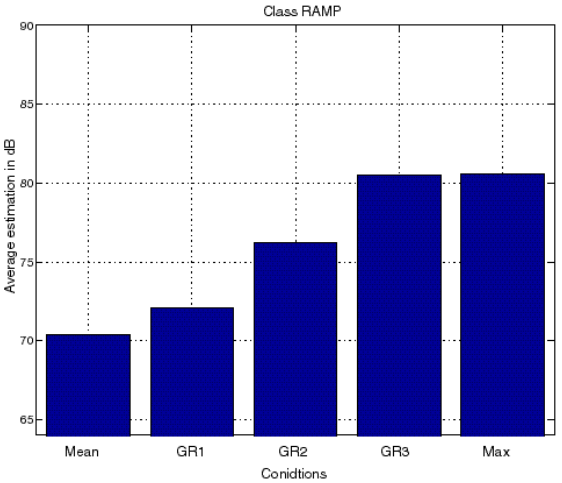


Figure 2 : Results for RAMP stimuli. Cross-modal matching ratings are expressed in equivalent dB (see Susini & McAdams, 2000).

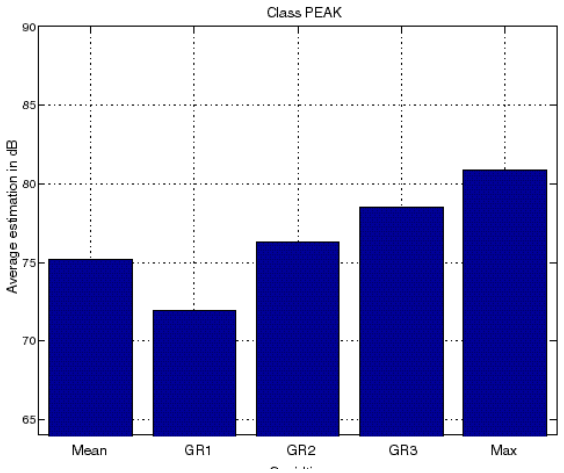


Figure 3 : Results for PEAK stimuli.

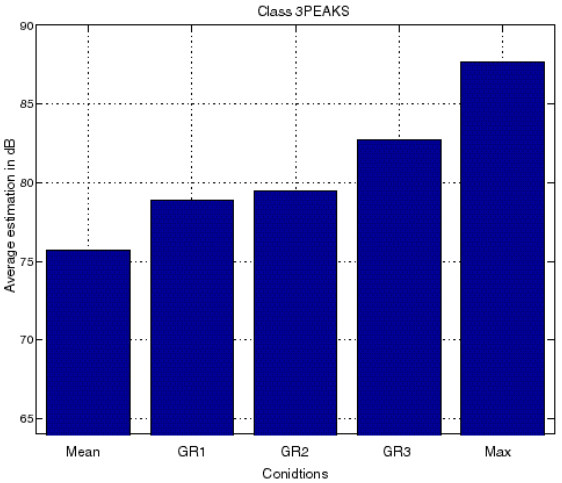


Figure 4 : Results for 3PEAKS stimuli.

**REFERENCES**


Jones, M. R. and Yee W. (1993) Attending to auditory events: The role of temporal organization. In S. McAdams & E. Bigand (eds.), Thinking in Sound: The Cognitive Psychology of Human Audition. Oxford: Oxford University Press, pp. 69-112.

Berglund, B., Berglund, U. and Lindvall, T. (1975) Scaling loudness, noisness, and annoyance of aircraft noise, Journal of the Acoustical Society of America, **57**, 930-934.

Susini, P., and McAdams, S. (2000). "Psychophysical validation of a proprioceptive device by cross-modal matching of loudness." ACUSTICA - acta acustica. **86:** 515-525.

Susini, P., McAdams, S. and Smith, B. K. (2002) "Global and continuous loudness estimation of time-varying levels." ACUSTICA - acta acustica, in press.