LISTENING TO ENVIRONMENTAL SCENES IN REAL TIME

Rivenez Marie[1,2], Drake Carolyn[1], Guillaume Anne[1,2], Detry Sebastien[1]

[1]Laboratory of Experimental Psychology
CNRS and University of Paris V, 71 av Edouard Vaillant, 92774 Boulogne Billancourt, France
[2]Cognitive Science Department
Aerospace Medicine Institute of Army Health Service, Bretigny sur Orge, France

ABSTRACT

Our perception of auditory environmental scenes depends on both the context in which the sources occur and the way in which we listen to them (if we are focused on some sources or not). This study investigates processing differences for a single source depending on its context of occurrence. A tone-detection paradigm adapted to an everyday listening context compares the ability to detect tones in a focused stream within three 10-second non-focused streams: an environmental scene, a white noise and a silence. We predict fluctuations in detection times as a function of the number of streams in the context and of the changes occurring in the non-focused stream.

1. INTRODUCTION

Everyday sound environments are usually extremely complex with multiple sound sources superimposed on background noises. Despite this complexity in the physical signal, our auditory perceptual system is able to "tune into" a conversation in a restaurant or music in a concert room, "filter out" irrelevant information, and "pick up" an incongruous event occurring elsewhere. How is this possible? In this paper we investigate the perceptual mechanisms underlying auditory analysis of environmental scenes. An environmental scene can be composed of sources and background associate with a particular meaning or can be composed of non identifiable sources and backgrounds. We focus particularly on changes in this perceptual organisation over time, its stability when other auditory events occur, and the role of attention in this perceptual organisation. From a practical viewpoint, we assess first the question of how to present sound events in a natural scene in order to make them clearly detectable and understandable and second the question of how non-focused events may disrupt the processing of focused events.

How do we perceptually organise sound events? It is now well established that the auditory system does not analyse sounds passively. Rather, sounds are organised into perceptual units called streams or sources [1]. Many principles of streaming have been demonstrated: Sounds that share similar physical characteristics are grouped into a single perceptual unit (simultaneous grouping) which are then linked over time within a stream (sequential grouping). Many studies have shown that streaming is influenced by the physical properties of the sounds, the result of bottom-up processing. For instance, the frequency and time separations of sounds are involved in streaming [2] [3] [4] [5] [1], as well as timbre, amplitude modulation rate and inter stimulus interval [6] [7] [8]. Streaming is also modulated by top-down processes such as expectancies, predictability [9] [10], learning and attention [11].

What is the role of attention in auditory streaming? This is currently a subject of vigorous debate. Bregman considers that streaming is an automatic, bottom-up process that does not involve attention [2]. On the other hand, many authors [10] [12] have shown that attention can be involved at a very early processing stage in both simultaneous and sequential grouping. In a study by Carlyon and colleagues [12], streaming was not possible when participants performed a separate, concurrent task. They demonstrated a deficit in auditory streaming in the controlateral lesion ear in neglect patient (deficit in attentional processing). This contribution of attention to streaming has been partially confirmed by neuropsychological [12] and electrophysiological studies [13]. Sussman et al. [13] have shown an MMN (electroencephalographic component based on an automatic, preattentive deviant detection system) when two tones were inverted in a melody for fast rates, even when participants were asked to ignore the melody, however, no MMN was elicited when the melody was presented at slow rates. They concluded that attention is necessary for streaming at slow but not at fast rates. Thus, it now seems likely that attention does play a role in streaming, but this role is probably highly task- and sequence-dependant. In the present study, we provide additional evidence of the role of attention in streaming with a new paradigm.

What happens when several streams occur simultaneously? The role of attention in perceptual organisation has been investigated with very simple sequences, but rarely with more complex scenes containing more than two sound sources or potential streams, even though complex scenes are the norm in everyday situations. Brochard et al. [11] examined the question of streaming with two, three and four subsequences (potential streams) by measuring temporal irregularity detection thresholds within one subsequence. These thresholds were taken as an indicator of the quantity of attention required for streaming. Thresholds were lower when there was only one subsequence than when there was more than one subsequence. Also, thresholds were not affected by a subsequent increase in number of subsequences (from two to four subsequences). These results indicate that stream segregation requires attentional effort but this effort is not proportional to the number of potential streams.
Another question emerges from this data: what is the structure of the focused and non-focused streams? Do we perceive one focused stream plus one non-focused stream or one focused stream plus several non-focused streams? Brochard et al. suggested that non-focused streams are not perceptually organised because the physical characteristics (frequency-tempo combinations) of the non-focused subsequences do not affect temporal irregularity detection thresholds, and was therefore not perceptually organised into separate streams. However, the frequency-tempo dimension may not be the most salient for testing the degree of organisation of the non-focused streams.

Furthermore, the weak degree of resemblance of this kind of paradigm to everyday situations may limit their generalizability. We propose a more realistic paradigm in order to test the degree of organisation of the non-focused streams.

Is the perceptual organisation of a focused stream stable, even when something happens in a concurrent non-focused stream? In the light of Brochard et al. results, we suggest an answer to this question. According to these authors, the focused stream should not be affected by changes in the structure of the non-focused stream, because the non-focused streams are not organised in as much depth as the focused one. Furthermore, if attention is necessary for streaming, non-focused streams should not be perceptually organised. Consequently a change in a non-focused stream should not affect the perceptual organisation of the focused stream, especially since non-focused streams are perceptually attenuated [14]. Thus, the focused stream has high perceptual stability as long as the focus of attention does not change stream. One aim of this study was to test the effect of a change in a non-focused stream on the perceptual organisation of the focused stream.

In sum, this research investigates three issues: 1) the effect of attention on auditory streaming, 2) the degree of organisation of the non-focused stream, and 3) the perceptual stability of the focused stream when something happens in the non-focused stream.

Rationale

The time required to detect a tone in a focused stream (tone detection time) was used as an indication of the attentional effort available at each point in time whilst listening to a complex auditory scene. Tones had to be detected in three different contexts: an environmental scene composed of a background and two sources occurring at different moments (three potential streams: tones + background + sources), white noise (two potential streams: tones + background) or a silence (one potential stream: tones). Listeners were instructed to focus on the tones, making the tones the focused stream and the background and sources the non-focused potential streams. This paradigm is a simple task, which does not require any particular effort on the part of the participants. It also provides an on-line measure of perceptual activity over a considerable time span.

Role of attention in streaming

The role of attention in streaming was assessed by looking at the time needed to perceptually organise each scene into streams. We hypothesised that tone detection times would be longer during this perceptual organisation since attentional resources are being directed elsewhere. We therefore predict that, in the contexts with more than one potential stream (environmental and white noise scenes), tone detection times should be longer at the beginning of the scene than after several seconds. Conversely, in the context with only one potential stream (silence), no such changes in detection times should be observed. Once the scenes are perceptually organised into the appropriate number of streams, no additional improvement in detection times should be observed because as soon as the background stream has been organised, it becomes an "old" stream [1], and the perceptual organisation of an old stream does not need to be continually updated if nothing new occurs.

Effect of a change in a non-focused stream on the perceptual organisation of the focused stream

We create changes in the non-focused stream by adding new sound sources that constitute a potential third stream. There are two possibilities. First, if tone detection within the focused stream is affected by changes in the non-focused stream, we can conclude that the non-focused streams have been partially organised. On the other hand, if tone detection within the focused stream is not affected by changes in the non-focused stream, we can conclude that the non-focused stream has not been organised.

2. METHOD

Identification task

An indication of the moment at which participants identified the environmental sounds was obtained with pre- and post-identification tests. Participants listened to the environmental scene used in the main experiment and pressed a key as soon as they had identified each of the sounds (background and two sources). In the main experiment, the scene was always the same[1]. The comparison of the pre- and post-tests provided an indication of learning effects.

Tone detection task

Stimuli. Three 10-second scenes were used: a natural scene, a white noise and a silence. Sounds were sampled at 44100 Hz. The environmental context scene was composed of background noise (a recording of a wood fire) and two sources (someone coughing and someone turning newspaper pages), one occurring 5 seconds (lasting 525 ms) and the other 7 seconds (lasting 950 ms) after the onset of the background noise. The environmental context was filtered at 5000 Hz with a Butterworth filter (order 10 equivalent to a lose of 60 dB per octave). Each sound (fire, coughing and newspaper) was equalized in loudness and presented at 70 dBA. The temporal envelope of this environmental context, which was quite homogeneous over the 10 seconds, was extracted using the Hilbert function in order to construct the noise scene. The noise was obtained by applying a white noise filtered at 5000 Hz (the same band pass as the natural scene) to the temporal envelope of the natural scene. This controlled for amplitude modulations occurring in the natural scene, which could have systematic effects on the tone detection task. It was presented at 70 dBA. The silence context was composed of no sound. The detection tones were 10 KHz pure tones lasting 50 ms (with a rise and decay of 5 ms). A 10 KHz tone is processed by a different filter (filter width = 8075-11025 Hz) from the 5000 Hz filtered scenes (filter width = 4475-5525 Hz) avoiding frequency masking (a difference between the two frequency bands of one octave). Tones were also presented at 70 dBA. An audiogram was performed for each participant before the tone detection paradigm in order to verify that participants could hear a 10-KHz tone.

Apparatus. Participants' task was to press a key each time they detected the tone that occurred randomly during the 10-second stimulus. The tone occurred 4 +/- 1 times in each stimuli. Each stimulus was repeated about 35 times to obtain one detection every 75 ms thus providing a high level of precision in the temporal course of the scene analysis. Between two trials, a sentence appeared inviting the participant to press a key to start the next trial. The three contexts were mixed and repeated three

[1] Only one environmental scene was used throughout the experiment due to the exploratory nature of this study. Subsequent studies are using a wider range of scenes.

times, leading to a total experimental time of about one and a half hours.

Participants. Fifteen volunteers participated to the experiment. They all reported normal hearing.

3. RESULTS

Identification times

Identification times for the background (fire) and the 2 sources (cough and newspaper) were analysed and compared before and after the main experiment. Mean identification times were longer for the background (1075 ms) than for sources (646 ms) ($F(1,12) = 14.39$; $p < .005$). A significant learning effect was found ($F(1,11) = 22.73$; $p < .001$): not surprisingly, identification times were longer before (972 ms) than after (612 ms) the main experiment. This learning effect was of the same magnitude for the three sources (fire, cough and newspaper): no significant differences in percentage decrease of the identification times (mean percentage = 39 % of gain relative to the first identification time).

Descriptive analysis

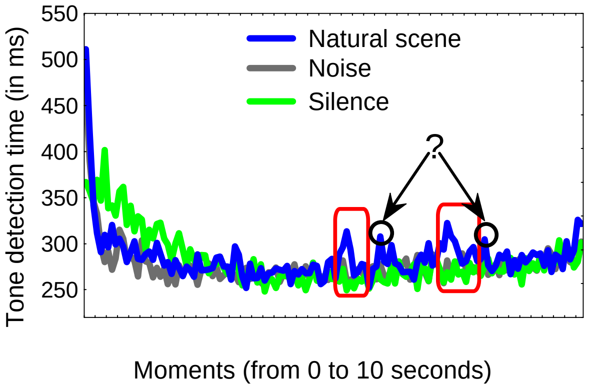

Figure 1: Mean tone detection times (in ms) for the three stimuli (blue curve for the environmental scene; grey curve for the noise; green curve for the silence) as a function of point in time in the 1O-second stimuli. The red boxes represent the two sources (cough and newspaper). The black circles represent unexpected increased detection times.

Figure 1 shows changes in tone detection times (TDT) over the 10-second stimuli for each of the three contexts. First, TDT varied as stream segregation progressed: at the start of the stimuli, TDT were faster for the 1-stream context (silence) than for the 2-stream context (natural scene and white noise). However, after several ms, TDT were slower for the 1-stream context than for the two-stream context. Second, within the environmental scene, TDT increased when a third stream occurred (red box). Surprisingly, TDT increased again around 900 ms after the start of the third stream (new sources - black circles). Three statistical analyses supported this description by focusing on: 1) the beginning of the scene, 2) the effect of the third stream, and 3) the unexpected increased detection times occurring after the additional sources.

Effect of stream segregation for a two-stream scene

Effect of segregation. In order to analyse the time course of streaming, we averaged TDT occurring at different points in time throughout the stimuli[1]. The first and second 150-ms-periods were compared. Figure 2 shows that during the first 150 ms, TDT in the 2-stream context (environmental scene and white noise: m = 413 ms) were slower than in the 1-stream context (silence: m = 356 ms). This difference in TDT started at the stimulus onset and lasted 150 ms. It diminished between 0 and 75 ms: TDT were 121 and 47 ms longer in the 2-stream than in the 1-stream context respectively. This effect was reversed during the next 150 ms: TDT in the 2-stream context were faster (m = 304 ms) than in the 1-stream context (m = 377 ms). A two-way Anova revealed a significant interaction between context and point in time ($F(2,32) = 24.60$; $p < .0001$).

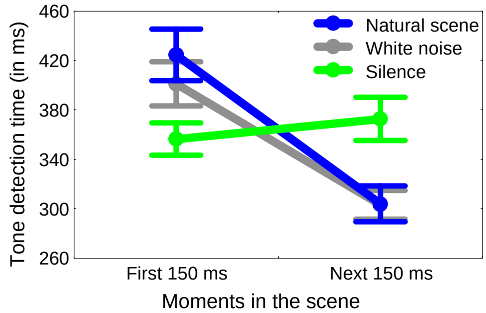

Figure 2: Mean tone detection times (in ms) for the three stimuli (blue curve for the environmental scene; grey curve for the noise; green curve for the silence) as a function of the point in time in the 1O-second stimuli: first and second 150 ms intervals. Error bars indicate standard errors.

Effect of silence context. The descriptive analysis led us to pay attention to TDT in the 1-stream context during the first three seconds of the scene. Figure 1 shows that TDT in the 1-stream context were slower than in the 2-stream context after the 150 first ms. In order to establish how long this effect lasted, we conducted post-hoc analyses (Tukey) which compared the different points in time in the 1-stream context. This statistical analysis revealed that the TDT decreased progressively up to 3075 ms. Beyond this point, the differences between TDT were not significant. The average TDT difference between the 1- and 2-stream contexts was 22 ms. We then compared the slope of the TDT from 150 ms to 3075 ms for the three stimuli. The effect of the number of streams on the slope was significant ($F(2,28) = 45.2$; $p < .0001$). The slope was significantly higher for the 1-stream context (m = -2.13) than for 2-stream context (-0.73 and –0.55 for the environmental scene and the white noise respectively) ($F(1,14) = 73.26$; $p < .0001$). This data revealed that the increase in TDT for silence context was not due to chance, indicating that different processes may be involved in tone detection in 1- and 2-stream contexts.

Effect of scene identification on stream segregation. We were also interested in possible effects of identification on stream segregation. As stated in the introduction, an environmental scene can be composed of sources that are easily identifiable and sources not identifiable. Our paradigm allowed us to test the effect of this factor on segregation, by comparing segregation for the natural context, leading to a specific identification (fire), and the white noise context, leading to no specific identification. TDT were 424 ms and 401 ms for the

[1] These points in time were selected a posteriori after descriptive analyses of the data.

natural and noise contexts respectively. However this difference was not significant.

Effect of stream segregation for the three-stream scene

The second question addressed in this study concerned the effect of a change in the non-focused stream. This change consisted in the appearance of a third potential stream: the two sources occurring in the environmental context. In order to answer this question, we conducted two Anovas: 1) a one-way Anova comparing TDT at different points in time around the sources[1] and 2) a two-way Anova comparing TDT for the sources in the natural scene (3-stream context) and those at the same points in time for the white noise (2-stream context). There were 5 points in time: t_1 = 100 ms before source onset; t_2 = 100 ms after source onset; t_3 = from 100 ms to 325 ms after source onset; t_4 = 100 ms before source offset and t_5 = 100 ms after source offset.

Figure 3 shows the evolution of TDT throughout the stimuli. Consider first the blue curve, representing the natural scene (3-stream context). TDT were stable between t_1 and t_2 (280 ms and 286 ms respectively), increased at t_3 (306 ms) and decreased to 278 and 277 ms for t_4 and t_5 respectively. The Anova revealed a significant main effect of point in time ($F(4,56) = 3.29$; $p < .05$). Planned comparisons revealed no significant differences between t_1, t_2, t_4 and t_5, but a significant difference between t_3 and all other points in time ($F(1,14) = 13.36$; $p < .005$). Thus TDT were slower (26 ms) 100 ms after the start of the source. This increase lasted 225 ms, after this point, TDT returned to baseline.

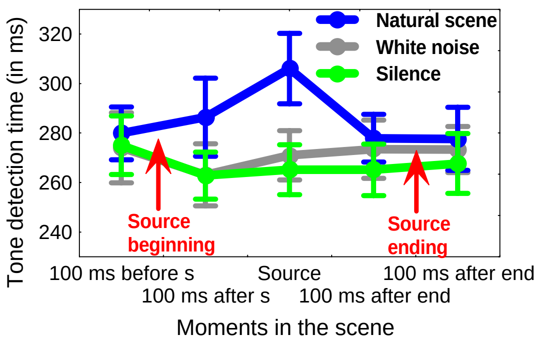

Figure 3: Mean tone detection time(in ms) for the three stimuli (blue curve for the environmental scene; grey curve for the noise; green curve for the silence) as a function of the point in time in the 1O-second stimuli: 100 ms before and after source onset, the following 225 ms, 100 ms before and after source offset. Note that for the white noise and silence, these points in time did not refer to any particular occurrence, but it was necessary to compare the different contexts at exactly the same point in time in order to get the strict effect of the third stream. Error bars indicate standard errors.

In order to verify that this effect was not due to any specific status of this moment in the 10-second stimuli (lose in attention for example), we compared TDT at t_3 in the natural and silent contexts. As expected, TDT were slower in the natural than in the silence context (difference = 41 ms) ($F(1,14) = 34.52$; $p < .0001$).

Finally, to verify that this effect was not due to some specific properties of the temporal envelope of the signal, we compared the increase in TDT at t_3 in the natural context with the same point in time in the white noise context (with the same temporal envelope as the natural scene). As expected, TDT were significantly longer in the natural scene than in the white noise context (difference = 35 ms) ($F(1,14) = 24.1$; $p < .0005$).

Is the 3-stream effect due to stream segregation? These data suggest that the third stream affected the perceptual organisation of the focused stream (TDT were slower when a third stream occurred in the non-focused stream). What is the origin of such an effect? Was it simply due to the fact that the detection of a change in the non-focused stream attracted attention, or was it due to a perceptual re-organisation of the third steam requiring attention? We cannot provide an complete answer to this question, but we can suggest some elements of response. We used the following logic. Increased detection times observed at the beginning of the stimuli were due to stream segregation. We therefore have a precise indication of the size and duration of this stream segregation effect: an increase of about 56 ms (2-stream context = 413 ms; 1-stream context = 357 ms) and lasted around 150 ms. The size of this increase for the third stream was of 35 ms (3-streams context = 306 ms; 2-stream context = 271 ms) (see Figure 4). A Student t-test opposing the effect of a 2-stream context (natural scene and white noise minus silence context for the first 150 ms) to the effect of a 3-stream context (natural scene minus white noise for t_3) was calculated and revealed no significant difference between the two increases. Furthermore, the increase lasted 150 ms after the stream onset for the 2-stream context and it lasted 225 ms (beginning 100 after the source start) for the 3-stream context. These data suggest that the increase in TDT induced by a third stream is due to stream segregation and not to a simple change detection.

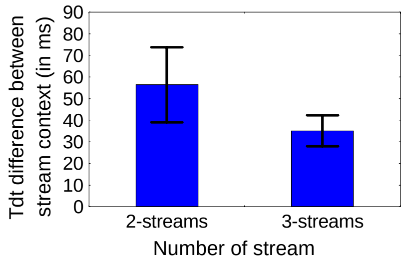

Figure 4: Effect of the number of streams (in ms) on TDT for the 2- and 3-stream contexts. The first bar indicates the difference in TDT between the 2- and 1-stream contexts (natural scene and white noise minus silence condition). The second bar indicates the difference in TDT between 3- and 2-stream contexts (natural scene minus white noise). Error bars indicate standard errors.

TDT increase after source offset

An unexpected effect emerged from the descriptive analysis: TDT increased after the source offset. In order to test the robustness of this effect, we carried out a two-way Anova: number of streams (3-stream vs 2-stream) and source (first and second). This Anova was done for the moments where increased TDT were observed, just after the sources. The two sources

[1] The two sources were averaged after a two-way Anova (2 sources * 5 moments) revealed a significant difference between the two sources ($F(1,14) = 7.68$; $p < .05$): mean tdt were longer for the second source (newspaper : m = 292 ms) than for the first source (cough: m = 279 ms). However there was not a significant interaction between the sources and the point in time, allowing us to collapse over the two sources.

were not averaged because the increase did not occur at the same time relative to the source (325 ms after the first source offset and 75 ms after the second source offset). TDT were longer for the natural scene (3-stream = 307 ms) than for the white noise (2-stream = 285 ms), but this effect was not significant.

4. DISCUSSION

The questions addressed in this study concerned the role of attention in streaming, the degree of organisation of a non-focused stream, and the perceptual stability of the focused stream while a change occur in the non-focused stream. This research produced two main results. Using a paradigm of tone detection in a focused stream, we first showed that tone detection times were longer at the start of the scene. This slowing down was greater when the scene was composed of 2-streams (natural scene and white noise) than of 1-stream (silence condition). Second, we showed that tone detection times slowed when a change occurred in the non-focused stream (appearance of a third stream in the natural scene).

Role of attention in streaming

Our results suggest that stream segregation needs a certain degree of cognitive effort. Actually, tone detection was longer at the beginning of the scene for a 2-stream context than for a 1-stream context. It is more difficult to organise tones as a same event when they are into a stream than when they are presented alone. We think that this segregation requires additional attentional effort, leading to slower reaction times. Our data suggest that stream segregation needs around 120 ms (size of the slow down for a 2-stream context relative to the 1-stream context for the first part of the scene). One could predict that stream segregation should occur at each point in time throughout the 10-second scene, leading to slower tone detection times throughout the scene. However it is not the case. This effect has been well explained by Bregman [1] under the name of “old plus new” phenomenon. When sounds have been organised into a stream that lasts several seconds, the stream becomes "old". If "new" sounds occur on this "old" stream, then they are instantaneously segregated from the "old" stream automatically. Consequently, since the background has been organised, it does not need to be re-organised each time, and tones are instantaneously segregated and their detection is faster. We suggest that a stream requires 150 ms in order to become “old” (duration of the slowing).

Another interesting point is that identification did not seem to affect stream segregation: the increased tone detection times did not differ for the natural and white noise scenes. However, as we used only one natural scene, this argument is rather limited. Other natural scene contexts should be tested in order to assess the effect of identification in stream segregation with this kind of paradigm.

Degree of perceptual organisation of the non-focused stream

The second question of this study concerned the extend to which the non-focused stream is organised. To assess this question, we measured tone detection times in the focused stream when a change occurred in the non-focused stream. This change consisted in the emergence of a new source, creating a new stream. We showed that the emergence of a new stream affected tone detection times giving rise to a 35 ms increase (relative to tone detection times in a 2-stream context, the white noise at the same moment). Thus, it seems that when subjects are focused on a particular stream, they cannot ignore what happens in a non-focused stream. An influence of the non-focused stream on the focused stream have been already studied, but a positive effect has been observed [2] [10]: non-focused streams can catch distracters and help the focused stream to be organised. In other words, similarity of non-focused tones helped focused stream selection. However, our situation is very different: first non-focused streams did not help focused stream detection (no facilitation was observed) and second, it seemed to inhibit the organisation of the focused stream (an increase in TDT was observed). Some studies have shown that non-focused information sustained a perceptual attenuation [15]. Attention may act like a band-pass frequency filter [10]: frequencies belonging to the non-focused stream would be perceptually attenuated and frequencies belonging to the focused stream would be enhanced. In our situation, the third stream belonged to a non-focused frequency (see method). Thus, it should be easy to inhibit any change in the non-focused stream. But it was not the case. The change was detected because tone detection changed precisely when the third stream occurred. The magnitude of the increase in tone detection times for the 3-stream context did not differ from the 2-stream context (at the beginning of the scene) and the duration was also the same (150 ms). This allows us to suggest that similar processes may be involved in these two conditions: the third stream as the background has been streamed. In other words, it is impossible to avoid stream segregation of sounds, even if they are in the non-focused stream.

Nevertheless, our arguments have a potential flaw: the focused stream is not continuous, thus participants were able to listen to and organise the non-focused stream in their own way. A new experiment is investigating this issue using a continuous stream as the focused stream. However, we cannot deny that participants' attention was focused on something other than the third stream, because tones could appear at each point in time. One could think that this contradicts a potential role of attention in streaming: the non-focused stream did not benefit from attention, even it seemed to be perceptually organised. Our interpretation is that our auditory system has a great ability in change detection. This does not require any attentional mechanism; it is definitely automatic. This could explain why the increase in detection times occurred only 100 ms after the start of the source. During this first 100 ms, a change was detected in the non-focused stream, but as this did not need attention, tone detection times did not increase. Finally, we proposed that, after a change was detected in the non-focused stream, attention is automatically enough to organise the new stream.

We also observed a second increase in detection times after the source. We apply the same logic to explain it: the end of the third stream created another change in the non-focused stream. This change was detected and generated a reorganisation of the non-focused stream, again requiring attention. However, this effect was not significant. It might be due to the fact that the disappearance of a source is generally harder to detect than the appearance of a new source (personal communication, Camus).

Perceptual stability of the focused stream

What about the perceptual stability of the focused stream? Our results showed that tone detection times increased when a third stream occurred, meaning that the non-focused stream has been organised. We suggest that the focused stream has lost its perceptual organisation while the third stream was organised. This is explained simply by the fact that attention was required for the third stream. Thus, we can conclude that the focused stream has a limited perceptual persistency: even if it has been organised, if it became an “old” stream and if it has been perceptually enhanced, it needs attention to keep it perceptual organisation.

What happened in the silence context?

An unexpected result was observed in the silence context: tone detection times were longer in the 1-stream context than in the 2-stream context in the first three milliseconds of the scene. In other words, the presence of a second stream facilitated the tone detection times. We interpret this result in terms of perceptually attenuation of the non-focused stream. As we said above, the non-focused stream is perceptually attenuated by 15 dB [14]. This generates a high contrast between the focused and non-focused streams, thus an enhancement of the focused stream. In the 1-stream context, the focused stream was alone, thus it did not benefit from any perceptually enhancement. This might lead to a weak salience of the focused stream, and consequently longer detection times. Nevertheless, detection times for the 1-stream context accelerated progressively and became as quick as in the 2-stream context after three seconds. This can be explained by the fact that after 3 seconds into the scene, most of the to-be-detected tones followed a first one. The first tone acted as a prime by preactivating the relevant frequency.

5. CONCLUSION

One of the interests of this research is that it proposes a new paradigm to study auditory scene analysis. Of course, this paradigm has limitations and needs control experiments, but it allows new discussions on the topic of stream segregation. First it has been demonstrated that attention plays an important role in stream segregation . Second stream segregation seems to be irrepressible even if it occurs in a non-focused stream. And third, the perceptual stability of a focused stream is highly dependent on attention. Applications of this study are rather easy. One of the advantages is that we used a paradigm that can be easily transposed to everyday life situations. In many everyday situations or in some specific work contexts, detecting a particular sound can be of prime importance. The speed of detection can be crucial, for example for pilots who have to react to many alarms in a very noisy context. The principles that we put to the fore should be helpful in auditory display in many ways: first, a sound or a source occurring at a particular moment will not be processed with the same difficulty if attention is required at this moment. Second, the same event will not be processed in the same way if it is presented alone, in a background noise or with other surrounding events. In sum, according to our results and because the auditory world is extremely rich, sounds display must be considered in their context of occurrence.

6. REFERENCES

[1] Bregman, A. S. (1990). Auditory Scene Analysis: The perceptual organization of sound. Cambridge, MA: MIT Press.

[2] Bregman, A. S. & Rudnicky, A. I. (1975) Auditory segregation: stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1 (3)*: 263-7.

[3] Jones, M. R. (1976).Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological review, 83,* 323-357.

[4] Jones, M. R., Kidd, G., & Wetzel, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance, 7(5)*:1059-73.

[5] Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, the Netherlands.

[6] Bregman, A. S., Ahad, P. A., Crum, P. A., & O'Reilly J. (2000). Effects of time intervals and tone durations on auditory stream segregation. *Perception and Psychophysics, 62 (3)*: 626-36.

[7] Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception and Psychophysics; 62 (5)*: 1112-20.

[8] Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *Journal of the Acoustical Society of America; 111 (3)*: 1340-8.

[9] Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns*. Journal of Experimental Psychology: Human Perception and Performance; 21 (2)*: 293-307.

[10] Alain, C., & Woods, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception and Psychophysics; 56 (5)*: 501-16.

[11] Brochard, R., Drake, C., Botte, M. C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: effect of number of simultaneous subsequences and frequency separation*. Journal of Experimental Psychology: Human Perception and Performance;* 25 (6): 1742-59.

[12] Carlyon, R. P., Cusak, R., Foxton, J. M., & Robertson, I. H. (2001). Effect of Attention and Unilateral Neglect on Auditory Stream Segregation. *Journal of the Experimental Psychology: Human Perception and Performance, 27 (1)*, 115-127., *Memory and Cognition, 20*, 1154-1171.

[13] Sussman, E., Ritter, W., & Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology; 36 (1)*: 22-34.

[14] Botte, M. C., Drake, C., Brochard, R., McAdams, S. (1997). Perceptual attenuation of nonfocused auditory streams. *Perception and Psychophysics; 59 (3)*: 419-25.

[15] Spiegel, M. F., & Watson, C. S. (1981). Factors in the discrimination of tonal patterns. III. Frequency discrimination with components of well-learned patterns. *Journal of the Acoustical Society of America; 69 (1)*: 223-30.