



HAL
open science

The shift of attention: Saliency modulates the local vs global processing of auditory scenes in musicians and non-musicians

Baptiste Bouvier, Patrick Susini, Emmanuel Ponsot

► **To cite this version:**

Baptiste Bouvier, Patrick Susini, Emmanuel Ponsot. The shift of attention: Saliency modulates the local vs global processing of auditory scenes in musicians and non-musicians. *JASA Express Letters*, 2025, 5 (1), 10.1121/10.0034822 . hal-04903612

HAL Id: hal-04903612

<https://hal.science/hal-04903612v1>

Submitted on 21 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

The shift of attention: Saliency modulates the local vs global processing of auditory scenes in musicians and non-musicians

Baptiste Bouvier,^{a)}  Patrick Susini, and Emmanuel Ponsot

STMS, IRCAM, Sorbonne Université, CNRS, Ministère de la Culture, 75004 Paris, France

baptiste.bouvier@ircam.fr, patrick.susini@ircam.fr, emmanuel.ponsot@ircam.fr

Abstract: This study addresses how saliency shapes the perceptual organization of an auditory scene. A psychophysical task that was introduced previously by Susini, Jiaouan, Brunet, Houix, and Ponsot [(2020). *Sci. Rep.* **10**(1), 16390] was adapted to assess how the ability of non-musicians and expert musicians to detect local/global contour changes in simple hierarchically-organized tone sequences is affected by the relative saliency of local information in the timbre dimension. Overall, results show that saliency enhanced local processing capacities, at the cost of global processing, suggesting a bottom-up reallocation of attention. Interestingly, for non-musicians, saliency caused a reversal of the basic global-over-local processing prioritization as it is typically observed in expert musicians. © 2025 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

[Editor: Alessandro Altoè]

<https://doi.org/10.1121/10.0034822>

Received: 6 November 2024 **Accepted:** 17 December 2024 **Published Online:** 17 January 2025

1. Introduction

Our senses are constantly inundated by an overwhelming quantity of information distributed over many dimensions including time and space, which our brain must process and organize in order to form a coherent scene of meaningful objects (Bizley and Cohen, 2013). The resulting perceptual organization is often addressed indirectly by characterizing how both bottom-up and top-down processes shape its composition. In the auditory modality, previous works have examined the influence of several bottom-up processes on the perception of complex sound scenes, such as the stimulus characteristics (temporal/frequency characteristics and relationships between auditory streams) shaping auditory scene analysis highlighted by the seminal work of Bregman (1994). The contribution of top-down processes was also addressed, by investigating for instance the effects of attentional capacities, auditory expertise, or prior musical knowledge on perceptual organization (Kaya and Elhilali, 2017; Kondo and Kashino, 2009; Moore and Gockel, 2012; Snyder *et al.*, 2012). Yet, the relative influence as well as the interactions that may exist between these different processes are still not fully understood (Kondo *et al.*, 2017).

A phenomenon typically involved in the perception of complex auditory or visual scenes concerns the hierarchy of processing between local and global levels of stimulus information. Navon (1977) studied the processing of large characters made up of smaller characters. His work showed that global characters are processed faster than local characters and are also less influenced by their characteristics; this primacy of holistic processing of visual information was termed the *global precedence effect*. Mevorach *et al.* (2006) showed that stimulus saliency can affect this initial hierarchical organization of local/global processing; when the local level is made visually more salient, participants perform better in detecting local changes. Thus, it seems that the hierarchical organization of visual information is flexible and that the global precedence effect can be reversed, depending on the most salient level of information. This local/global paradigm was subsequently transposed to the auditory modality. Justus and List (2005) first adapted the stimuli to the auditory modality, switching from spatial visual perception to temporal auditory perception. They proposed melodies of nine successive notes, whose pitch could be modified at the level of a single note (local) or a group of notes (global). With these stimuli, Bouvet *et al.* (2011) showed that participants were faster and more accurate at detecting global variations. Further studies (Ouimet *et al.*, 2012; Black *et al.*, 2017) subsequently confirmed this “global precedence” effect in auditory information processing.

The latest adaptation of this paradigm was proposed by Susini *et al.* (2020). This study showed that the global precedence effect is modulated by musical expertise. It is well known that musical practice is associated with different abilities in the perception and processing of sound information (Herholz and Zatorre, 2012; Talamini *et al.*, 2017). In particular, results from several studies revealed differences between musicians and non-musicians in the organization of auditory

^{a)} Author to whom correspondence should be addressed.

streams in the frequency dimension (van Noorden, 1975; Bey and McAdams, 2002, 2003; Wenhart *et al.*, 2019) or the temporal dimension. With regard to the temporal dimension, Ouimet *et al.* (2012) and Black *et al.* (2017) showed that the global precedence effect was reduced in musicians, and suggested that this reduction was driven by their enhanced processing of local temporal information. This interpretation was further supported and extended by Susini *et al.* (2020) and Susini *et al.* (2023), who showed that the global advantage¹ of non-musicians was reversed not only in expert musicians but also in amateur musicians. Even moderate musical practice thus seems to be associated with the development of enhanced analytical listening skills at the local level (Bever and Chiarello, 1974), and thus favors the detection of musical interval modifications, independently of changes in the melodic contour. This ability to direct attention to the desired level is an example of a *top-down* process at work in auditory information processing: the musician participants could have learnt how to better direct his auditory attention to the desired level of information.

In contrast, the influence of *bottom-up* processes on the hierarchical organization of local/global processing in the auditory modality has not been investigated yet. Can auditory saliency modulate this organization as it was observed in Mevorach *et al.* (2006) for vision? A salient sound is defined as a sound having the ability to capture listener's attention (Itti and Koch, 2001; Tsiami *et al.*, 2016; Zhao *et al.*, 2019; Kaya *et al.*, 2020); this effect is considered a bottom-up process. Loudness is the most obvious auditory feature that helps make a sound salient (Liao *et al.*, 2016; Huang and Elhilali, 2017; Tordini *et al.*, 2016). However, other auditory features also appear to be able to modulate attention (Bouvier *et al.*, 2023; Bürgel *et al.*, 2024; Bürgel and Siedenburg, 2023; Kaya *et al.*, 2020; Tordini *et al.*, 2016): sounds of varying brightness and roughness, edge frequencies or frequency micro-modulations, trigger and modulate attentional capture.

The first question addressed in this work concerns the effect of auditory saliency on the organization of auditory information processing. In particular, we asked whether and to which extent saliency can affect the temporal analysis of a sound scene in the framework of the local-global auditory paradigm introduced in previous studies (Susini *et al.*, 2020; Susini *et al.*, 2023). More precisely in this context, can local saliency be associated with a perceptual reorganization favoring local over global information? Does saliency affect the overall perception of the sound scene? In the present study, the saliency of the local level of auditory sequences was modulated by manipulating a timbre attribute called brightness (related to the sound's spectral centroid). The second question addressed concerns how this potential effect might interact with musical expertise. In other words, is the effect of saliency strong enough to counteract for example the specific "detail-oriented cognitive style of processing" (Wenhart and Altenmüller, 2019) exposed in expert musicians with the local/global paradigm? To address this aspect, the present work involved participants with either no musical training or high musical expertise.

2. Methods

2.1 Participants

Twenty participants were initially recruited. Final inclusion for the study was confirmed once the criteria introduced in our previous studies (see the following) were all met, based on participants' responses to a questionnaire addressing musical experience and abilities. This left 17 participants: 11 non-musicians (four women, mean age: 31.4 ± 11.0 years) and six expert musicians (one woman, mean age: 41.4 ± 15.9 years). The details of the inclusion procedure are reported in Appendix A within the supplementary material.

Non-musicians were participants without any musical training or practice. A more specific questionnaire was addressed to the expert musicians, regarding musical educational history and practice, using an adapted version of the Goldsmiths Musical Sophistication Index (Gold-MSI) questionnaire (Müllensiefen *et al.*, 2014). The criteria used for the group of expert musicians were as follows: participants having solid musical training in French institutions such as the Conservatoire National à Rayonnement Régional (CRR), considering themselves musicians, with daily practice, more than six years of theoretical and instrumental musical learning, and playing with other musicians in bands or orchestral ensembles. Answers to the questionnaire are reported in Appendix A in the supplementary material.

The sample size of the present study was based on sensitivity measures reported in Appendix A in the supplementary material. We also verified that the power of the present study was higher than 0.8.

None of the participants reported hearing problems. They gave written consent prior to the experiment and were remunerated for their participation.

2.2 Stimuli

The structure of the stimuli was very similar to Susini *et al.* (2020) and Susini *et al.* (2023), which is detailed in Appendix C (see supplementary material), except for a few modifications that are reported in the following. Each stimulus consists of nine notes, segmented into three triplets of three notes. The local level is defined as the pitch structure within the triplets, and the global level is the pitch structure formed by the average pitch of the three triplets (see Fig. 1).

As compared to Susini *et al.* (2020) and Susini *et al.* (2023) where the notes were made of pure tones, here the notes follow the harmonic structure of Bouvier *et al.* (2023) and Bouvier (2024): each note having a fundamental frequency f_0 has n harmonics (n in $[1, 20]$), the n th harmonic f_n having a frequency $n \cdot f_0$ and weight $1/n^2$. Thus, a variation in α modifies the sound's spectral centroid (SC), and hence its perceived brightness; for a dull note, $\alpha = 5$, and a bright

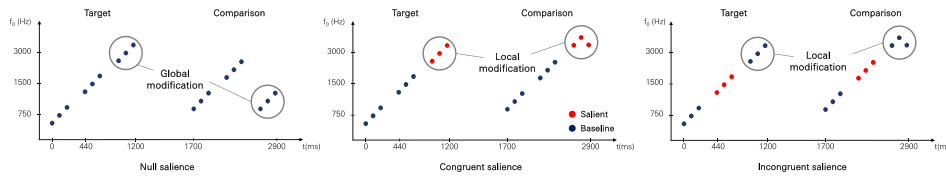


Fig. 1. Examples of stimuli in the three experimental conditions, here illustrated for an ascending profile with a modification on the third triplet, for the three conditions: null, congruent, and incongruent salience. A global upward pitch transposition is observed between the target and the comparison melodies.

note, $\alpha = 1.5$. Note levels are normalized in loudness across all frequencies using the ISO226 equal-loudness curve at 70 dB sound pressure level (SPL) (on MATLAB).

Note duration is 100 ms, intervals between notes within each triplet are 10 ms, and intervals between triplets are 120 ms, giving sequences of 1200 ms. The center of gravity (mean on a log-frequency scale) of the pitch of the second triplet is chosen according to a random uniform distribution ([400–1000] Hz); note that compared to Susini *et al.* (2020) the upper limit of this range is restricted to avoid potentially strident sounds made of very high harmonic frequencies.

The sequences are then structured to respect specific musical intervals: there is always a difference of four semi-tones between two consecutive tones within a triplet, and there is always a difference of one octave between the pitch centers of gravity between two consecutive triplets. This construction of the stimuli on a musical scale was made to best highlight the performance of expert musicians, accustomed to this type of interval (Susini *et al.*, 2020) fitting a diatonic musical scale, largely used in Western music.

2.2.1 Target stimuli

The present study considered the two main temporal profiles employed in previous studies (Justus and List, 2005; Bouvet *et al.*, 2011; Ouimet *et al.*, 2012; Susini *et al.*, 2020), namely, ascending or descending monotonic pitch profiles (hereafter [A] and [D]). Each triplet is indicated by C_j , corresponding to the center of gravity of the fundamental frequencies of the three sounds within a triplet, with j indicating its position in the sequence, from one to three. For each target stimulus, the value of C_2 is first randomly selected from a uniform distribution between 400 and 1000 Hz. Next, the values of C_1 and C_3 are set within ± 1 octave of C_2 .

2.2.2 Comparison stimuli

For the comparison stimuli, C_2 is also chosen from a uniform random distribution between 400 and 1000 Hz. There is therefore always at least one overall pitch transposition (i.e., a transposition of the entire stimulus) between the target and comparison stimuli. Listeners had to compare target and comparison stimuli in terms of pitch profile and had to ignore this overall pitch transposition. This pitch roving procedure ensures that listeners focus on local/global pitch contours rather than the pitch of the sequences per se to make their judgment (Susini *et al.*, 2020). Four types of modifications can additionally be applied:

- no modification (No);
- a local modification (L): the alteration of the pitch profile within a triplet (transposition of a single note within the modified triplet);
- a global modification (G): the alteration of the global pitch profile (transposition of an entire triplet);
- both local and global modification (L + G), applied simultaneously on the same triplet;

Each modification can occur on the first or third triplet, with equal probability.

2.2.3 Salience conditions

On each trial, the pair of stimuli presented may or may not be affected by a salience manipulation: 2/3 of trials contain a salience manipulation.

Different conditions are thus distinguished:

- the *null salience* condition: in 1/3 of the trials, there is no salience manipulation;
- the *congruent salience* condition: in 1/3 of the trials, the modified triplet is made salient;
- the *incongruent salience* condition: in 1/3 of the trials, one of the two non-modified triplets (with equal probability for each) is made salient;

An example is shown in Fig. 1, for an ascending melody with a local modification on the third triplet.

2.3 Apparatus

Sounds were presented to listeners via Beyerdynamic DT-770 PRO headphones (Beyerdynamic, Heilbronn, Germany) and a Focusrite Scarlett 2i2 sound card (Focusrite, High Wycombe, UK). The experimental setup was calibrated at a level of 70 dB SPL using a Brüel & Kjaer 2238 Mediator sound-level meter (Brüel & Kjaer, Virum, Denmark), coupled with the mounting plate provided for circumaural headphones. The experiment took place in an Industrial Acoustics Company (IAC) double-walled soundproof booth. The test interface was coded with Max (v8) on a Mac Mini.

2.4 Procedure

Participants took part in two distinct tasks in separate sessions: a *local* session and a *global* session. In each trial, stimuli were presented diotically to participants in two successive intervals, with one target stimulus followed by one comparison stimulus, separated by 500 ms. Participants were asked to perform a “similar-different” discrimination task, focusing either on the local level or the global level depending on the session. In the local session, they had to determine whether the pitch profiles of the three triplets were similar or different in the target and comparison stimuli, independently of the global profile. In the global session, they had to determine whether the global profile (i.e., the $C_1 C_2 C_3$ organization) was identical or not, independently of the local profiles of each triplet.

At the end of each trial, participants gave their answers by pressing the “similar” or “different” buttons. They had as much time as they wished to respond. Participants were given visual feedback (correct/incorrect) on each trial. The following trial began after a fixed 500 ms delay after each response. The type of session (local/global) was counterbalanced across participants.

Given the four variables—two profiles ([A], [D]), four modification conditions (No, L, G, L + G), two pitch modification positions (first or third triplet), three salience conditions (null salience, congruent salience, incongruent salience)—there were 48 different stimulus configurations.

In order to derive individual scores with reasonable precision, the 48 configurations were repeated ten times each per participant for each session, leading to a total of 480 trials per session and participant. Trials were never identical, as the target height was always drawn from a random uniform distribution. Each session was divided into five blocks of 96 trials and lasted approximately 1 h 30 min. A break was provided after each block, allowing participants to leave the room and relax at their leisure. Before each session, participants were familiarized with the stimuli and the task. First, they were presented with auditory examples and visual analogies created for the specific purpose of the experiment (see [Susini et al., 2020](#)), followed by a block of training trials. The training was validated by the experimenter if participants performed clearly above chance. Participants did not report fatigue effects, and no further learning effects or drops in performance were observed during the main task (see Appendix E of the supplementary material for details).

3. Analyses and results

3.1 Analyses

Two types of analyses were conducted on the results collected in the experiment.

(i) Results were analyzed as a whole following a Signal Detection Theory (SDT) approach to characterize the overall effect of saliency on the perceptual sensitivity in the local and global tasks. We calculated confusion matrices to derive sensitivity (d') and decision criterion (c) values for each participant in each task, as a function of the modalities of the different factors. For each task, the participants' responses were ranked according to the condition:

- Local task: hits = percentage of “similar” responses in conditions No and G; false alarms = percentage of “similar” responses in conditions L and L + G.
- Global task: hits = percentage of “similar” responses in No and L conditions; false alarms = percentage of “similar” responses in G and L + G conditions.

When proportions of Hits or False Alarms were equal to 0 or 100% in a condition, the values were replaced by $1/N$ or $(N-1)/N$ respectively (N being the number of trials) to derive the sensitivity and decision criterion (in line with the analyses of [Susini et al., 2020](#)). The maximum sensitivity that can be reached is therefore 6.2.

Sensitivity results were analyzed based on a $2 \times [2 \times 3]$ factorial design: one inter-participant factor “Group” (Musicians|Non-Musicians) and two intra-participant factors “Task” (Local|Global) \times “Salience” (Null|Congruent|Incongruent).

(ii) Results were also analyzed by contrasting the scores in specific conditions in order to specifically assess the global advantage and the effects of interference between local and global levels. Two indices were computed following [Susini et al. \(2020\)](#) and [Susini et al. \(2023\)](#): the GA (“Global Advantage”) and the GL (“Global-to-Local interference”). If we denote S_{tc} the average score (percentage of correct answers) of a participant in task t and condition c , with l and g referring to local and global tasks/conditions, then:

- the global advantage index was calculated as the difference between global and local global and local scores:

$$GA = 0.5 \cdot (S_{gl} + S_{gg}) - 0.5 \cdot (S_{ll} + S_{lg});$$

- the global-local interference index was calculated as the difference between global-local and local-global interference effects: $GL = (S_{ll} - S_{lg}) - (S_{gg} - S_{gl})$.

3.2 Results

Figure 2 shows the distribution of results from every individual of the two tested groups plotted in the (GA, GL) plane. Only the results of the “null salience” condition are shown in this plot because this is the condition directly comparable to the paradigm used in our previous study (Susini et al., 2020). Despite a non-negligible inter-individual variability within each group, individuals from the non-musicians group mainly cluster in the top-left corner of that plane, while expert musicians are located in the center of the plot. Thus, descriptively, Fig. 2 shows that the results of the “null salience” condition were very similar to those obtained by Susini et al. (2020), replicating previous differences between non-musicians and expert musicians.

A mixed analysis of variance (ANOVA) was performed to assess the effects of the between-subjects factor “group” and the within-subjects factors “task” and “salience” on sensitivity (d'). See Table 1 in Appendix A of the supplementary material.

Results showing sensitivity in each task as a function of the salience conditions, for non-musicians and expert musicians, are presented in Fig. 3. T-tests were performed to compare each of these conditions with the null salience condition. The p -values resulting from the multiple tests then underwent a Benjamini-Hochberg correction (known as “false discovery rate”) and are denoted p_{corr} in the following.

For non-musicians, the presence of locally salient sounds led to a significant degradation of performance in the global task, whether congruent [$T(10) = 4.99, p < 0.001, p_{corr} < 0.001, \text{cohen-d} = 1.51, \text{power} = 1.0$] or incongruent [$T(10) = 4.65, p < 0.001, p_{corr} < 0.001, \text{cohen-d} = 1.4, \text{power} = 0.99$].

In the congruent salience condition, there was a significant improvement in performance in the local task [$T(10) = 2.39, p = 0.019, p_{corr} = 0.025, \text{cohen-d} = 0.72, \text{power} = 0.72$]. In fact, in this condition, their sensitivity was greater ($d' = 1.77$) in the local task than in the global task ($d' = 1.26$). In the incongruent condition, performance in the local task was not significantly affected.

For expert musicians, the presence of locally salient sounds also degraded performance in the global task, significantly in the congruent salience condition [$T(5) = 2.54, p = 0.026, p_{corr} = 0.026, \text{cohen-d} = 1.04, \text{power} = 0.70$]. The effect on the local task was not significant, even in the incongruent salience condition. This result should be treated with caution, as it would appear that the performance of some experts may have reached ceiling performance on this particular task.

The variations in decision criterion (c) induced by salience, albeit significant, remained small compared to the sensitivity values (d'), which indicates that the present results can be primarily accounted for by a change in sensitivity. The analysis of the changes in decision criterion is reported in Appendix D (see the supplementary material).

The effect of salience can be visualized in the (GA, GL) plane introduced previously. Figure 4 plots the distribution of participants according to their group in the null salience condition (dots) and the congruent salience condition (crosses). Almost all individuals from the non-musician group exhibit a clear shift to the left (global advantage dimension); this illustrates the inversion of the global-to-local advantage modification in the congruent salience condition, i.e., a more local bias. Expert musicians also exhibit a shift in the same direction the horizontal dimension, albeit to a lesser extent. More precisely, for non-musicians, the shift to the left of the GA dimension was consistently observed in all the 11 participants tested, and the size of this shift was as larger than 30% in the majority of these individuals. For expert musicians, a similar shift to the left of the GA dimension was consistently observed in all the six participants tested, but here the size

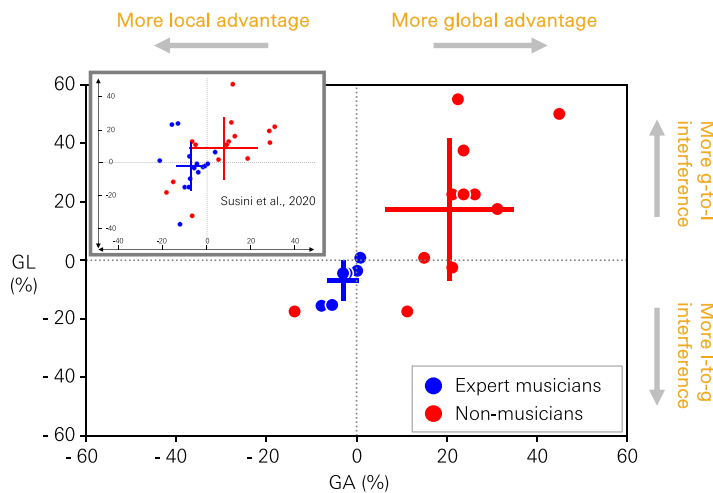


Fig. 2. Individual results obtained in this study, here presented in the (GA, GL) plane introduced in Susini et al. (2020). Error bars, 95% confidence intervals. Top left, results replotted from Susini et al. (2020).

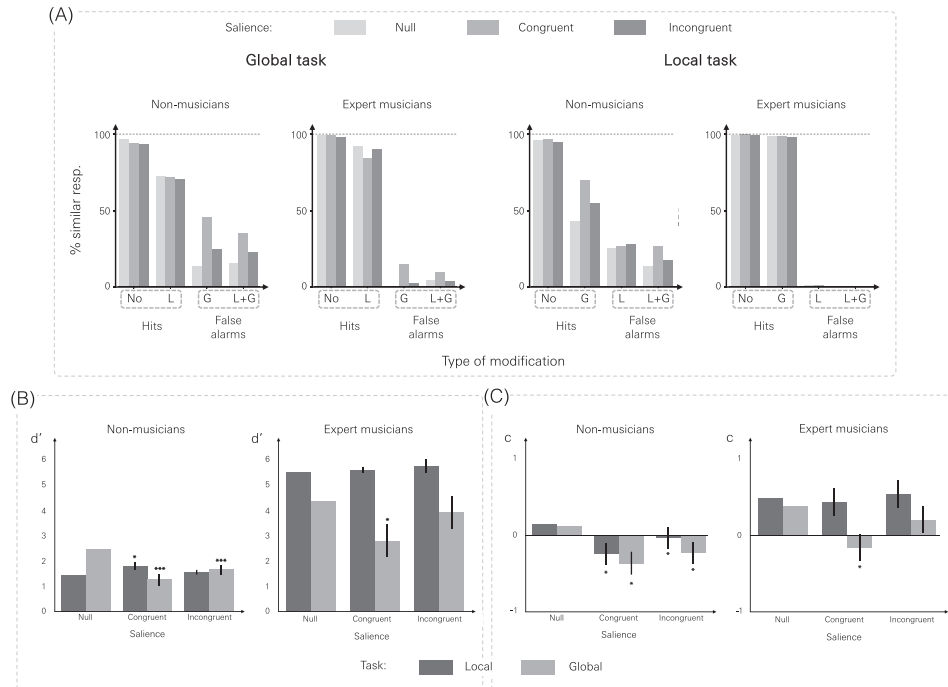


Fig. 3. (A) Percentage of similar responses (further classified as “Hits” and “False Alarms” depending on the condition; see the text) for non-musicians and expert musicians in the two tasks. (B) and (C), sensitivity (d') and decision criterion (c) in the local task (gray) and the global task (light gray) for non-musicians (left) and expert musicians (right) as a function of the salience condition. Error bars: standard error of the distribution of participants’ scores between each condition and the condition with null salience. Significance of differences between each condition and the condition without salience is indicated by stars ($*p_{corr} < 0.05$, $***p_{corr} < 0.001$).

of this shift was much more modest and never exceeded 25%. The fact that all individuals from the two groups exhibit a behavior shift in the same direction with saliency strongly suggests that the observed effect is robust, despite the relatively low sample size of the present study (Ince et al., 2022).

4. Discussion

4.1 Effect of salience on the organization of local/global processing

We observed that the performance of non-musicians in the global task is decreased when the modified triplet is salient, but also when one of the other two triplets is salient. These data suggest that local salience draws attention to the local

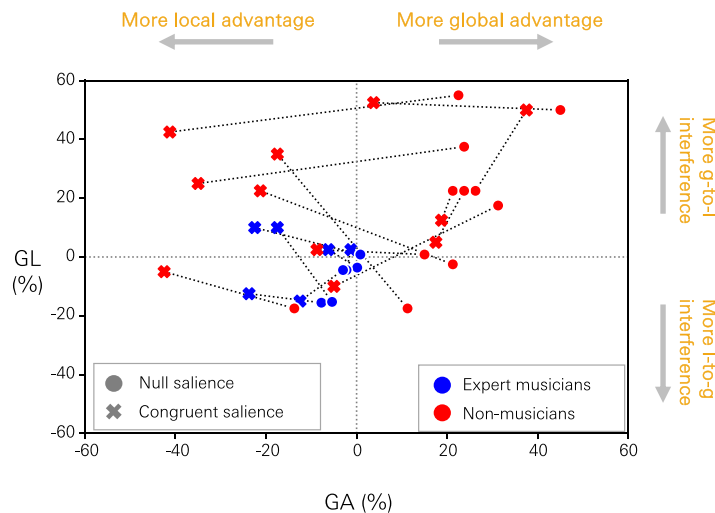


Fig. 4. Distribution of participants in the (GA, GL) plane, in the null salience condition and congruent salience condition.

level, but this “local emphasis” comes at the cost of a reduction in global performance. In return, when salience is congruent with the local modification to be detected, listeners are better at processing local changes. Thus, by appropriate timbre manipulations at the local level, the global advantage initially observed in naive listeners can be reversed. There is thus a reorganization of the auditory local/global temporal processing due to the presence of local salience. This result is consistent with the observed reorganization of spatial information processing in the visual modality by [Mevorach et al. \(2006\)](#).

4.2 Interaction with processes underlying musical expertise

Results for expert musicians in the null salience condition are mostly located in the middle of the (GA, GL) plane (Fig. 2), which suggests that expert musicians are, even when no particular information is made salient, already able to filter out information that occurs at the other level from which their attention is directed to (e.g., local modification in the global task, and vice versa). This result is consistent with those already observed previously ([Susini et al., 2020](#)). Interestingly, our data show a significant influence of local salience in the global task - but only in the congruent condition. Attentional capture at the local level thus appears to disrupt processing at the global level for musicians and non-musicians in a similar fashion. Yet, as compared to non-musicians, the effect of salience on performance in the local task was not significant for expert musicians, but this was likely due to the fact that most individuals already had performance close to ceiling in the null salience condition. That task difficulty was kept identical for expert musicians and the small sample size of this group constitute two aspects that limit the potential generalization of these results. Further studies where the difficulty of the task could be adjusted for each individual, for instance by using more complex/faster tone sequences, would be helpful in addressing this issue. In addition, the decision to provide participants with feedback on their responses (correct/incorrect) was initially made with the idea to help keep them motivated and focused on the task. One may think that this choice could have an influence on the results, for instance, this might have affected the spontaneous response strategy of non-musician participants more than that of musicians who are more confident and robust in their responses. Yet, a specific study comparing the results with and without feedback is needed.

Overall, the present results should be interpreted as a reorganization of priority rules or underlying local/global temporal processing driven by stimuli characteristics.

5. Conclusion and perspectives

This study shows that attentional capture affects the typical global/local hierarchy of temporal sound information processing, by enhancing the processing of the saliency-increased local information. The primacy of holistic processing appears to be reversed when elements are salient at the local level, which parallels previous results in the visual modality ([Mevorach et al., 2006](#)). This effect of salience is also observed in expert musicians, albeit to a lesser extent, and was limited by ceiling performance with our protocol. It is nonetheless interesting to observe that this bottom-up effect would even exacerbate the already more “detail-oriented cognitive style of processing” ([Wenhart et al., 2019](#)) acquired by expert musicians when no particular information is made salient ([Susini et al., 2020; 2023](#)). To conclude, our results demonstrate that bottom-up salience strongly shapes the organization principles of auditory scene analysis. The local/global paradigm provides an interesting tool for future research on this topic.

Supplementary material

See the supplementary material for information about participants, task, and further analyses.

Acknowledgment

E.P. was supported by a grant from the Agence Nationale de la Recherche (ANR-22-CE28-0010). We would like to thank the reviewers as well as the associate editor of JASA-EL, Alessandro Altoè, for their interesting comments and suggestions, and Nicolas Misdariis and Catherine Marquis-Favre for useful comments on this work.

Author Declarations

Conflict of Interest

The authors declare no competing interests.

Ethics Approval

All experiments were approved by the Institut Européen d'Administration des Affaires (INSEAD) IRB, in accordance with the American Psychological Association Ethical Guidelines. Participants gave written informed consent, received financial compensation for their participation, and were debriefed and informed about the purpose of the research after the experiment. All data files and, in particular, the questionnaire have been named anonymously.

Data Availability

The datasets generated during the current study are available at https://github.com/BouvierBaptiste/the_dance_of_attention_local_global_musicians.git.

References

¹“Advantage” can be interpreted as “bias.”

- Bever, T. G., and Chiarello, R. J. (1974). “Cerebral dominance in musicians and nonmusicians,” *Science* **185**(4150), 537–539.
- Bey, C., and McAdams, S. (2002). “Schema-based processing in auditory scene analysis,” *Percept. Psychophys.* **64**, 844–854.
- Bey, C., and McAdams, S. (2003). “Postrecognition of interleaved melodies as an indirect measure of auditory stream formation,” *J. Exp. Psychol. Hum. Percept. Perform.* **29**(2), 267–279.
- Bizley, J. K., and Cohen, Y. E. (2013). “The what, where and how of auditory-object perception,” *Nat. Rev. Neurosci.* **14**(10), 693–707.
- Black, E., Stevenson, J. L., and Bish, J. P. (2017). “The role of musical experience in hemispheric lateralization of global and local auditory processing,” *Perception* **46**, 956975.
- Bouvet, L., Rousset, S., Valdois, S., and Donnadieu, S. (2011). “Global precedence effect in audition and vision: Evidence for similar cognitive styles across modalities,” *Acta Psychol.* **138**(2), 329–335.
- Bouvier, B. (2024). “Sailance auditive: De la caractérisation psychoacoustique à la perception de l’environnement sonore” (Auditory salience: From psychoacoustics to environmental perception), Ph.D. thesis, Sorbonne Université, Paris, France.
- Bouvier, B., Susini, P., Marquis-Favre, C., and Misdariis, N. (2023). “Revealing the stimulus-driven component of attention through modulations of auditory salience by timbre attributes,” *Sci. Rep.* **13**(1), 6842.
- Bregman, A. S. (1994). *Auditory Scene Analysis* (MIT Press, Cambridge, MA), Vol. 198.
- Bürgel, M., Mares, D., and Siedenburg, K. (2024). “Enhanced salience of edge frequencies in auditory pattern recognition,” *Atten. Percept. Psychophys.* **86**, 2811–2820.
- Bürgel, M., and Siedenburg, K. (2023). “Salience of frequency micro-modulations in popular music,” *Music Percept.* **41**(1), 1–14.
- Herholz, S. C., and Zatorre, R. J. (2012). “Musical training as a framework for brain plasticity: Behavior, function, and structure,” *Neuron* **76**(3), 486–502.
- Huang, N., and Elhilali, M. (2017). “Auditory salience using natural soundscapes,” *J. Acoust. Soc. Am.* **141**(3), 2163–2176.
- Ince, R. A., Kay, J. W., and Schyns, P. G. (2022). “Within-participant statistics for cognitive science,” *Trends Cogn. Sci.* **26**(8), 626–630.
- Itti, L., and Koch, C. (2001). “Computational modelling of visual attention,” *Nat. Rev. Neurosci.* **2**(3), 194–203.
- Justus, T., and List, A. (2005). “Auditory attention to frequency and time: An analogy to visual local–global stimuli,” *Cognition* **98**(1), 31–51.
- Kaya, E. M., and Elhilali, M. (2017). “Modelling auditory attention,” *Philos. Trans. R. Soc. B* **372**(1714), 20160101.
- Kaya, E. M., Huang, N., and Elhilali, M. (2020). “Pitch, timbre and intensity interdependently modulate neural responses to salient sounds,” *Neuroscience* **440**, 1–14.
- Kondo, H. M., and Kashino, M. (2009). “Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming,” *J. Neurosci.* **29**(40), 12695–12701.
- Kondo, H. M., van Loon, A. M., Kawahara, J. I., and Moore, B. C. (2017). “Auditory and visual scene analysis: An overview,” *Philos. Trans. R. Soc. B* **372**(1714), 20160099.
- Liao, H. I., Kidani, S., Yoneya, M., Kashino, M., and Furukawa, S. (2016). “Correspondences among pupillary dilation response, subjective salience of sounds, and loudness,” *Psychon. Bull. Rev.* **23**, 412–425.
- Mevorach, C., Humphreys, G. W., and Shalev, L. (2006). “Opposite biases in salience-based selection for the left and right posterior parietal cortex,” *Nat. Neurosci.* **9**(6), 740–742.
- Moore, B. C., and Gockel, H. E. (2012). “Properties of auditory stream formation,” *Philos. Trans. R. Soc. B* **367**(1591), 919–931.
- Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). “The musicality of non-musicians: An index for assessing musical sophistication in the general population,” *PLoS One* **9**(2), e89642.
- Navon, D. (1977). “Forest before trees: The precedence of global features in visual perception,” *Cogn. Psychol.* **9**(3), 353–383.
- Ouimet, T., Foster, N. E., and Hyde, K. L. (2012). “Auditory global-local processing: Effects of attention and musical experience,” *J. Acoust. Soc. Am.* **132**(4), 2536–2544.
- Snyder, J. S., Gregg, M. K., Weintraub, D. M., and Alain, C. (2012). “Attention, awareness, and the perception of auditory scenes,” *Front. Psychol.* **3**, 15.
- Susini, P., Jiaouan, S. J., Brunet, E., Houix, O., and Ponsot, E. (2020). “Auditory local–global temporal processing: Evidence for perceptual reorganization with musical expertise,” *Sci. Rep.* **10**(1), 16390.
- Susini, P., Wenzel, N., Houix, O., and Ponsot, E. (2023). “Psychophysical characterization of auditory temporal and frequency streaming capacities for listeners with different levels of musical expertise,” *JASA Express Lett.* **3**(8), 084402.
- Talamini, F., Altoè, G., Carretti, B., and Grassi, M. (2017). “Musicians have better memory than nonmusicians: A meta-analysis,” *PLoS One* **12**(10), e0186773.
- Tordini, F., Bregman, A. S., and Cooperstock, J. R. (2016). “Prioritizing foreground selection of natural chirp sounds by tempo and spectral centroid,” *J. Multimodal User Interfaces* **10**, 221–234.
- Tsiami, A., Katsamanis, A., Maragos, P., and Vatakis, A. (2016). “Towards a behaviorally-validated computational audiovisual saliency model,” in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 20–25, Shanghai, China, pp. 2847–2851.
- van Noorden, L. P. A. S. (1975). “Temporal coherence in the perception of tone sequences,” Ph.D. thesis 1 (Research TU/e / Graduation TU/e), Institute for Perception Research, Eindhoven, Technische Hogeschool, Eindhoven.
- Wenhardt, T., and Altenmüller, E. (2019). “A tendency towards details? Inconsistent results on auditory and visual local-to-global processing in absolute pitch musicians,” *Front. Psychol.* **10**, 31.
- Wenhardt, T., Hwang, Y. Y., and Altenmüller, E. (2019). “Enhanced auditory disembedding in an interleaved melody recognition test is associated with absolute pitch ability,” *Sci. Rep.* **9**(1), 7838.
- Zhao, S., Yum, N. W., Benjamin, L., Benhamou, E., Yoneya, M., Furukawa, S., Dick, F., Slaney, M., and Chait, M. (2019). “Rapid ocular responses are modulated by bottom-up-driven auditory salience,” *J. Neurosci.* **39**(39), 7703–7714.