

Philippe Lalitte ^{*1}^{*} Université de Bourgogne Franche-Comté, LEAD (UMR 5022), France¹philippe.lalitte@u-bourgogne.fr

Contribution de l'analyse de scène auditive à l'analyse de la performance

RÉSUMÉ

L'analyse de la performance de musiques savantes écrites des XX^e-XXI^e siècles pose des problèmes méthodologiques spécifiques à ce répertoire (Lalitte 2015). Cette communication se donne pour objectif d'examiner en quoi les concepts de l'analyse de scène auditive (Bregman, 1990) peuvent être exploités pour l'analyse de la performance d'œuvres du répertoire précité. McAdams (2015) distingue trois types de groupement impliqués dans la formation d'une scène auditive : 1) le groupement simultané qui permet d'organiser la surface acoustique en événements distincts ; 2) le groupement séquentiel qui intègre en flux auditifs les événements successifs ; 3) le groupement par segmentation qui découpe les flux d'événements en cellules, motifs et phrases. Afin de réaliser les processus de groupement, le système auditif utilise de façon heuristique un certain nombre d'indices pour procéder à ces groupements. On peut supposer que la perception qu'a l'auditeur d'une performance est, en partie, déterminée par l'interaction entre ces processus « ascendants » (automatiques/inconscients) et des processus « descendants » (attention, mémoire à court et long terme). Il s'agira d'une part de déterminer les concordances possibles entre les indices de groupement et les paramètres manipulés par l'interprète (tempo, dynamique, articulation, phrasé, vibrato, timbre, synchronisation, etc.) et d'autre part de comprendre comment ces facteurs interagissent avec les processus descendants pour former une image auditive de la performance. Les hypothèses théoriques seront illustrées par quelques exemples pris dans le répertoire du XX^e siècle.

1. INTRODUCTION

L'analyse de la performance de musiques savantes écrites des XX^e et XXI^e siècles pose des problèmes méthodologiques spécifiques à ce répertoire (Lalitte 2015). Plus que dans un autre répertoire, l'interprète a un rôle crucial pour transmettre à l'auditeur une image auditive (McAdams 1984) la plus cohérente possible avec des œuvres fréquemment qualifiées de perceptivement complexes. Cette constatation est particulièrement cruciale dans la musique polyphonique où le ou les interprète(s) doivent clarifier le plus possible certains éléments de structure à un niveau temporel local ou plus global. De son côté, le musicologue qui étudie la performance musicale (dans ses aspects d'exécution et d'interprétation) peut contribuer à une meilleure compréhension des enjeux perceptifs de l'œuvre et de leurs implications dans le rendu sonore et expressif de la performance. Les principes développés par l'Analyse de Scène Auditive (Bregman 1990, 1994 ; McAdams et Bregman 1979) constituent un ensemble de ressources fiables qui peuvent être transférées à l'analyse de la performance.

Si l'ASA a donné lieu à une exploitation dans les domaines de l'analyse d'œuvres (Lalitte 2009, 2011, 2013 et 2017 ; McAdams, 2015) et de la théorie musicale (Huron 1989 et 2016 ; Moutain 2009 ; Pressnitzer *et al.* 2001), elle n'a pas encore été mise à profit pour l'analyse de la performance. Nous nous sommes donné pour objectif d'examiner en quoi et dans

quelle mesure les concepts de l'ASA peuvent être transférés à l'analyse de la performance d'œuvres des XX^e et XXI^e siècles. Nous avons d'une part cherché à déterminer les concordances possibles entre les indices de groupement et les paramètres d'interprétation (tempo, dynamique, articulation, phrasé, vibrato, timbre, synchronisation, etc.) et d'autre part tenté de comprendre comment ces facteurs interagissent avec les processus descendants (attention, mémoires à court et long terme) pour former une image auditive cohérente de la performance.

Afin d'illustrer les propositions théoriques, nous avons pris quelques exemples dans les *Dix pièces pour quintette à vent* de Ligeti enregistrées par les solistes de l'Ensemble Intercontemporain (Sophie Cherrier, flûte ; Didier Pateau, hautbois ; Jérôme Comte, clarinette ; Jens McManama, cor et Paul Riveaux, basson) le 18 janvier 2013 dans l'auditorium du Musée d'art moderne et contemporain de Strasbourg ¹ dans le cadre d'un projet scientifique du GREAM ². Lors de cet enregistrement, nous leur avons demandé de jouer les pièces dans trois versions : « neutre » (en respectant le texte mais sans intention expressive), « expressive » (comme ils joueraient en concert) et « sur-expressive » (en exagérant) ³. L'un des paradigmes les plus utilisés dans pour l'étude de l'expressivité de la performance consiste à demander à un (ou plusieurs) musicien(s) de jouer des extraits d'une même pièce dans différentes versions. Celles-ci peuvent être toutes expressives, mais il est également possible d'inclure une version « neutre », c'est-à-dire une version qui serait conforme à une exécution stricte de la partition, sans aucune intention expressive. La version neutre sert alors de base pour mesurer les déviations et analyser l'expressivité. Carl Seashore (1938) est le premier avoir émis l'idée que l'expressivité de l'interprète se traduit par des déviations de certains paramètres par rapport à un niveau « neutre ». Nous avons repris ce paradigme expérimental maintes fois utilisé dans les études sur la performance musicale (Barthet *et al.* 2010 ; Canazza *et al.* 1997 ; Davidson 1993 ; Repp 2002 ; Vieillard *et al.* 2011).

¹ Nous remercions chaleureusement les cinq musiciens de l'EIC qui se sont prêtés au jeu d'enregistrer des versions différentes d'une même pièce.

² Le numéro 10 de la revue *Musimédiane*, coordonné par Pierre Michel et moi-même, est consacré aux résultats de ce projet de recherche du GREAM.

³ Aucune autre consigne d'interprétation n'a été spécifiée, ce sont les musiciens eux-mêmes qui ont déterminé plus ou moins explicitement les paramètres d'interprétation qu'ils souhaitaient manipuler pour produire ces trois versions.

2. ANALYSE DE SCÈNE AUDITIVE : INDICES DE GROUPEMENT ET PARAMÈTRES D'INTERPRÉTATION

Les nombreux travaux dans le champ de l'ASA depuis une quarantaine d'années se sont attachés à montrer comment se forment les processus d'organisation auditive dans le cerveau à l'écoute d'un environnement sonore ou plus spécifiquement musical. Selon Bregman, le signal acoustique est traité par deux types de groupement :

- le groupement simultané permet à l'auditeur d'organiser la surface acoustique en événements distincts chacun provenant d'une seule source (ainsi nous fusionnons les harmoniques d'un hautbois pour former une image auditive cohérente et reconnaissable de l'instrument) ;
- le groupement séquentiel permet à l'auditeur d'intégrer les événements successifs produits par une source sonore en un flux auditif distinct (dans le cas d'une musique polyphonique, notre oreille est ainsi à même de séparer les différentes voix et de suivre le sujet d'une fugue).

McAdams (2015) ajoute un troisième type de groupement : le groupement segmentationnel où les flux d'événements sont regroupés en unités perceptives plus larges (motif, phrase, thème, etc.). Cependant, nous ne traiterons ici que des deux premiers groupements dans leurs rapports avec la performance.

Afin de réaliser les processus de groupement, le système auditif utilise de façon heuristique un certain nombre d'indices. La plupart du temps, les indices convergent vers une même solution. Parfois, cependant, les indices entrent en conflit les uns avec les autres créant une image auditive équivoque. Les principaux indices de groupement simultané servent à distinguer les différentes sources entre elles. Le transfert des principes de l'ASA vers l'analyse de la performance nécessite de faire correspondre les paramètres d'interprétation aux indices de groupement. En d'autres termes, quels sont les paramètres utilisés par les interprètes susceptibles d'impacter les indices de groupement ?

Dans l'environnement acoustique naturel, il est par exemple improbable que deux sources sonores distinctes aient des composantes fréquentielles parfaitement synchrones. Du point de vue évolutionniste, il est important pour l'individu de pouvoir distinguer les sources sonores entre elles, notamment dans un environnement hostile où il peut être attaqué par un prédateur ou peut courir un danger provenant d'une catastrophe naturelle. En musique, les choses sont plus complexes. Dans certains cas, les sources doivent être distinctes. L'auditeur doit ainsi pouvoir distinguer les différents instruments ou voix d'un ensemble. Dans d'autres cas, les sources sont appelées à fusionner pour former une seule image auditive. Le compositeur peut ainsi composer des alliages de timbres qui fusionnent afin de créer une sonorité impossible à obtenir avec un seul instrument. L'interprète, quant à lui, va chercher à favoriser, selon les cas, la séparation ou la fusion des sources. Pour cela, il dispose de plusieurs moyens d'action pour agir sur les indices de groupement simultané (Tableau 1) : le synchronisme des attaques peut être contrôlé par la précision rythmique et le degré de synchronisation entre les musiciens, l'harmonicité peut dépendre de la justesse de l'intonation et de l'accord des instruments, la co-modulation des amplitudes peut être impactée par les dynamiques et les coups d'archet, comme la co-modulation des fréquences par le vibrato. Enfin, la disposi-

tion des instruments ou des voix sur scène, notamment la distance qui les séparent, est en mesure d'affecter l'indice de localisation spatiale.

Indices de groupement simultané	Interprète(s)
Synchronisme des attaques	Précision rythmique et synchronisation
Harmonicité	Justesse, accord
Co-modulation des amplitudes	Dynamiques, coups d'archets
Co-modulation des fréquences	Vibrato
Localisation spatiale	Disposition scénique

Tab. 1. Indices de groupement simultané et moyens d'action de l'interprète.

Fig. 1. Début de la pièce n° 7 des *Dix pièces pour quintette à vent* de Ligeti (© Schott 6304).

Prenons l'exemple de la pièce n° 7 des *Dix pièces pour quintette à vent* de Ligeti (Figure 1). Nous sommes en présence de cinq sources (5 instruments) qui tendent à fusionner sur la base d'un cluster chromatique de cinq hauteurs de durée brève joué *sfz*. Le cluster est répété avec les mêmes hauteurs, mais il est varié à partir de la quatrième occurrence par des permutations de timbres.

Les interprètes seront en mesure d'obtenir un son fusionné (une masse sonore homogène) en privilégiant la précision rythmique et l'équilibre des dynamiques entre les 5 instruments. D'après London (2004), au delà de 100 ms, les asynchronismes sont clairement perceptibles : l'auditeur perçoit plus un arpegge qu'un son fusionné. Nous avons mesuré les écarts de temps entre les attaques d'un même accord⁴. Dans les versions expressives, les écarts entre les attaques des cinq musiciens tombent en moyenne en dessous de 100 ms (entre 58 et 80 ms), alors que dans les versions neutres et sur-expressives ils sont supérieurs à 100 ms. Les équilibres dynamiques entre les instruments constituent un autre indice de fusion ou de distinction des sources. Si certaines hauteurs de l'agrégat ressortent trop, relativement aux autres, la fusion perceptive risque d'être perturbée. L'auditeur entend alors deux ou plusieurs sources et non une masse sonore homogène. L'analyse de la saillance de hauteurs sur les sept premiers accords de la pièce n° 7 (Tableau 2) montre que dans la version expressive une seule note ne sort pas (*si*), alors que les deux autres versions sont beaucoup moins homogènes. Dans la version neutre, *la* et *si* bémol sont saillants, alors que *sol* et *sol* dièse sont faibles et *si* est très faible. Dans la version sur-expressive, *si* bémol est très saillant alors que *sol* et *la* sont faibles et *sol* dièse et *si* sont très faibles. Ces indices différencient clairement les trois versions de la pièce n° 7 enregistrées par les musiciens de

⁴ Les analyses du signal ont été réalisées avec MIR Toolbox (Lartillot et Toiviainen, 2007).

l'EIC. Par conséquent, la précision rythmique et l'équilibre des dynamiques ont un impact sur la cohérence de l'image sonore.

	G	G#	A	A#	B
E	0,92	1,00	0,82	0,64	0,08
N	0,66	0,46	0,80	1,00	0,17
S	0,44	0,21	0,55	1,00	0,24

Tab. 2. Saillance des hauteurs (moyenne des sept premiers accords).

Le groupement séquentiel se fonde soit sur le principe de similarité ou de continuité des événements (hauteur, timbre, sonie ou position spatiale), soit sur le principe de proximité temporelle entre les événements (Tableau 3). L'oreille a, par exemple, tendance à grouper dans un même flux auditif les notes conjointes, proches temporellement, non séparées par des silences, de même intensité, de même timbre et provenant d'une même position spatiale. En manipulant les indices de groupement séquentiel, l'interprète a en quelque sorte le pouvoir de rendre plus lisible une polyphonie complexe ou de surligner une voix virtuelle (enchevêtrée dans une ligne mélodique ou à l'intérieur d'une suite d'accords). Afin de manipuler les indices de similarité/continuité, le musicien peut s'appuyer sur des éléments tels que l'articulation, le timbre, la dynamique, l'ornementation, le phrasé, la pédalisation, les coups d'archets, etc. Et pour jouer sur l'indice de proximité temporelle, il peut utiliser des éléments tels que le tempo, le timing, le rubato, les césures et les *fermate*.

Indices de groupement séquentiel	Interprète(s)
Similarité/continuité entre les propriétés des événements	Timbre, dynamique, articulation, phrasé, ornementation, portamento, glissando, pédalisation, coup d'archet...
Proximité temporelle entre les événements	Tempo, rubato, timing, césure, fermata, etc.

Tab. 3. Indices de groupement séquentiel et moyens d'action de l'interprète.

Prenons l'exemple du premier processus de la pièce n° 8 (mes. 1 à 11). La texture est constituée d'une polyphonie à trois voix qui part d'un intervalle commun (tierce mineure, *ré₃/fa₃*), chaque instrument ayant son propre pattern rythmique (Figure 2). Au fur et à mesure que le nombre de hauteurs par voix augmente et remplit l'espace chromatique, les valeurs rythmiques diminuent produisant une texture de plus en plus dense. Le processus se termine lorsque les trois voix se rejoignent sur un trémolo de deux notes (*fa dièse/sol*). Les interprètes ont pour tâche d'accompagner ce processus. Au début, ils peuvent faciliter la formation de trois flux auditifs, les plus distincts possibles, en jouant sur les indices de dissimilarité rythmique et timbrique et de continuité temporelle. À la fin du processus, ils peuvent tirer parti de la similarité fréquentielle et de la proximité temporelle pour conforter la perception d'un seul flux par l'auditeur. L'équilibre dynamique entre les instruments et la stabilité dynamique à l'intérieur de chaque voix contribuent également aux processus de ségrégation et de fusion des flux auditifs.

8
Allegro con delicatezza (sehr gleichmäßig, ohne jede Betonung) (very evenly, without any accentuation)
 (♩ = 72)
 *) fließend, alle neuen Einsätze sehr weich / fluently, every new entry

Flauto
 Clarinetto in Sib
 Fagotto

Fig. 2. Début et fin (mes. 11) du premier processus de la pièce n° 8 (© Schott 6304).

La comparaison des trois versions de cette pièce montre de nettes différences dans l'interprétation. La version expressive maintient la continuité rythmique avec peu de micro-silences et un très bon équilibre dynamique (Figure 3). Dans la version neutre, les quarts de soupir, légèrement plus long, tendent à casser la continuité. Les rythmes moins précis et l'instabilité dynamique diminuent la sensation de continuité. Dans les versions neutre et sur-expressives, la clarinette passe au premier plan, ce qui tend à gêner la ségrégation en trois flux auditifs distincts.

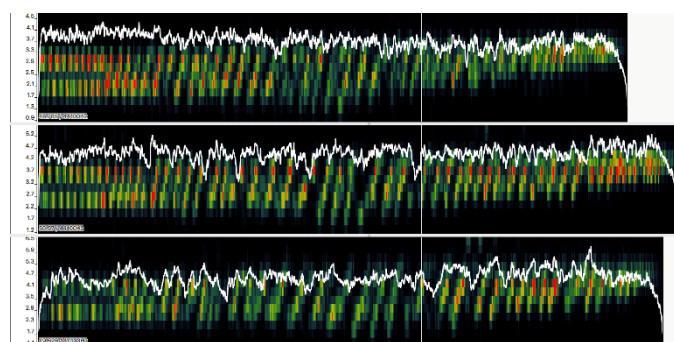


Fig. 3. Premier processus de la pièce n° 8 (mes. 1-11) : versions expressive (en haut), neutre (au milieu) et sur-expressive (en bas). Spectrogramme à Q constant et loudness (courbe en blanc). Figure obtenue avec Sonic Visualiser (Cannam, Landone et Sandler, 2010).

3. ANALYSE GUIDÉE PAR LES PROCESSUS DESCENDANTS

Les processus que je viens de décrire sont dits « ascendants » car ils relèvent d'une analyse primitive effectuée par le cerveau, basée sur des propriétés acoustiques générales, sans hypothèse a priori sur les sources en présence et les structures

musicales impliquées. Mais, selon Bregman (1990) il faut néanmoins tenir compte des processus dits « descendants » associés, pour certains d'entre eux, à un traitement conscient et volontaire. D'une part, toutes les connaissances antérieures stockées en mémoire à long terme dont certaines sont rappelées lors de l'écoute vont former des attentes chez l'auditeur. Celles-ci peuvent d'ailleurs être associées à une interprétation donnée familière au sujet, c'est pourquoi il lui sera parfois difficile d'apprécier une nouvelle interprétation lorsqu'il la découvre. D'autre part l'empan et la vitesse de traitement en mémoire à court terme (ou mémoire de travail), ainsi que les processus attentionnels endogènes guident la focalisation auditive. Ainsi, Van Noorden (1975) a démontré que la ségrégation des flux auditifs peut dépendre de ce que l'auditeur cherche à entendre. Plusieurs auteurs ont trouvé qu'il est plus facile de détecter une mélodie familière intercalée à des sons distracteurs présentés dans le même registre si l'auditeur connaît au préalable cette mélodie (Dowling 1973 ; Dowling *et al.* 1987 ; Bey et McAdams 2002). Par ailleurs, Brochard, Drake, Botte et McAdams (2000) ont montré que l'attention a tendance à se focaliser sur une seule source sonore, les autres voix étant intégrées dans une unité perceptive indifférenciée, et que la focalisation de l'attention sur une voix intérieure ou virtuelle nécessite un effort d'inhibition plus important.

Comment l'interprète peut-il faciliter les processus attentionnels ? Prenons l'exemple de la pièce n° 5 jouée en quatuor (hautbois *tacet*). La texture est formée de trémolos (doubles ou triples coups de langue) et d'accents qui doivent faire émerger les hauteurs d'une mélodie « fantôme » (Figure 4). D'un point de vue de l'analyse de scène auditive, nous sommes en présence d'une compétition entre l'organisation simultanée (les timbres) et l'organisation séquentielle (la mélodie fantôme). L'auditeur peut percevoir cette pièce comme une polyphonie jouée par quatre sources différentes (quatre flux auditifs) ou une mélodie (un flux auditif) émergeant d'un fond indifférencié (une texture granuleuse). La deuxième solution étant celle souhaitée par le compositeur, le rôle des interprètes serait de focaliser l'attention de l'auditeur sur l'organisation séquentielle afin qu'il groupe les notes accentuées en un seul flux auditif (la mélodie). Pour ce faire, les interprètes sont à même, par exemple, d'augmenter l'écart dynamique entre les accents et la texture granuleuse et/ou d'opter pour un tempo le plus rapide possible (tout en sachant qu'il y a ensuite deux changements de tempo à 132 bpm)⁵.

Presto staccatissimo e leggiero - Oboe tacet - Stets Tonrepetitionen, im ganzen Stück.
Tone repetitions throughout the piece.

Fig. 4. Début de la pièce n° 5 (© Schott 6304).

⁵ En augmentant le tempo, on réduit la proximité temporelle entre les notes et de ce fait on facilite le groupement des notes en un seul flux auditif.

L'analyse du tempo et de l'amplitude des accents a été effectuée avec Sonic Visualiser (Tableau 4). Le tempo de la version neutre est le plus lent et les accents ressortent moins nettement de la texture. Par ailleurs, l'équilibre dynamique entre les quatre instruments n'étant pas optimal, l'auditeur aura tendance à se focaliser sur l'organisation simultanée (les timbres séparés). Dans la version sur-expressive, le tempo est au contraire très rapide et la dynamique globalement plus forte. Par conséquent, cette version tend à focaliser l'attention sur l'organisation séquentielle, mais la vitesse étant tellement rapide que les sons de la mélodie fantôme se chevauchent et ne forment pas une image auditive cohérente. C'est dans la version expressive que la différence entre le fond (la texture granuleuse) et la mélodie fantôme produite par les accents est la plus nette. L'auditeur tend cette fois à percevoir un seul flux auditif (la mélodie plus l'arrière plan textural).

	E	N	S
Tempo	131,23 (7,44)	125,33 (4,50)	156,19 (8,15)
Amplitude des attaques	0,031 (0,118)	0,020 (0,107)	0,019 (0,116)

Tab. 4. Moyennes et écart-types du tempo et de l'amplitude des attaques de note.

4. CONCLUSION

On peut supposer que la perception qu'a l'auditeur d'une performance est, au moins en partie, déterminée par l'interaction entre les processus « ascendants » (automatiques) et les processus descendants, volontaires, qui dépendent des traitements attentionnels et mnémoniques. Il semble donc que l'interprète puisse jouer un rôle facilitateur pour la perception d'une image auditive cohérente d'œuvres perceptivement complexes en manipulant plus ou moins explicitement les indices de groupements simultanés et séquentiels, voire les indices de groupement segmentationnel. L'analyse de la performance pourrait, par conséquent, s'appuyer à la fois sur les principes de groupement de l'analyse de scène auditive et sur les théories de l'attention et de la mémoire pour mieux comprendre les enjeux perceptifs de œuvres et pour mettre en évidence l'impact perceptif de telle ou telle performance. Au-delà de l'analyse musicale, le recours à cette approche peut profiter à la pédagogie et la pratique de la performance, notamment dans le champ de la musique après 1945.

MOTS-CLÉS

Analyse de la performance, analyse de scène auditive, groupements auditifs, image auditive, mémoires à court terme et à long terme, musique savante écrite des XX^e et XXI^e siècles, perception, processus attentionnels.

RÉFÉRENCES

- BARTHET (Mathieu), DEPALLE (Philippe), KRONLAND-MARTINET (Richard) et YSTAD (Sølvi), 2010, « Acoustical Correlates of Timbre and Expressiveness in Clarinet Performance », *Music Perception*, 28, n° 2, p. 135-154.
- BEY (Caroline) et MCADAMS (Stephen), 2002, « Postrecognition of Interleaved Melodies as an Indirect Measure of Auditory Stream Formation », *Perception & Psychophysics*, 64, n° 5, p. 844-854.
- BREGMAN (Albert S.), 1990, *Auditory Scene Analysis*. Cambridge (MA), The MIT Press.

- , 1994, « L'analyse des scènes auditives : l'audition dans des environnements complexes », dans MCADAMS (S.) et BIGAND (E.), dir., *Penser les sons : psychologie cognitive de l'audition*. Paris, Presses Universitaires de France, p. 11-39.
- BROCHARD (Renaud), DRAKE (Carolyne), BOTTE (Marie-Claire), et MCADAMS (Stephen), 2000, « Perceptual Organization of Complex Auditory Sequences : Effect of Number of Simultaneous Sequences and Frequency Separation », *Journal of Experimental Psychology : Human Perception and Performance*, 26, n° 6, p. 1742-1759.
- CANAZZA (Sergio), DE POLI (Giovanni), RINALDI (Stefano) et VIDOLIN (Alvise), 1997, « Sonological Analysis of Clarinet Expressivity », dans LEMAN (M.), dir., *Music, Gestalt, and Computing : Studies in Cognitive and Systematic Musicology*. Berlin/Heidelberg, Springer-Verlag, p. 431-439.
- CANNAM (Chris), LANDONE (Christian) and SANDLER (Mark), « Sonic Visualiser : An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files », dans DEL BIMBO (A.), CHANG (S.-F.) et SMEULDERS (A.), *Proceedings of the 18th ACM international conference on Multimedia*. New York (NY), Association for Computing Machinery. (<<https://dl.acm.org/doi/proceedings/10.1145/1873951#issue-downloads>>, accédé le 27/03/2023.)
- DAVIDSON (Jane W.), 1993, « Visual Perception of Performance Manner in the Movements of Solo Musicians », *Psychology of Music*, 21, p. 103-113.
- DOUGLAS (Chelsea), NOBLE (Jason), MCADAMS (Stephen), 2016, « Auditory Scene Analysis and the Perception of Sound Mass in Ligeti's *Continuum* », *Music Perception*, 33, n° 3, p. 287-305.
- DOWLING (Jay), 1973, « The Perception of Interleaved Melodies », *Cognitive Psychology*, 5, p. 322-337.
- DOWLING (Jay), LUNG (Kitty Mei-Tak), et HERRBOLD (Susan), 1987, « Aiming Attention in Pitch and Time in the Perception of Interleaved Melodies », *Perception & Psychophysics*, 41, n° 6, p. 642-656.
- HURON (David), 1989, « Voice Denumerability in Polyphonic Music of Homogeneous Timbres », *Music Perception*, 6, n° 4, p. 361-382.
- , 2016, *Voice Leading. The Science behind a Musical Art*. Cambridge (MA), The MIT Press.
- LALITTE (Philippe), 2009, « Rythme et espace chez Varèse », *Fili-grane*, n° 10, p. 247-265.
- , 2011, « Acoustique et perception dans l'esthétique de Marc-André Dalbavie », dans REYNAUD (C.), dir., *Color Marc-André Dalbavie*, Scérén CNDP-CRDP, p. 30-34.
- , 2013, « La totalité mélodique chez Varèse », dans Bosis (B.), dir., *Mélodie et fonction mélodique comme objet d'analyse*, Le Vallier, Éditions Delatour, p. 149-170.
- , 2015 *Analyser l'interprétation de la musique du xx^e siècle. Une analyse d'interprétations enregistrées des Dix pièces pour quintette à vent de György Ligeti*. Paris, Hermann.
- , 2017, « Temps, texture et analyse de scène auditive dans *Eight Lines* de Steve Reich », *Musicologies Nouvelles*, 4, p. 66-76.
- LARTILLOT (Olivier) et TOIVIAINEN (Petri), 2007, « A Matlab Toolbox for Musical Feature Extraction From Audio », Actes du *10th International Conference on Digital Audio Effects*. Bordeaux. (<https://dafx.labri.fr/main/dafx07_proceedings.html>, accédé le 27/03/2023.)
- LONDON (Justin), 2004, *Hearing in Time : Psychological Aspects of Musical Meter*. Oxford/New York, Oxford University Press.
- MCADAMS (Stephen) & BREGMAN (Albert S.), 1979, « Hearing Musical Streams », *Computer Music Journal*, 3, n° 4, p. 26-43.
- MCADAMS (Stephen), 1984, « The Auditory Image : A Metaphor for Musical and Psychological Research on Auditory Organization », dans CROZIER (W. R.) et CHAPMAN (J. C.), dir., *Cognitive Process in the Perception of Art*. Amsterdam, North-Holland, p. 289-323.
- , 2015. *Perception et cognition de la musique*. Paris, Vrin.
- MOUTAIN (Rosemary), « Auditory Scene Analysis and Electroacoustics », Actes du colloque international *Electroacoustic Music Studies Network – EMS09*. (<<http://www.ems-network.org/ems09/proceedings.html>>, accédé le 27/03/2023.)
- PRESSNITZER (Daniel), SUIED (Clara) et SHAMMA (Shihab A.), 2011, « Auditory Scene Analysis : the Sweet Music Ambiguity », *Frontiers in Human Neuroscience*, 5, n° 158, p. 1-10.
- REPP (Bruno), 2002, « The Embodiment of Musical Structure : Effects of Musical Context on Sensorimotor Synchronization with Complex Timing Patterns », dans PRINZ (W.) et HOMMEL (B.), dir., *Common Mechanisms in Perception and Action : Attention and Performance XIX*. Oxford/New York, Oxford University Press, p. 245-265.
- Seashore (Carl), 1938, *The Psychology of Music*. New York (NY), McGraw-Hill.
- VIEILLARD (Sandrine), ROY (Matthieu) et PERETZ (Isabelle), 2011, « Expressiveness in Musical Emotions », *Psychological Research*, 76, n° 5, p. 641-653.