

# **Le bruissement des nuées**

*Articulation morphologique de textures  
sonores par nuages de grains  
auto-organisés*

Mémoire de fin d'études

Ryo Baldet  
Ecole Nationale Supérieure Louis-Lumière  
Promotion Son 2021

**Directeur de mémoire interne** : Laurent Millot  
**Directeur de mémoire externe** : Thierry Coduys  
**Rapporteur** : Eric Urbain

version du 06/11/2021 à 23:40

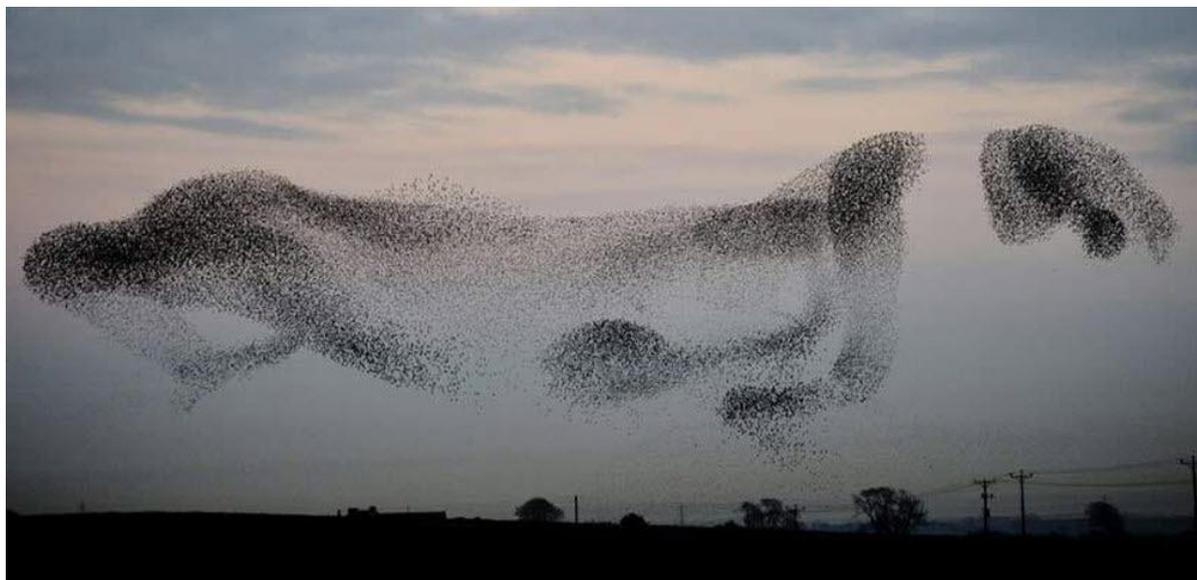
# Table des matières :

<b>Introduction</b>	<b>5</b>
<b>Partie 1. Les masses sonores comme nouveau paradigme de composition : entre écriture et perception</b>	<b>11</b>
La musique massique dans le paysage de la musique contemporaine des années 50-60	12
Des métronomes en masse : introduction à la musique massique par le Poème Symphonique pour 100 métronomes de György Ligeti (1963)	12
Les partitions graphiques comme moyens de formalisation de l'écriture massique	15
Etude de cas de Volumina pour orgue de Ligeti	15
Etude de cas de Metastasis pour orchestre de Xenakis	17
Plasticité de la masse : ductilité et transformations continues	19
La masse comme matière ductile	19
Quelques précisions sur la notion de sonorité en musique	19
Spectromorphologie et transformations continues	20
L'écriture des masses comme changement de paradigme perceptif	21
Emergence et Gestaltisme : quelques notions préalables	21
Perception des masses : analyse de scènes auditives	23
Complexité paradoxale : entre sensation d'homogénéité et de multiplicité	23
Regroupement par intégration simultanée : analyse des glissandi de cordes dans les dernières mesures de Metastasis	24
Sur la notion d'entitativité	25
Sur la notion de distances perceptives : modulations par changement du point d'écoute	26
Analogie des attributs de la Gestalt visuelle à l'analyse des scènes auditives	28
La densité comme attribut favorisant la sensation de fusion par intégration séquentielle : étude de cas à travers le Poème Symphonique pour 100 métronomes	29
Le rôle de la composante transitoire dans la détection d'unités au sein d'un flux	31
A propos de l'association d'un motif rythmique à une unité	32
Synthèse et définition de masses sonores :	33
<b>Partie 2. Les phénomènes naturels comme modèle de composition : articulation de textures sonores par modélisations physiques</b>	<b>35</b>
Composer avec le son.... composer le son, du grain à la texture, paradigme granulaire	37
La pensée granulaire	37
Micro-sons et grain	38
La pensée granulaire dans la musique	38
Caractérisation du micro-temps	41
Du grain à la texture	44
Perception de la texture : le coefficient de remplissage	45
Le timbre de la texture	46

Théorie de l'émergence sonologique	47
Synthèse de textures sonores	47
Manipulation de masses sonores par modélisation de systèmes inspirés de phénomènes naturels : de la stochastique aux systèmes auto-organisés	49
Stochastique, entre ordre et chaos	50
Limites de la stochastique dans la formation de sonorités émergentes	54
Systèmes auto-organisés comme orchestrateurs des nuages de grains : le cas de l'algorithme boids	55
Boid et vie artificielle	55
Description de l'algorithme	56
Boids comme instrument musical	59
L'approche directe :	60
L'approche relative :	61
L'approche par modèles physiques :	62
<b>PARTIE 3. Le bruissement des nuées : implémentation du synthétiseur de nuées et monstration publique</b>	<b>65</b>
Implémentation de l'algorithme de l'orchestrateur de nuées	66
Boids comme instrument musical expressif	66
Contrôle du système	66
Données extraites du système	67
Synthèse concaténative par corpus	68
Choix du corpus de sons non-voisés	71
Organiser l'espace timbral du corpus sonore par descripteurs audio	72
Système de sonification	75
Vitesse	77
Angulation relative	77
Accélération	77
Regroupement spatial	78
Bilan	78
Structuration globale de l'algorithme	78
Interface boids :	78
Moteur de synthèse sonore	79
Module d'analyse mubu.process	79
Module de corrélations	81
Module de sélection : mubu.knn	81
Module de lecture : mubu.concat~	81
Le bruissement des nuées : monstration publique	83
Choix du dispositif de diffusion sonore	84
Adaptabilité du dispositif de diffusion sonore	92
Les concerts improvisés	96
Retour d'expérience	97
La déambulation comme modulation de la perception de l'émergence des masses	97

Perception de mouvements de masse	97
Sonorités évoquées	98
Perception d'une entité "vivante"	99
Biais perceptifs induits par l'interface graphique boids	100
Le synthétiseur de nuées comme instrument	101
<b>Conclusions et perspectives</b>	<b>102</b>
<b>Annexes</b>	<b>104</b>
Descripteurs audio utilisés :	104
Réponses au questionnaire à l'issue du concert	105
Que vous ont évoqué les sons que vous avez entendus ?	105
Avez-vous ressenti des mouvements de masse, de groupe de sons ?	106
Avez-vous ressenti des présences d'une ou plusieurs entités "vivantes" ? Si oui, décrivez.	107
Le rapprochement et l'éloignement des hauts-parleurs ont-ils changé la perception de ce que vous avez écouté ? Si oui, comment ?	108
Avez-vous observé l'interface graphique utilisée pour la manipulation des sons ? Si oui, votre ressenti sur les sonorités générées a-t-il changé après l'avoir observé ?	109
<b>Bibliographie</b>	<b>111</b>

# Introduction



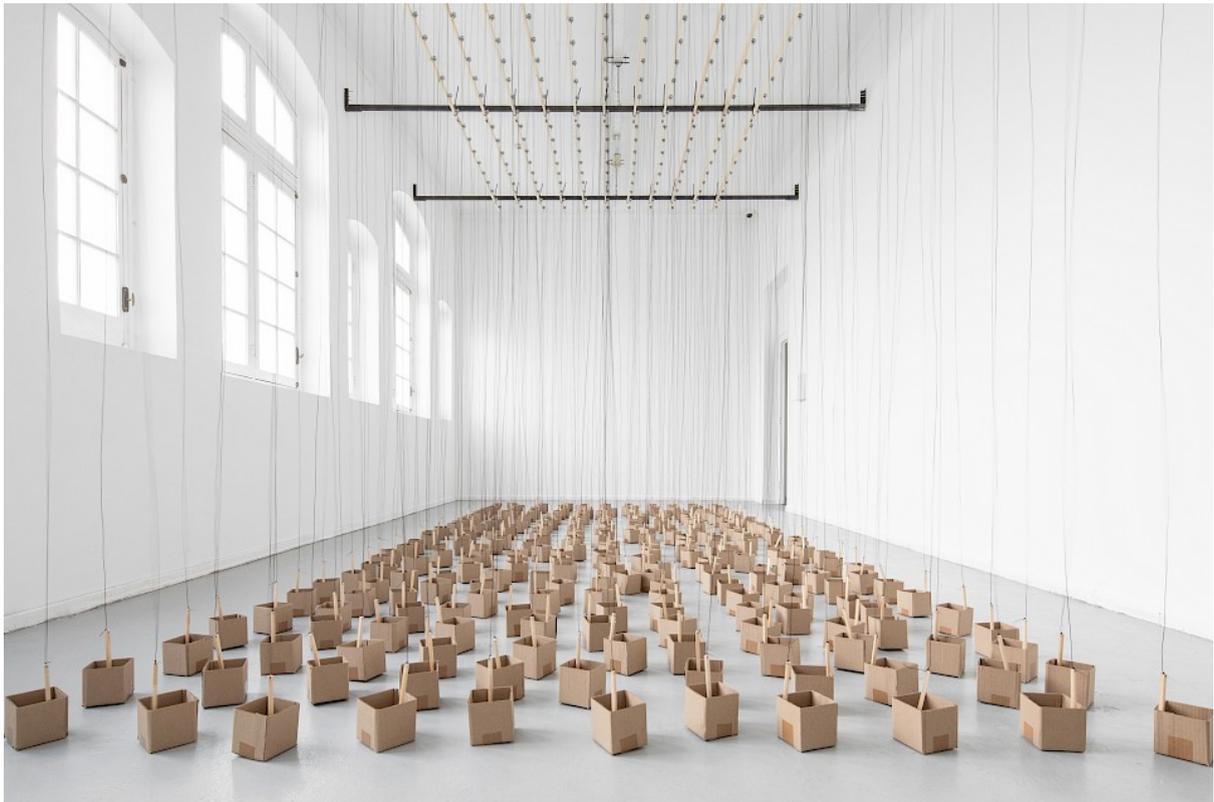
*Photographies de milliers d'étourneaux aux alentours de Gretna Green, Ecosse, 2014.  
Owen Humphreys, Press Association*

Des points noirs sont disséminés dans le ciel, semblables à un nuage grumeleux tiré d'un tableau pointilliste. Se regroupant par milliers, les étourneaux dessinent ensemble une masse qui semble flotter dans les airs, comme une matière liquide en suspens, s'agitant nerveusement et déformant sans cesse ses propres contours. La déviation de la direction de leur vol altère leur forme individuelle, passant de taches allongées à de fines lamelles, ou vice-versa. Cette variation de densité des couleurs noires sur fond clair semble donner à cet amas informe un volume tangible. Le mouvement des uns et des autres se déverse de proche en proche telle une onde de choc, faisant miroiter par vagues successives la surface de cette nappe perlée.

Vu de loin, on peine à discerner les étourneaux de manière individuelle. Ensemble, ils dessinent un superorganisme à la surface granulée, aux densités plus ou moins marquées en fonction des zones sur lesquelles se fixent les regards, laissant les points noirs disparaître pour laisser place à une grande image holistique.

Ces nuées, aussi appelées murmurations, émergent de la communication entre les oiseaux au sein du groupe grâce à leur instinct grégaire. Ces animaux sociaux forment des structures dites auto-organisées, dont l'effet global résulte de la multitude d'interactions simples et locales entre eux. Car si les étourneaux semblent se donner en spectacle aux yeux de l'observateur hypnotisé, c'est pour se protéger des prédateurs qu'ils se rassemblent par milliers. En formant ces réseaux de solidarité, ils mutualisent leur force pour former une défense collective, réduisant le danger pour eux-mêmes en essayant de rejoindre le centre du groupe.

Il se tisse sous mes yeux, un parallèle séduisant entre ces minuscules points noirs volants et le concept de "grains sonores", de "sonorités granulées". Mon rattention se déporte à présent vers les cohortes d'automates mis en espace dans les installations de l'artiste Suisse Zimoun.



297 prepared dc-motors, 1247m rope, wooden sticks 19 cm, cardboard boxes 10x10x10cm, 2019.  
Museum of Contemporary Art MAC Santiago de Chile, 2019.

Disposées au sol au centre d'un grand espace d'exposition blanc, de petites boîtes en carton semblent avoir proliféré par centaines à la manière d'une culture de champignons. Si l'installation impressionne visuellement au premier abord par la quantité des éléments dont elle est constituée, ce n'est pas seulement pour leur aspect scénographique que l'artiste les a convoqués.

A chaque boîte en carton cubique est associé un bâtonnet en bois, suspendu au plafond par une longue ficelle noire. A l'extrémité haute de celles-ci se trouvent des moteurs, qui, par leur agitation mécanique les remuent de façon frénétique. A la manière d'un instrument percussif, les baguettes viennent alors frapper les recoins des boîtes, émettant un rudimentaire bruit de grelottement.

Cet assemblage me rappelle les *fūrin*, carillons éoliens japonais que l'on retrouve dans les jardins au niveau de la devanture des portes lorsque vient l'été, et qui témoignent du passage du vent en émettant un léger tintement. Mais si les carillons révèlent une brise légère rafraîchissante en période caniculaire, le son émis par les centaines de boîtes en carton n'est au contraire pas sans rappeler le son d'une épaisse averse.

La scénographie de l'installation m'invite à imaginer ce rapprochement : le réseau de moteurs fixés au plafond est semblable à un nuage par lequel les gouttes se laissent tomber selon les lignes verticales dessinées par les ficelles noires, jusqu'à venir s'écraser sur le sol. Là, ils émettent un faible et unique son percussif, disparaissant aussitôt qu'il est apparu.

A partir de ces instruments au fonctionnement élémentaire, les mouvements désordonnés des bâtonnets confèrent à cet orchestre percussif une apparente complexité qui semble imiter, par leur comportement chaotique, les processus de génération de ces systèmes naturels. Regroupés ensemble au sein du même espace, les minuscules grelottements de ces centaines d'automates s'entremêlent et s'élèvent en un gigantesque brouhaha. Perdu dans la masse, le son individuel et éphémère des gouttelettes disparaît pour laisser place à un autre son, d'une autre nature, qui renvoie à la globalité de cet ensemble.

Lorsque l'on tend une oreille vers la fenêtre pour écouter la pluie, entendons-nous une somme de sons de gouttes frappant le sol, selon une distribution aléatoire dans le temps et dans l'espace, ou entendons-nous *la pluie* à proprement parler ? De la même manière, lorsque le vent fait chanter les arbres dans la forêt, entendons-nous une somme de dizaine de milliers de feuilles se frottant les unes aux autres, ou entendons-nous *le bruissement des feuilles* ?

J'ai l'intuition que le phénomène visuel et l'installation sonore se conjuguent dans la manière dont ils surgissent à la surface de notre attention. Si la nuée d'oiseaux apparaît comme un nuage granuleux, les sonorités pluviales émises par les automates de Zimoun semblent être leur pendant sonore. Dans les deux cas, nous sommes face à des phénomènes émergents, dont la forme globale résulte du regroupement des éléments dont ils sont constitués.

Dans chacun de ces cas, nous sommes confrontés à des *masses*. Celles-ci apparaissent à partir de l'instant où les unités élémentaires qui les composent, oiseaux comme micro-sons percussifs, ne sont plus perçues en tant qu'individus, mais contribuent par leur regroupement au phénomène émergeant dans son ensemble.

En musique, certains compositeurs tels que Xenakis ou Ligeti, ont fréquemment utilisé des procédés de composition mobilisant de grands ensembles de sons, des masses sonores, leur permettant de créer des textures imitant les effets des phénomènes naturels tels que la grêle, les tempêtes, les phénomènes sismiques ou cataclysmes mais aussi des grouillements d'insectes ou des exclamations de foules.

En jouant sur des effets perceptifs, ils parviennent à faire fusionner certaines entités distinctes en une masse tandis que d'autres conservent leur identité, ou alors à faire en sorte qu'une unité se détache du reste du groupe telle une voix émergeant d'un brouhaha. Multipliant l'invention de nouveaux procédés d'écriture, ces compositeurs ont mis en évidence le fait qu'il était possible de produire des sonorités extrêmement riches et variées grâce aux masses, en fonction de leur organisation et de leur comportement, inatteignables par des modes de composition traditionnels.

Car cette bascule vers le paradigme de l'écriture des masses posait de nouveaux enjeux compositionnels : il n'était pas question d'écrire une à une chaque unité sonore mais de trouver des moyens de les articuler ensemble, à la manière d'un chef d'orchestre dirigeant son ensemble instrumental. Comment composer à grande échelle, par la formulation d'intentions générales, avec des milliers d'entités sonores ?

C'est par la modélisation de systèmes naturels, et des lois statistiques qui les décrivent, que Xenakis trouve certaines solutions, qu'il expose dans son ouvrage Musiques Formelles en 1963. En transposant dans le domaine musical des opérations mathématiques de gestion d'ensemble d'entités, il énonce les prémices de procédés de compositions des masses par voie algorithmique.

Avec l'arrivée des outils informatiques et la musique électroacoustique, certains compositeurs tels que Curtis Roads ou Agostino Di Scipio ont continué à alimenter ces réflexions théoriques, notamment par le biais de la composition granulaire.

Cependant, la question des moyens d'articulation morphologique de masses sonores, notamment dans le cadre de manipulation temps-réel, reste encore aujourd'hui en suspens.

Alors que je travaillais sur un tout autre sujet de mémoire, le visionnage d'une vidéo sur le vol collectif d'étourneaux a fait germer une intuition : pourrait-on s'inspirer du phénomène des nuées afin de manipuler des nuages de grains sonores ?

A l'image de l'installation de Zimoun, par laquelle émane un nuage de micro-sons, j'entrevois un lien qui unit ces masses volantes aux masses sonores, une réciprocité par laquelle ils pourraient se répondre et se compléter : il est indéniable qu'il existe une résonance dans la manière dont nous percevons ces effets de masse dans le domaine visuel comme dans le domaine sonore.

En identifiant chaque oiseau à un son élémentaire, serait-il possible de faire émerger des masses de sons au comportement grégaire ? De la même manière que l'a fait Xenakis, serait-il possible de transposer une modélisation de nuée afin d'orchestrer des sons ? Cette translation d'effets synesthésiques peut-elle se concrétiser en une implémentation exploitable dans le cadre d'un jeu instrumental ? Par exemple par la création d'un instrument numérique synthétisant en temps-réel des nuées sonores ?

*Quelles types de sonorités émergeront alors ?*

Dans une première partie, je propose l'étude d'un corpus musical d'œuvres de Ligeti et de Xenakis, en m'appuyant sur des théories de l'analyse de scènes auditives. Cette étude me sert à essayer de comprendre et présenter les mécanismes de perception du phénomène d'émergence des masses sonores et pour voir de quelle manière les unités sonores élémentaires se fondent dans la masse ou, au contraire, émergent de celle-ci.

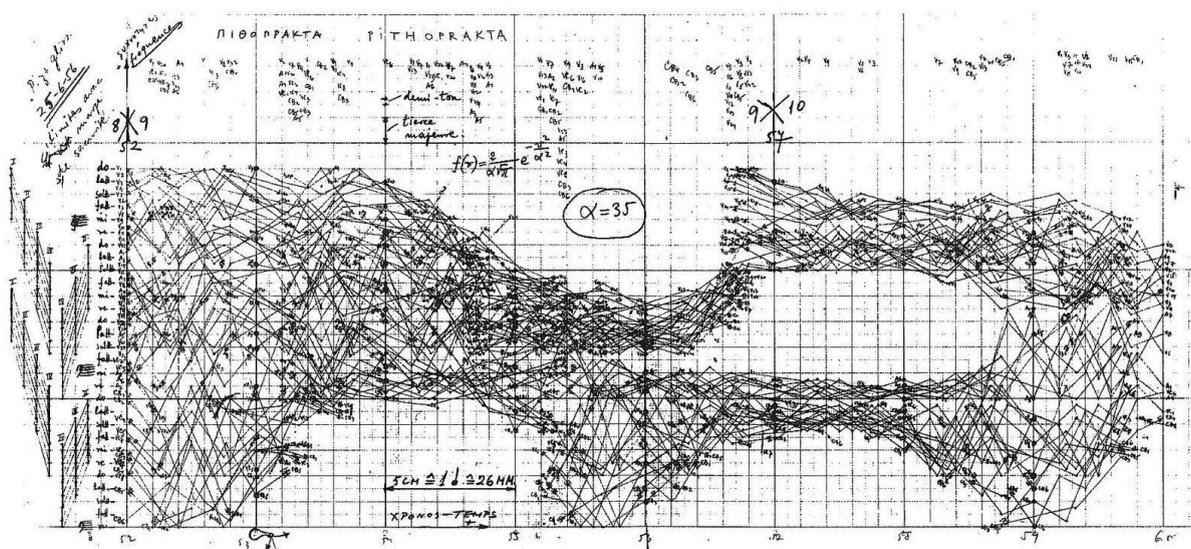
Dans la deuxième partie, je procède à l'étude des possibilités de composition de textures sonores par masses de micro-sons. A partir de l'analyse de phénomènes naturels et des travaux théoriques de Xenakis sur la formulation d'une pensée granulaire, je chercherai à identifier ce qui caractérise un micro-son et à comprendre ce qui se cache sous le terme polysémique de texture sonore. Grâce à l'étude des premières hypothèses de Xenakis sur la stochastique et avant de m'intéresser à la modélisation des nuées avec l'algorithme boids, je me pencherais alors sur les moyens de manipulation des masses par voie algorithmique.

Enfin, dans la dernière partie, j'exposerai les détails de ma proposition d'implémentation du synthétiseur de nuées. Je procéderai d'abord à la description de l'algorithme, de l'interface visuelle des nuées au moteur de synthèse granulaire, élaboré entièrement au sein de l'environnement Max/MSP. Puis, je terminerai mon développement par la description et le retour d'expérience de la monstration publique de la Partie Pratique de Mémoire, dans le cadre d'une série de concerts, en exposant mes choix de scénographie et de dispositif de projection sonore.

# Partie 1. Les masses sonores comme nouveau paradigme de composition : entre écriture et perception



*Photographie issue de la série "Ornithographies" de Xavi Bou*



*Extrait de partition graphique de Pithoprakta - Xenakis (1957)*

## I. La musique massique dans le paysage de la musique contemporaine des années 50-60

### A. Des métronomes en masse : introduction à la musique massique par le Poème Symphonique pour 100 métronomes de György Ligeti (1963)

100 métronomes mécaniques sont disposés, à l'arrêt, sur des tables de la petite scène de l'hôtel de ville de Hilversum aux Pays-Bas, devant un public silencieux et impatient. Les dix exécutants attendent patiemment le signal du chef, le compositeur György Ligeti lui-même, pour pouvoir s'affairer à déclencher chaque métronome le plus rapidement possible. Nous sommes le 13 septembre 1963 et le compositeur autrichien présente pour la première fois sa pièce conceptuelle Poème Symphonique pour 100 métronomes.

Une fois ceux-ci déclenchés, les exécutants sortent de scène, laissant derrière eux une cohorte de sonorités tictacquantes. Tous accordés sur des *tempi* différents, les métronomes battent la mesure à leur rythme propre. La pulsation désordonnée de cette multitude résultent en un nuage percussif chaotique. Cette sonorité émergente, que l'on pourrait caractériser comme une *texture*, semble être contenue dans des contours statistiquement stables, mais à l'intérieur de laquelle le grouillement des sons se renouvelle à chaque instant. Peu à peu, le brouhaha continu s'effrite pour laisser place à une jungle rythmique complexe, certains métronomes réglés sur des tempi rapides cessant leur battue par dissipation d'énergie après quelques minutes. Ainsi, les sons des métronomes les plus lents, de plus en plus isolés, deviennent identifiables par leur balancement régulier. La pièce s'achève lorsque tous les battements de métronome ont cessé, environ 19 minutes après leur déclenchement.



Première de Poème symphonique pour 100 métronomes à l'hôtel de ville de Hilversum (1963), 10 exécutants attendent le signal du compositeur pour déclencher les 100 métronomes.<sup>1</sup>

<sup>1</sup> La première du concert de 1963 est consultable via ce lien :

<https://gaudeamus.nl/en/jubileum/ligetis-poeme-symphonique-the-premiere-in-1963/>

La partition comme mode d'emploi de la mise en route des métronomes, ou l'absence de l'intervention humaine (mis à part pour le déclenchement), sont autant de procédés conceptuels reflétant les influences typiquement Cagiennes du début des années 60, ou du mouvement Fluxus. Il s'agira de la dernière œuvre de Ligeti en relation avec ce mouvement dont il a été membre durant quelques années. Car, si le compositeur a délaissé la présence d'interprètes humains pour laisser place aux instruments mécaniques autonomes, c'est parce qu'ils se suffisent à eux-mêmes pour produire les jeux d'effets perceptifs escomptés. En effet, cette période propice aux expérimentations compositionnelles a amené Ligeti à faire usage de procédés brouillant les repères des formes et sonorités connues, par l'influence notamment des théories de l'information :

*J'avais en tête de nombreuses grilles superposées, des motifs moirés, qui donneraient lieu à des structures rythmiques changeantes. De plus, au début des années 60, la théorie de l'information était à la mode, et il y avait beaucoup de radotage sur "l'esthétique de l'information". Je me suis donc dit que j'allais produire une grille rythmique si épaisse au début qu'elle semblait pratiquement continue : cela signifierait flou et désordre. Pour ce faire, j'avais besoin d'un nombre suffisamment important de métronomes - 100 était une estimation. [...] Dès que certains des métronomes sont arrêtés, des motifs rythmiques changeants apparaissent, en fonction de la densité du tic-tac, jusqu'à ce que, à la fin, il ne reste plus qu'un seul métronome émettant un tic-tac lent, dont le rythme est alors régulier.<sup>2</sup> (Ligeti 1997)*

Ligeti était tout au long de sa vie obnubilé par l'idée d'atteindre des sonorités statiques, suspendues dans le temps, caractérisées le plus souvent par des notes tenues et l'absence de silences au sein d'un assemblage de tissus polyphoniques. L'utilisation d'un ensemble d'instruments percussifs, dont les sons sont de nature discontinue dans le temps, dans le but de construire des sonorités continues, se posait donc comme un véritable défi.

Comment transformer un ensemble de métronomes, instrument métrique par excellence, dictant une référence temporelle stable et dynamique sur laquelle le musicien vient se synchroniser, en un nuage trouble et statique qui anéanti tout repère temporel ?

Comme il le mentionne lui-même, le facteur primordial qui permet à cet effet d'apparaître réside dans la convocation d'un nombre suffisant de métronomes jouant en même-temps. En effet, la grande densité des tic tac joués au début de la pièce leur assure une répartition aléatoire relativement homogène dans le temps, permettant de percevoir une sonorité globale quasiment continue, qui comporte une certaine stabilité chaotique. Lorsqu'ils sont tous activés, il semble presque impossible d'isoler un métronome du reste de l'ensemble, à moins de le voir, quand bien même leur nature périodique nous permettrait de les identifier. Ils s'effacent collectivement à ce stade derrière la masse.

---

<sup>2</sup> Citation originale : "I had in mind numerous superimposed grids, moiré patterns, which would result in changing rhythmic structures. Also, at the beginning of the 60s, information theory was fashionable, and there was a lot of drivel about "information aesthetics." So I thought, I'll produce a rhythmic grid that is so thick at the beginning that it seems practically continuous: that would signify blurring and disorder. To do this I needed a sufficiently large number of metronomes - 100 was just an estimate. [...] As soon as some of the metronomes have run down, changing rhythmic patterns emerge, depending on the density of the ticking; until, at the end, there is only one, slowly ticking metronome left, whose rhythm is then regular".

Cette œuvre me rappelle de manière analogue l'installation de Zimoun évoquée dans l'introduction. On y retrouve les mêmes concepts : un grand nombre d'automates mécaniques semblables entre eux et aux fonctionnements rudimentaires, disposés dans un même espace, émettant des sons ponctuels de manière incontrôlée qui résultent en une distribution aléatoire des événements sonores dans le temps.

Une des grandes différences entre ces deux œuvres réside dans leur temporalité : si dans l'installation de Zimoun, les automates sont entretenus continuellement, les métronomes du poème symphonique cessent de battre lorsque l'énergie du mécanisme s'est entièrement dissipée.

Dans les deux œuvres, ni les métronomes ni les automates ne sont utilisés comme s'ils étaient simplement dupliqués, scandant une certaine battue en étant synchronisés ensemble. Ils sont au contraire exploités d'abord parce que la manière dont ils sont conçus permet des décalages temporels entre eux, garantissant ce foisonnement chaotique. Ensemble, ils forment donc plus qu'une simple somme d'instruments : ils transforment par leur densité le phénomène percussif discontinu en une texture entretenue, massive, émergeant de la cohue.

Ces changements d'états sont caractéristiques du phénomène d'émergence, illustrant l'apparition d'un son d'une nature nouvelle, perçu comme un tout et ne préexistant pas à l'assemblage des parties qui la compose : l'existence du nuage de sons en tant qu'entité homogène ne préexiste pas à la mise en action des 100 métronomes.

Comme Ligeti, le compositeur Iannis Xenakis a également expérimenté en profondeur ces effets à partir de la composition de masses sonores. Il théorise dans son traité *Musiques Formelles* (1963) le concept de sonorités du second ordre pour qualifier ces illusions perceptives.

Les deux compositeurs continueront longuement à utiliser ces écritures par grands ensembles sonores, qu'ils soient orchestraux ou générés par dispositifs électroacoustiques. Ces musiques dites "massiques" marquent un tournant dans l'histoire de la musique contemporaine, en s'opposant notamment à l'esthétique sérialiste très en vogue à cette période. Entre des aplats de notes par clusters répartis dans tout l'effectif orchestral, et des nuages de *pizzicati* de cordes, Xenakis et Ligeti seront les principaux initiateurs de ce mode de composition, proposant des évolutions sonores globales par l'écriture de grands mouvements d'ensemble.

## B. Les partitions graphiques comme moyens de formalisation de l'écriture musicale

### 1) Etude de cas de Volumina pour orgue de Ligeti<sup>3</sup>

La cas de Volumina pour orgue, composé en 1961 par Ligeti, est particulièrement intéressant parce qu'il déploie sur la partition graphique les intentions du compositeur à vouloir écrire des sonorités globales de manière formelle. Les effets de « surfaces » sont clairement perceptibles dans l'utilisation de clusters successifs, et s'expriment sur la partition comme des figures pleines, ou dans lesquelles sont entremêlées des amas de lignes. Ces aplats d'encre sont la traduction presque littérale « d'aplats » de notes, qui se manifestent à l'écoute par la perception de blocs très denses et compacts. Il écrit en 1968 dans le texte de présentation pour la nouvelle version de son oeuvre pour orgue :

*"J'imaginai une musique quasiment amorphe, dans laquelle les sons n'avaient plus de fonction individuellement mais où les accumulations et les agglomérations de sons ainsi que les rapports de volume de ces collectifs sonores seraient constitutifs de la forme. Densifications, dissolutions, divers mouvements internes dans les agglomérats sonores, événements tectoniques comme des avalanches, des entassements, des sédimentations, mais aussi des processus atmosphériques comme l'évaporation, l'essoufflement et autres phénomènes semblables articulent cette forme qui, vue dans sa globalité, est de nature continue."<sup>4</sup>*

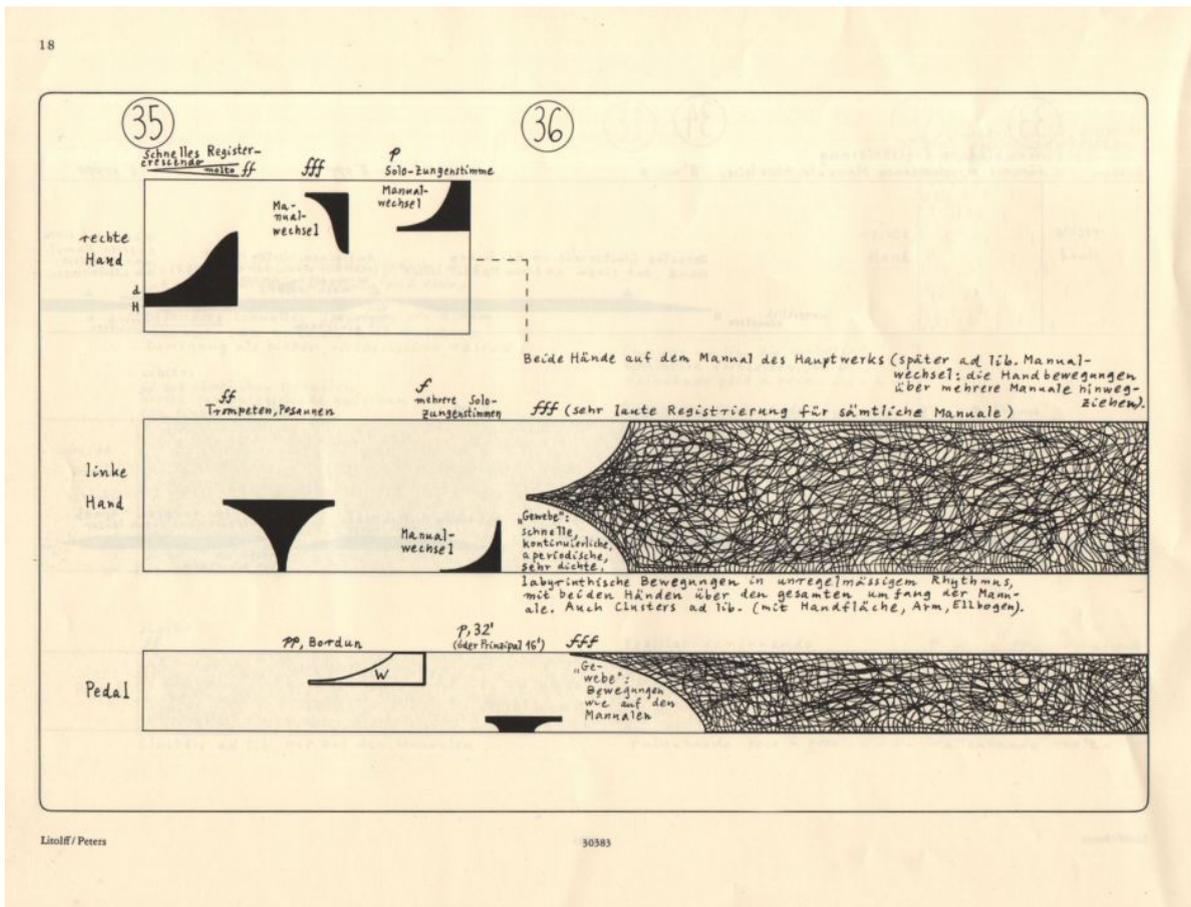
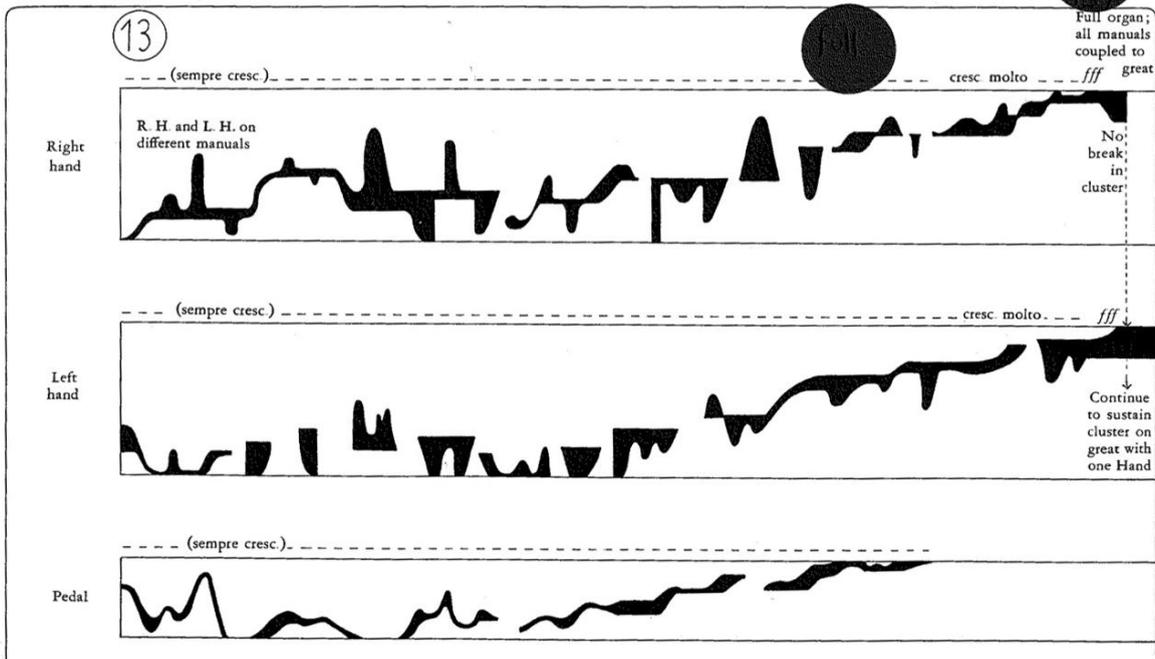
Là encore, il semble très difficile de démêler les notes individuelles prises dans les grands clusters. Cette saturation de l'espace harmonique est volontairement produite pour effacer la valeur tonale de chaque note jouée : seule la perception des formes globales délimitées par leurs contours, ainsi que les grouillements chaotiques à l'intérieur de celles-ci importent. On retrouvera des procédés d'écriture similaires dans Atmosphère pour orchestre (1961) ou Requiem (1963-65), mais aussi dans de nombreuses pièces de Xenakis telles que Metastasis (1954) ou Pithoprakta (1957).

---

<sup>3</sup> La pièce suivie par la partition graphique est consultable sur youtube via ce lien :

<https://www.youtube.com/watch?v=G7bdwarV6SQ>

<sup>4</sup> György Ligeti, « Über Volumina », écrit en février 1968 comme texte de présentation pour la création de la nouvelle version, le 8 mars 1968 à Kiel. Traduit par Catherine Fourcassier in György Ligeti, *L'Atelier du compositeur*, Éditions Contrechamps, 2013.



Extraits de partitions graphiques de Volumina pour orgue (1968)

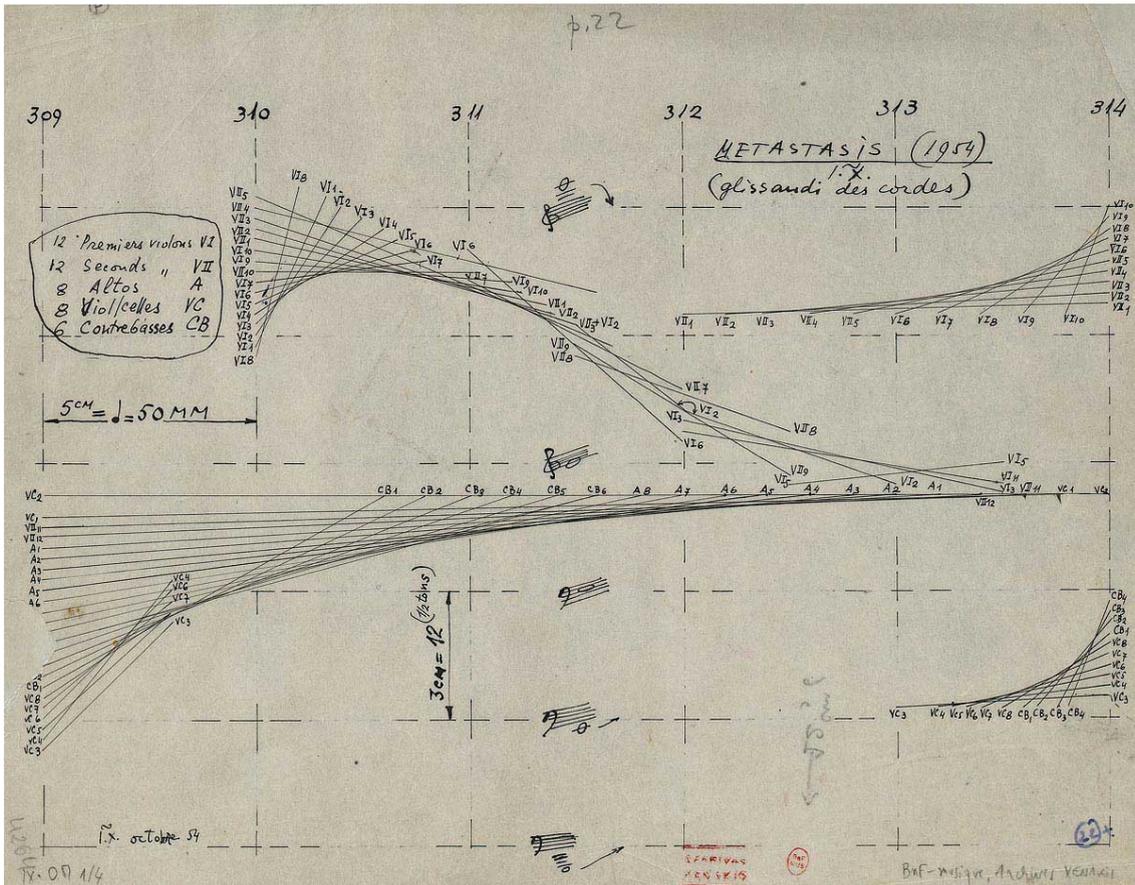
## 2) Etude de cas de Metastasis pour orchestre de Xenakis

Pour sa pièce Metastasis, composée en 1954, Xenakis écrit une voix indépendante presque pour chaque instrument de l'ensemble orchestral. Dans les premières et dernières mesures de la pièce, il attribue, pour chaque voix de cordes un segment de droite, assimilé à un glissando linéaire.

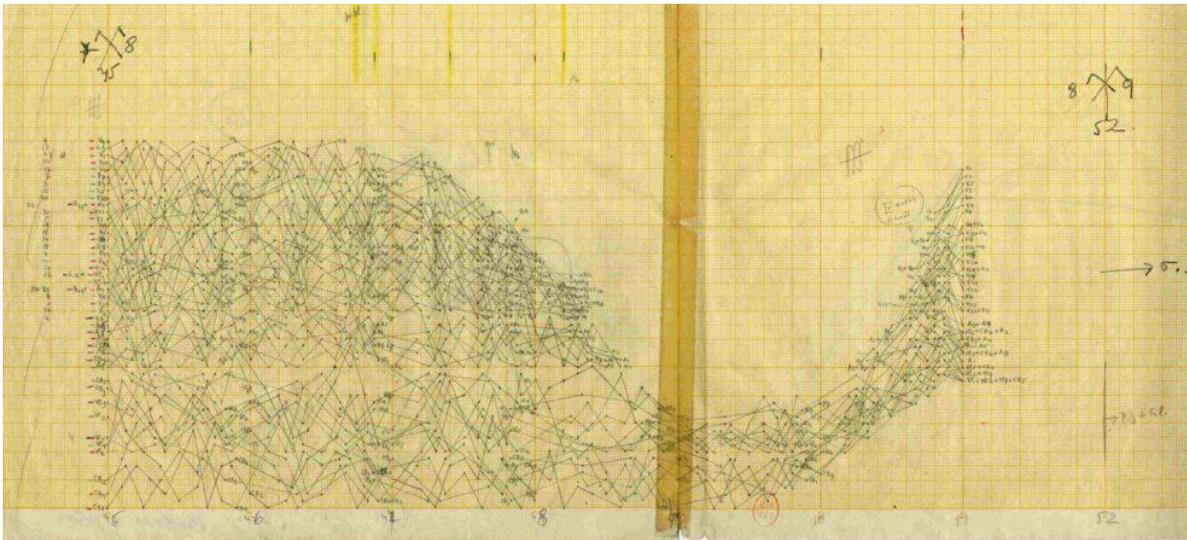
Elles sont inscrites au sein du modèle des partitions traditionnelles avec en abscisse le temps et en ordonnée la hauteur des sons, en s'abstrayant de la grille du système de notation musicale tempérée. Ces entrelacements de sons en de gigantesques clusters permettent des glissements de masse vers des points de convergence, ou au contraire permettent la dispersion vers des voies divergentes.

De quelques coups de traits dessinés à la règle sur papier millimétré, Xenakis parvient ainsi à étirer ou à compresser la matière sonore, sculptant l'amas de sons en une forme définie. La diversité des hauteurs jouées simultanément engendre une sonorité très complexe, qui produit une sensation de multiplicité, mais dont il est difficile voire impossible d'isoler une voix, prise dans la cohorte.

Ainsi, l'effet massal du mouvement collectif l'emporte sur la perception des trajectoires individuelles.



Extrait de partitions graphiques de Metastasis (1953-54)



Extraits de partitions graphiques de Pithoprakta (1957),  
au sein desquelles Xenakis utilise les mêmes principes de superpositions de glissandi.

## C. Plasticité de la masse : ductilité et transformations continues

### 1. La masse comme matière ductile

Dans ces œuvres, les différentes voix sont « aplanies » par superposition sur un plan, permettant aux compositeurs de dessiner la forme dans sa globalité. Les partitions graphiques permettent à Xenakis ou Ligeti de visualiser et de déployer ces globalités émergentes, résultantes à la fois métaphorique et formelle de l'imbrication de toutes les voix en amas de notes.

Celles-ci se fondent dans une mélasse complexe, au sein de laquelle elles forment un ensemble indissociable. Cette dissolution réduit l'écriture des masses à un volume déformable et modelable comme un tout, et qui peut se dilater et se compresser à la volonté du compositeur. Ainsi, le travail du compositeur se rapproche de celui du plasticien, façonnant la matière sonore comme une matière ductile.

### 2. Quelques précisions sur la notion de sonorité en musique

Dès 1936, et alors que le paysage musical d'avant-guerre était dominé par le néo-classicisme, le compositeur Edgar Varèse pose la question de la composition des mouvements de masses au sein de l'écriture musicale. Pionnier dans la composition de sons plutôt que l'écriture de notes, utilisant des sonorités bruitées ou non-conventionnelles d'instruments classiques, il propose une nouvelle définition de la musique dans son manifeste The Liberation of Sound (1966) :

*“Lorsque de nouveaux instruments me permettront d'écrire la musique telle que je la conçois, en prenant la place du contrepoint linéaire, le mouvement des masses sonores, ou plans mouvants, sera clairement perçu. Lorsque ces masses sonores entreront en collision, des phénomènes de pénétration ou de répulsion sembleront se produire. Certaines transmutations ayant lieu sur certains plans sembleront être projetées sur d'autres plans, se déplaçant à des vitesses et des angles différents. Il n'y aura plus l'ancienne conception de la mélodie ou du jeu des mélodies. L'œuvre entière sera une totalité mélodique. L'œuvre entière coulera comme coule une rivière.”<sup>5</sup>*

Cette définition préalable de la musique comme “son organisé” coïncide avec l'écriture des musiques massiques. L'exploration de notions telles que les “textures” ou les “masses” sonores nécessitent une abstraction en regard de la puissance symbolique de la note, qu'elle soit représentée dans un système tonal ou atonal<sup>6</sup>.

---

<sup>5</sup> citation originale : “When new instruments will allow me to write music as I conceive it, taking the place of the linear counterpoint, the movement of sound masses, or shifting planes, will be clearly perceived. When these sound masses collide the phenomena of penetration or repulsion will seem to occur. Certain transmutations taking place on certain planes will seem to be projected onto other planes, moving at different speeds and at different angles. There will no longer be the old conception of melody or interplay of melodies. The entire work will be a melodic totality. The entire work will flow as a river flows”.

<sup>6</sup> Le concept de “post-tone” music, qui n'a d'équivalent français, et utilisé par analogie avec le terme “post-tonal” (atonal en français), semble être le terme le plus approprié pour parler de musique comme “son organisé”. Il est défini dans l'article “Sound mass, auditory perception and 'post-tone' music” (Noble & McAdams 2020) comme “une musique au sein de laquelle le timbre joue un rôle primordial. A la place de notes, d'accords harmoniques ou

Dans toutes ces explorations musicales, il s'agit de faire abstraction des composantes musicales sur lesquelles reposent les fondations de la musique "savante" occidentale, c'est-à-dire de la puissance symbolique de la note au sein d'un système tonal ou atonal. Cette abstraction est nécessaire afin que la perception de l'auditeur puisse se polariser sur les évolutions timbrales continues des sonorités émergentes.

Il ne s'agit donc pas de manipuler des notes, mais d'organiser des sons, dans toute leur diversité comme le terme le laisse supposer, afin de construire d'autres sonorités. Je m'appuierais dans le cadre de ce mémoire sur la définition que propose le musicologue Makis Solomos :

*"Cette notion désigne une entité globale intégralement construite et composée de l'intérieur, issue de la dissolution des dimensions classiques du son (hauteur, rythme, timbre...), c'est-à-dire de la perte de leur autonomie. La recomposition et la fusion de ces dimensions génèrent la sonorité en tant que surplus, qualité émergente. Elle découle également de la fusion de la forme et du matériau. Le choix du terme sonorité indique que nous sommes proches de la notion de « son », mais qu'il s'agit d'un son composé, articulé, construit, c'est-à-dire d'un artefact, et non pas du son naturel – encore moins du « timbre »". (Solomos 2013, p.216)*

### 3. Spectromorphologie et transformations continues

La construction de ces sonorités continues révèle, dans le contexte musical des années 50, les prémices d'une pensée spectromorphologique. Il s'agit non plus de construire des compositions selon des "événements" dynamiques, mais plutôt de s'intéresser à la manière dont se sculpte le spectre sonore dans le temps. Ici aussi, la hauteur tonale et le rythme sont délaissés afin de se concentrer sur les évolutions des qualités timbrales.

Ces transformations, ou interpolations, sont définies par Smalley comme *"la perception du changement d'une identité sonore à une autre"* (Smalley 1993). Ces lentes transformations continues de matériaux sonores dans le temps constitueront les *modus operandi* de courants musicaux dans les années 70 comme la musique spectrale, dont Gérard Grisey (avec *Partiels*, 1975) ou Tristan Murail (avec *Désintégrations* 1982-83) sont les précurseurs, et au sein desquelles des techniques microtonales d'orchestration sont utilisés pour synthétiser des sons complexes.

Le concept de sonorités émergentes, ou sonorités de second-ordre, résonne de manière analogue avec ces principes de transformations d'identités de sons. Mais si dans la pensée spectromorphologique les sonorités se transforment dans le temps, les sonorités du second-ordre, elles, font surface par le changement de densité de sons mobilisés.

---

*de rythmes, la musique 'post-tone' s'attache à la spectromorphologie, la perception auditive, les objets sonores, et le son pour lui-même".*

## II. L'écriture des masses comme changement de paradigme perceptif

En utilisant des masses de notes comme des unités musicales, Xenakis et Ligeti font basculer le langage musical dans un nouveau paradigme. Comme dans tout langage musical, ce basculement nécessite des attributs propres à ce nouveau mode d'écriture afin de pouvoir qualifier et quantifier ces effets. Par conséquent, la musique dite "massique" doit être considérée comme un style musical à part entière, avec ses propres principes d'organisation et ses propres modes d'émergence.

Cependant, aucune théorie musicale n'existe à ce jour pour nous donner des clés de compréhension des intrinsèquement induits par cette écriture. Ils devront être analysés par le truchement d'études perceptives, notamment par l'analyse des scènes auditives, afin d'en comprendre les mécanismes d'apparition.

Comment percevons-nous ces "effets de masse" ? Comment les composer ? Finalement, quels sont les facteurs qui déterminent l'existence de ces phénomènes émergents ?

### A. Emergence et Gestaltisme : quelques notions préalables

En 1946, André Souris, compositeur et musicologue belge, évoque dans le chapitre «La forme sonore», de son essai Conditions de la musique (Souris 1946), le phénomène de transformation subite et inattendue de la *qualité* déclenchée par le changement survenu dans la *quantité* :

*« Il se construit en musique des phénomènes comparables à ceux du monde physique. Un accroissement ou un décroissement de quantité peut, à un certain moment, déclencher une transformation subite de la qualité. »*

Cette observation résonne avec les sonorités caractérisant les œuvres citées précédemment, au sein desquelles des qualités sonores continues et globales émergent de la mobilisation de masses.

Cette réflexion s'inscrit dans un contexte de remise en question de la perception des formes musicales avec notamment l'avènement de la Gestalttheorie (théorie de la forme) au début du XXe siècle. Cette école de la Gestalt défend une vision de la psychologie de la perception éminemment phénoménologique. Elle s'attache à l'étude de la manière dont se construisent des modèles mentaux à partir de l'entremêlement de connexions qui se forment entre les éléments perçus par nos interfaces sensorielles. Pour les gestaltistes, les processus de représentation mentale traitent les phénomènes qui nous apparaissent comme des formes globales plutôt que comme la juxtaposition d'éléments simples.

En psychologie cognitive, on retrouve souvent l'adage : "l'ensemble est plus grand que (ou différent de) la somme de ses parties". En d'autres termes, cette expression révèle que l'ensemble a des propriétés ou des fonctions qui diffèrent de celle des parties dont il est constitué, prises individuellement.

Les travaux des Gestaltistes sur la perception visuelle ont permis de mettre en évidence le fait que ces changements d'état sont avant tout une affaire de perception. Si des galets sont alignés sur une plage pour former une lettre, nous la verrons probablement telle quelle, puisque la manière dont elle se structure nous renvoie un message qui fait sens à nos yeux, comme une information signifiante de haut niveau. Mais si nous déplaçons seulement quelques galets de leur position initiale, de sorte que la lettre ne soit plus reconnaissable, alors nous ne percevons probablement qu'un ensemble de galets isolés.

De la même manière, il existe un pendant sonore de ces principes d'organisation, étudiées sous le modèle de l'analyse des scènes auditives (*auditory scene analysis*) entamées par Alan Bregman dans les années 90 (Bregman 1990).

Il s'agit de s'intéresser à la manière dont nous percevons des éléments isolés au sein d'une scène sonore, grâce aux principes de dissociation ou de regroupement. Par exemple, lorsque l'on est plongé dans un environnement bruyant dans lequel de nombreuses conversations ont lieu simultanément, nous avons la capacité d'isoler perceptivement celle sur laquelle notre attention se porte. Cette capacité de dissociation d'unités sonores au sein d'un même flux (*stream*) est bien connue des travailleur-euses du son sous le nom de *cocktail party effect*. Et, de la même manière que les galets, notre cerveau regroupe des éléments sonores isolés afin de les traiter comme un seul et même tout, comme c'est le cas dans les pièces de musiques massiques.

Notre environnement est construit selon des ensembles entremêlés dans des structures organisationnelles et hiérarchiques, et dont les propriétés sont significatives seulement parce qu'elles sont identifiées comme un unique ensemble cohérent.<sup>7</sup> Ce qui importe alors pour comprendre la manière dont ces structures émergentes apparaissent, ce n'est pas seulement la nature des éléments qui les constitue mais aussi et surtout la manière dont ils sont organisés et interconnectés entre eux.

Il s'agira alors d'utiliser l'analyse des scènes auditives dans le but de nous éclairer sur la manière dont nous percevons les œuvres mobilisant des sons en masse, comme celles de Ligeti.

Quels sont alors les facteurs qui déterminent l'apparition du phénomène d'émergence dans le champ de la composition ? Et dans quelle mesure pourrait-on le composer ?

Une majorité d'éléments de réponses peut être trouvée par le prisme de l'analyse des scènes auditives appliquées à ces œuvres de musique massique.

---

<sup>7</sup> Bregman, A. S. (1990). *Auditory Scene Analysis*. The MIT Press, p.138 : "But what does this have to do with perception? The answer is this: Because nature allows structures to behave in ways that are derived from the properties of parts but are unique to the larger structure, perception must calculate the properties of structures, taken as wholes, in order to provide us with valuable information about them."

## Perception des masses : analyse de scènes auditives

Malgré l'importance de l'adoption du « paradigme des masses » dans la recherche analytique et théorique, considéré comme procédé compositionnel à part entière, les études perceptives concernant ce style musical n'ont été menées que relativement récemment (Noble 2018, Noble & MacAdams 2020, Noble et al 2020, Douglas et al 2016).

Ces études proposent des analyses de larges corpus de musiques apparentées au style classique, par l'utilisation de tests perceptifs auprès d'un échantillon de participant·e·s à la fois familier·e·s et non familier·e·s de ces musiques, et parviennent à tirer certaines conclusions en corrélation avec des procédés d'analyse des scènes auditives.

Il s'agira maintenant d'examiner les *attributs*, les éléments sémantiques qui vont nous permettre de qualifier ce qui définit l'esthétique des masses sonores.

### Complexité paradoxale : entre sensation d'homogénéité et de multiplicité

Toutes les pièces du corpus de musiques classiques évoquées jusqu'ici, malgré la diversité des sonorités qu'elles proposent, se rejoignent perceptivement sur plusieurs points.

Les masses sont avant tout caractérisées par leur aspect « complexe<sup>8</sup> ». Ces flots de sons, du fait de leur grande densité, défient nos capacités de traitement d'informations auditives. Les intrications rythmiques, les successions rapides de sons, le nombre de sources sonores ou d'événements grandissants sont des facteurs augmentant ce ressenti. À ces surcharges de stimuli auditifs s'ajoutent la perte de tout repère tonal ou rythmique sur lesquels l'auditeur·rice pourrait se raccrocher. Les masses flottent dans un espace-temps sans métrique, brouillant par la même occasion la perception du temps passé à les écouter.

Cependant, malgré leur caractère complexe, nous retenons tout de même une impression globale d'homogénéité des différentes voix qui les composent, sans pour autant que celles-ci fusionnent en un unique percept auditif, c'est-à-dire qu'elles soient identifiées et réduites à une seule entité sonore. De plus, si l'on discerne les contours de ces amas, cela ne nous empêche pas de percevoir le fourmillement interne des sons, qui semblent témoigner, du fait de leur activité, d'une grande multiplicité. Au contraire, c'est bien la perception de ces entrelacs complexes qui permettent de définir et identifier ce qui pourrait s'apparenter à une masse sonore.

Comment expliquer ce paradoxe entre perception d'une homogénéité globale et sensation de multiplicité ?

Jason Noble nous explique qu' *"au-delà d'un certain point, nous ne sommes plus en mesure d'analyser tous les stimuli sonores entrants comme des unités, et l'augmentation de la complexité acoustique se traduit par une diminution de la complexité perçue. Les détails*

---

<sup>8</sup> D'après l'étude de Jason Noble dans "Semantic Dimensions of Sound Mass Music" 2020, le terme complexité serait associé à la sensation de la difficulté de percevoir individuellement les différents sons qui composent les masses, à leurs interactions et activités internes. p.220

*acoustiques s'annulent les uns les autres par superposition, et se fondent ainsi en une perception globale*" (Noble & McAdams 2020, pp.7-8).

Lasse Thoresen décrit ce phénomène comme une *"complexité paradoxale"*, qu'il définit comme étant un attribut d'un objet sonore contenant *"une myriade de détails, mais ayant un caractère global perceptivement simple"* (Thoresen 2015, p.457).

Face à cette saturation perceptive, nous aurions donc tendance à relier les composantes sonores distinctes entre elles pour les réduire en une entité homogène. Bien évidemment, plusieurs facteurs semblent amener à amplifier cette sensation de fusion, comme ils pourraient amplifier une sensation de dissociation (ou fission). Quels sont-ils ?

Il faudra pour les déterminer procéder à l'analyse de notre corpus d'œuvres massiques par le prisme de l'analyse des scènes auditives. On utilisera les termes d'**unité** pour caractériser une composante distinctive du groupe, ainsi que le terme **flux** pour caractériser au contraire ce qui s'apparente au reste de l'ensemble homogène.

Il s'agira alors de déterminer les facteurs qui provoquent une sensation de regroupement de plusieurs unités en un flux, ou au contraire une sensation de dissociation d'une unité vis-à-vis du flux.

#### Regroupement par intégration simultanée : analyse des glissandi de cordes dans les dernières mesures de Metastasis

Si nous reprenons l'exemple de la dernière partie de Metastasis, les glissandi des cordes sont entremêlés dans une superposition de mouvements ascendants et descendants, selon de larges ambitus, qui viennent se réduire progressivement dans les dernières mesures en un point de convergence : un unisson. La diversité des hauteurs jouées au début amène à une sensation de multiplicité et de complexité, qui se réduit quelques instants avant l'unisson, liant les différentes voix par leur proximité de hauteur.

Ces sensations de dissociations puis d'intégrations d'unités au sein d'un même flux peuvent s'expliquer par le fait que nous avons tendance à regrouper des unités ayant des composantes fréquentielles voisines : il s'agit du principe de regroupement par intégration spectrale (Bregman 1990, p.654). Un autre facteur favorisant la sensation d'intégration provient de la similarité des timbres mobilisés. Les sons des différentes unités sont facilement assimilables au même groupe car ils appartiennent à la même famille d'instruments : nous entendons des glissandi joués par un ensemble de cordes frottées.

Mais ces qualités intrinsèques aux unités sonores elles-mêmes n'expliquent pas tout. Plutôt que d'entendre des différences de hauteurs, nous percevons avant toute chose la manière dont elles évoluent dans le temps. En effet, dans cet extrait nous entendons principalement d'abord un mouvement global descendant superposé à un mouvement global ascendant, qui se transforment progressivement en un entremêlement plus incertain, qui enfin converge vers l'unisson. L'identification de ces différentes phases se distingue donc d'abord par la manière dont se « comportent » les différentes voix au sein de ces mouvements globaux, plutôt que sur la distinction de différences de hauteurs relatives.

Si nous écoutons un extrait dans lequel les notes étaient jouées de manière continue et statique, sans glissandi, alors le regroupement par intégration spectrale aurait prédominé. Mais dans le cas d'un phénomène dynamique, c'est la sensation de mouvement de groupe qui prime.

Les glissandi sont liés par une dynamique commune : il s'agit du principe de regroupement par intégration de sorts communs (common fate integration) (Bregman 1990, p.292).

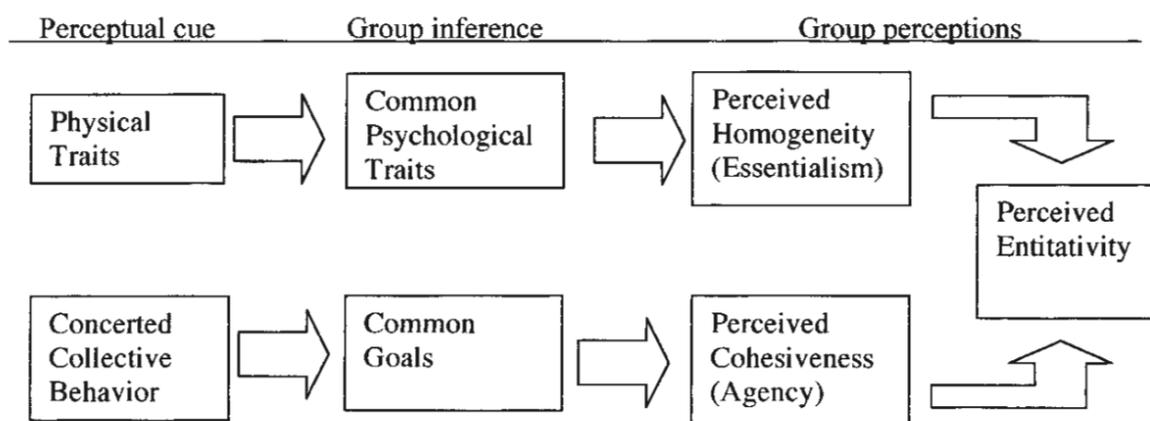
Pour illustrer ces principes par des éléments plus tangibles, nous pouvons faire une analogie avec les principes perceptifs qui régissent la perception d'individus humains au sein d'un groupe : le phénomène d'entitativité.

### Sur la notion d'entitativité

Dans leur étude intitulée Birds of a Feather and Birds Flocking Together: Physical Versus Behavioral Cues May Lead to Trait- Versus Goal-Based Group Perception, les autrices proposent la distinction de deux types de regroupement par inférences perceptives du phénomène d'entitativité : une inférence physique et une inférence comportementale.

Dans la première il s'agit de distinguer ou de regrouper des individus au sein d'un groupe par leur aspect physique : par exemple, si des individus issus d'un groupe scolaire portent le même uniforme, nous les assignerions automatiquement au même groupe.

Dans la deuxième il s'agit d'observer les dynamiques de groupe, leur comportement : par exemple, si des individus au sein d'une manifestation portent des actions communes, nous les assignerions également au même groupe, bien qu'ils puissent avoir des aspects physiques très divers. Ces regroupements seront d'une part perçus comme homogènes par leur caractérisation essentialiste, et d'autre part perçus comme comportant une certaine cohésion de groupe par leur caractérisation comportementale.



Schématisation du modèle dualiste de la perception de l'entitativité (tiré de Wai-man Ip 2006)

Nous pouvons dans une certaine mesure transposer ce modèle à celui de l'analyse des scènes auditives. Les aspects intrinsèques aux unités sonores, c'est-à-dire le timbre, le niveau sonore, la hauteur (si elle peut être clairement perçue), etc. peuvent être assimilés à des caractères essentialistes (plus ces attributs seront proches plus la masse sera perçue comme homogène<sup>9</sup>), tandis que les dynamiques de groupes peuvent être assimilés à des caractères comportementaux.

Ainsi, de la même manière que l'exemple des manifestants, une masse de sons hétérogènes (dont les hauteurs, par exemple, diffèrent) pourra être perçue comme ayant une certaine cohésion si les unités qui la composent ont une dynamique de mode de jeu similaire, assimilé à un caractère comportemental (par exemple un glissando ascendant).

Ce modèle va être particulièrement utile lorsque nous commencerons à aborder les corrélations perceptives entre le modèle issu de modélisations de nuées et le moteur de synthèse sonore pour l'élaboration de "l'orchestrateur de nuées".

Si les glissandi de Xenakis sont aisément appréhendables, par le fait que leurs composantes sont clairement identifiables, il semble moins évident de comprendre les phénomènes perceptifs en jeu dans le Poème Symphonique pour 100 métronomes de Ligeti ou dans l'installation de Zimoun.

En effet, dans ces deux œuvres, les masses d'instruments génèrent leurs sons percussifs de manière aléatoire, ne permettant pas de leur attribuer un regroupement comportemental, et leur nature sonore discontinue ne permet pas d'appliquer les principes de regroupement de sonorités simultanées comme c'était le cas avec glissandi. Quels sont les facteurs qui permettent

#### Sur la notion de distances perceptives : modulations par changement du point d'écoute

Au début du Poème Symphonique comme dans l'installation de Zimoun, rappelons-nous que les centaines d'instruments mécaniques produisent ensemble un tumulte de sons percussifs. Face à cette saturation de sons, notre perception délaisserait alors la compréhension de la manière selon laquelle se structurent les unités sonores à une échelle microscopique, pour se concentrer sur la forme globale à une échelle macroscopique, comme si nous nous tenions à une certaine distance par une prise de recul.

La distinction de ces deux échelles est cruciale afin de comprendre ces phénomènes émergents, et il semblerait que la notion de « distance » soit l'élément conditionnel qui permettrait d'effectuer des allers-retours entre l'un et l'autre.

Si on reprend le cas des nuées d'étourneaux, on s'aperçoit que les mêmes types de phénomènes perceptifs se produisent : nous voyons à la fois des milliers d'unités discrètes qui se détachent du ciel et à la fois une entité globale homogène. Cette ambivalence résulte des interactions internes perçues comme complexes, puisqu'il est impossible de suivre les

---

<sup>9</sup> D'après l'étude de Jason Noble dans "Semantic Dimensions of Sound Mass Music" 2020, le terme homogénéité serait largement associé au timbre. p.221

mouvements individuels de chaque oiseau, et entraîne une réduction de notre attention sur le mouvement dans sa globalité.

L'identification des unités serait alors plus saisissable si nous pouvions opérer un « zoom perceptif », comme si nous venions nous rapprocher d'elles, afin de les isoler du reste du groupe. Cette réduction de distance entre l'unité perçue et le point de vue permet d'inverser la perception du rapport "unité sur groupe", permettant de passer d'une observation *macro* à une observation *micro*.

De manière analogue, si nous reprenons cette fois l'exemple de l'installation de Zimoun, nous pouvons imaginer lors d'une déambulation au sein de l'espace d'exposition un déplacement de notre point d'écoute vers un des automates. Ce rapprochement permet l'isolation d'une source sonore par rapport au reste du groupe, comme si nous inversions un rapport « signal utile sur bruit ».

Ainsi, le déplacement de notre point d'écoute comme de notre point de vue affecterait grandement la perception de l'homogénéité de ces masses. Quels sont alors les facteurs perceptifs qui permettent de dissocier ces unités en opérant ces déplacements ?

Si nous parvenons à isoler perceptivement un des 297 automates en s'en approchant, c'est parce que nous arrivons avant tout à relier un son, un stimuli sonore, à sa source. Cette assimilation source-cause permet l'identification d'une "voix" grâce à ses qualités intrinsèques. Bien évidemment, un des facteurs qui facilite largement cette corrélation provient de la complémentarité de la perception visuelle. Nous arrivons à identifier la source d'une génération sonore par action mécanique tout simplement parce que nous la voyons. Mais si nous faisons abstraction de ce biais visuel, nous pourrions en nous approchant suffisamment d'une source arriver à l'identifier et l'isoler car "nous l'entendrions plus proche que les autres". Cela veut dire que nous arriverons à l'identifier par des attributs essentialistes tels que son timbre, son niveau sonore, mais aussi sa position dans l'espace.

Tous ces attributs ne sont cependant pas absolus, mais sont perçus de manière relative au reste de l'ensemble, puisqu'ils varient dès lors que l'on déplace notre point d'écoute. Le niveau sonore par exemple est un facteur de dissociation dès lors que l'écart relatif entre l'unité perçue et le flux dépasse un certain seuil, de la même manière que la hauteur dans le cas des glissandi.

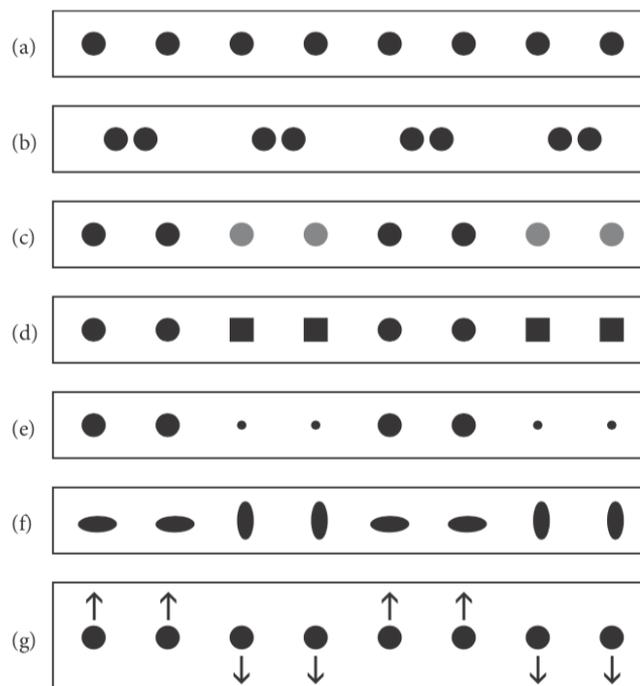
En réalité, tout est une affaire de distances lorsque l'on traite des principes de regroupement. Si l'on assimile les écarts perceptifs mentionnés plus hauts à des distances relatives, alors nous pourrions parler de distance spatiale, mais aussi de distance de hauteur, de niveau sonore, etc. Ainsi, les unités sonores se regroupent en un flux à partir du moment où les distances perceptives qui les séparent restent en dessous d'un certain seuil de différenciation. Autrement dit, les unités sonores auront tendance à se regrouper lorsque leurs attributs partageront des valeurs similaires (niveaux sonores ou timbres proches, proximité spatiale, etc..).

Cette déclaration tombant sous le coup de l'évidence, il est surtout important de rappeler le fait que ces distances relatives, et donc par conséquent les conditions d'émergence de sonorités globales, sont modulables dans une certaine mesure par le déplacement du point d'écoute. Il s'agira d'un point crucial dans l'élaboration de l'espace de déambulation lors de la monstration du projet de partie pratique de mémoire.

### Analogie des attributs de la Gestalt visuelle à l'analyse des scènes auditives

Les travaux de l'école de la Gestalt ont permis de mettre en évidence ces mêmes principes de regroupement, et c'est d'ailleurs à partir de ceux-là que se sont construites les recherches sur l'analyse des scènes auditives (Bregman 1990).

De la même manière que les qualités intrinsèques aux unités sonores (telles que le timbre, niveau sonore, le positionnement dans l'espace), les unités visuelles se dotent d'attributs telles que la forme, la couleur, la taille, ou encore le positionnement dans l'espace. Et similairement, les attributs comportementaux correspondent à la manière dont elles se déplacent au cours du temps.



#### Principes de regroupement Gestalt :

- (a) les unités sont séparées car elles sont à équidistance
- (b) les unités sont regroupées par leur proximité spatiale
- (c) à (f) les unités sont regroupées par la similarité de leur forme et taille
- (g) les unités sont regroupées par leur comportement dynamique

Si nous revenons au cas des nuées d'étourneaux et appliquons ces principes de regroupement, alors nous pouvons en tirer certaines conclusions sur la manière dont nous les percevons. La proximité de leur couleur, forme et taille, de leur proximité spatiale, mais aussi et surtout leurs mouvements dynamiques communs sont des éléments qui permettent l'émergence d'une entité globale.

Par ailleurs, si la nuée est constituée d'un très large nombre d'étourneaux, alors des dynamiques locales peuvent commencer à émerger, créant des variations complexes différentes selon les zones considérées, mais faisant toujours partie d'un même tout, car toujours regroupées par leur similarité physique, leur proximité et leur continuité spatiale, comme peuvent l'être les différents membres d'un même corps.

### La densité comme attribut favorisant la sensation de fusion par intégration séquentielle : étude de cas à travers le Poème Symphonique pour 100 métronomes

S'il existe bien un attribut qui permet de quantifier les masses, il s'agit de la *densité*. Cette unité, bien que très largement employée pour désigner les sensations de masse, n'est pas réellement quantifiable car elle dépend du contexte d'écoute. Il s'agit d'un concept polysémique, utilisé souvent métaphoriquement (associé aux mots blocs, nuages, noeuds, etc.) ou pour désigner à la fois des superpositions de voix, des grands débits rythmiques, des compressions spatiales (comme si les voix se situaient très proches les unes des autres), etc. (Noble & McAdams 2020, p.4, Noble et al 2020, p.220).

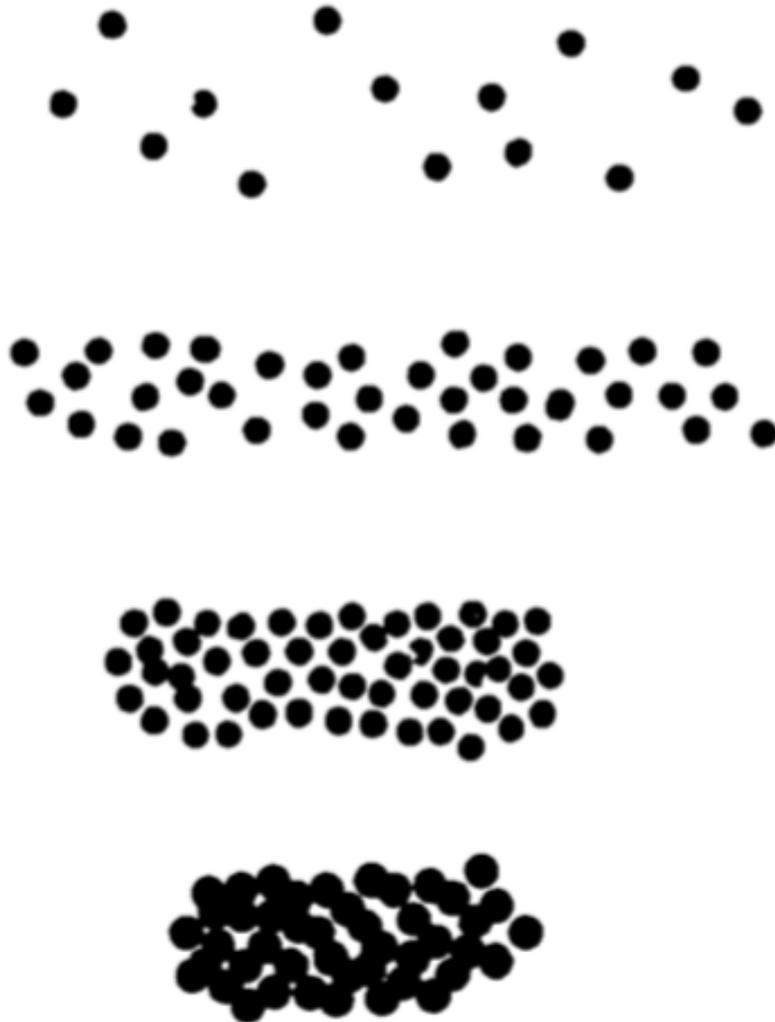
La densité semble être un paramètre primordial qui conditionne la perception d'un regroupement de sons distincts en une seule et même unité. Comment cela peut-il s'expliquer par l'analyse des scènes auditives ?

Quelques minutes après le commencement du Poème Symphonique pour 100 métronomes, un délitement se produit progressivement au sein du nuage de tic tac, laissant apparaître des « trous », zones temporelles dans lesquelles la consistance semble vaciller. L'arrêt successifs des métronomes réglés aux *tempi* les plus rapides fait peu à peu diminuer la densité globale, produisant des discontinuités rythmiques. Cet émiettement donne la sensation d'une dispersion des sons en de multiples unités distinctes, et la perception d'une sonorité globale continue semble se dissoudre petit à petit.

Ces sensations de dissociations d'unités peuvent s'expliquer par le fait que nous avons tendance à regrouper des unités se trouvant temporellement les unes à la suite des autres : il s'agit du principe de regroupement par intégration séquentielle (Bregman 1990, p.47). Ici encore nous pouvons utiliser la notion de distance temporelle séparant les unités entre elles afin de relier ces effets à ceux mentionnés précédemment.

Lorsque la densité de tic tac se réduit, il se produit statistiquement un allongement moyen d'écart temporel des unités sonores entre elles.

De manière analogue à la perception de la densité d'unités visuelles au sein d'une scène, nous pouvons corréler la similarité des effets que produisent ces écartement temporels aux effets que produisent les écartements spatiaux dans le champ de la vision.



*L'augmentation progressive de la densité produit un effet de regroupement d'unités*

Si pour le champ de la vision l'axe des abscisses correspondrait à une dimension de l'espace, elle pourrait être assimilée à une échelle temporelle dans le domaine sonore.

Cette sensation d'augmentation de densité comme une « augmentation d'événements sonores dans le temps » peut résulter en une sensation de continuité d'une forme globale émergente.

Nous retrouvons ce même phénomène avec par exemple les roulements entretenus de caisse claire. L'accélération des battements de baguette contre la membrane, rendus possibles par un jeu de rebonds très courts et rapprochés, permet d'obtenir un effet sonore ressemblant à un bourdonnement régulier, et dont les impacts sont rendus diffus. Les impacts comme des unités distinctes parviennent à se fondre dans le flux, par augmentation de la densité rythmique, et forment une texture sonore. La notation musicale du roulement de caisse claire est par ailleurs écrite comme une note tenue et suspendue.

**Rolls**

Open roll      Open roll (written out)      Buzz roll/Closed roll      Buzz roll/Closed roll (written out)

Single-stick buzz    Crushed buzz    Slowing down    Speeding up    Crescendo roll    *fp* Drop-away buzz

Différentes notations musicales du roulements de caisse claire, entre triples croches et notes tenues

Le rôle de la composante transitoire dans la détection d'unités au sein d'un flux

Même si chaque unité sonore se démarque d'une autre par le fait que ses attributs soient différents d'une autre, il semblerait que des caractères "doux" ou "saillants" de sons leurs permettent respectivement de davantage se fondre, ou au contraire émerger de la masse.

Cela peut être expliqué par le fait que notre système de perception auditive est sensible aux sons à forte composante transitoire. Au sein d'un flux composé d'un séquençage d'unités discrètes, plus une unité aura une composante transitoire forte, plus elle aura tendance à se démarquer et à être perçue comme une unité indépendante du flux. *A contrario*, plus une unité aura une attaque lente, plus elle aura tendance à s'intégrer dans la masse<sup>10</sup> (Bregman 1990, p.71).

<sup>10</sup> "It is likely that it is really the rate of rise in intensity that controls the formation of units. A possible illustration of this fact is our perception of a rapid series of guitar tones. A plucked string sound has a very sudden attack but decays gradually so that rapidly played tones will overlap in time. Despite this, in a rapid series of plucked guitar notes, each one sounds like a discrete event. In our laboratory, we have synthesized these tones backwards so that the onset is slow and the decay rapid, maintaining the same overlap in time. When we listen to these, the discreteness of the individual sounds is less definite (and the pitches are less precise). Apparently, abrupt rises in intensity tell the auditory system that a new sound has joined the mixture and that it should begin the analysis of a new unit." p.71

Ainsi, nos oreilles seront très sensibles aux sons percussifs, dont l'enveloppe dynamique contient une attaque et une résonance très courtes, comme peuvent l'être les métronomes de Ligeti. Malgré leur grande densité, les tic tac ont en réalité du mal à fusionner entre eux pour faire émerger un son véritablement continu, et la moindre variation de densité sera perceptible.

#### A propos de l'association d'un motif rythmique à une unité

Pour terminer notre analyse du Poème Symphonique, il est également important de noter que la nature périodique des métronomes permet de les identifier comme des unités sonores distinctes. En effet, nous avons aussi tendance à regrouper des unités ensemble dès lors que nous percevons une structure périodique, un motif rythmique (Bregman 1990, p.411).

Lorsque la majorité des métronomes se sont arrêtés, nous avons la possibilité d'associer un "tic tac" à un métronome car nous en percevons le motif répétitif. La dissociation par localisation dans l'espace est un autre facteur pouvant faciliter cette identification.

Ce caractère périodique doit néanmoins satisfaire certaines conditions, notamment le fait que sa fréquence ne doit pas être trop basse, car elle ne permettrait pas alors la perception d'une continuité entre les motifs joués.

Prenons par exemple le cas d'une autre installation de Zimoun : 127 prepared dc-motors, sticks, 30cm, 2015<sup>11</sup>. Ici les automates émettent un son de manière régulière grâce à l'utilisation de moteurs, comme les métronomes, mais leur période de révolution étant longue (environ 7 secondes), il est impossible (si l'on ferme les yeux) d'associer un motif rythmique à un automate. L'identification par une corrélation rythmique source-cause est donc compromise si la fréquence d'émission est trop faible.

Ces considérations vont être importantes lorsque nous allons traiter de l'implémentation de l'algorithme de "l'orchestrateur de nuées", notamment dans l'identification d'une voix par motif rythmique.

---

<sup>11</sup> Captation de l'installation visionnable via ce lien : <https://vimeo.com/126609082>

### Synthèse et définition de masses sonores :

En analysant le corpus par le prisme de l'analyse de scènes auditives, nous avons pu comprendre la manière dont nous avons tendance à regrouper des unités disparates en un ensemble, un flux, en identifiant les facteurs. La perception de fusion ou de fission d'unités vis-à-vis du flux peuvent être quantifiées par ces facteurs : fréquentiel, timbral (notamment le caractère percussif), niveau sonore, localisation spatiale, dynamiques communes, perception de motifs, etc.

Ainsi, nous avons pu identifier deux types d'intégrations d'unités au sein d'un flux : simultané (correspondant à la fusion de sons continus, comme les clusters) et séquentiel (correspondant à la fusion de sons discontinus, comme les nuages de grains).

Nous avons ainsi compris que la densité agissait comme un attribut quantitatif permettant de changer la nature qualitative de la sonorité composée. Plus la densité de la masse sera grande, plus il y aura de chance qu'une sonorité composée émerge.

Cet ensemble ainsi constitué, fusionné, ou plutôt *composé*, peut contenir des qualités qui lui confèrent une certaine homogénéité, une certaine autonomie, et par laquelle on peut la désigner en tant que tout. C'est le cas par exemple de "nuages de sons", ou "pluie de sons" désignant les sonorités issues de métronomes de Ligeti ou des pizzicati de Xenakis. En ce sens, il constitue une sonorité émergente.

Cependant, malgré l'autonomie et l'homogénéité de celles-ci, ces masses donnent la sensation d'une multiplicité, nous percevons toujours qu'elles sont constituées de multiples parties qui la composent. C'est cette ambivalence, la complexité paradoxale, qui caractérise la nature des masses sonores.

Enfin nous proposons une définition synthétique des masses sonores, proposée par Noble & McAdams (2020, p.9) :

*Masse sonore : Unité auditive dense et perceptivement homogène intégrant de multiples événements ou composantes sonores tout en conservant une impression de multiplicité. Bien que leurs corrélats acoustiques puissent être très complexes, les masses sonores sont homogènes sur le plan perceptif car elles résistent à la ségrégation perceptive d'un ou de plusieurs paramètres (par exemple, la hauteur, le rythme, le timbre).<sup>12</sup>*

Ces qualités semblent se conjuguer avec celles observées dans les nuées d'étourneaux. Les tentatives de corrélations par ces principes de regroupement entre le domaine visuel et le domaine sonore seront abordées dans la Partie 3, dans le cadre de la création du *synthétiseur de nuées*.

---

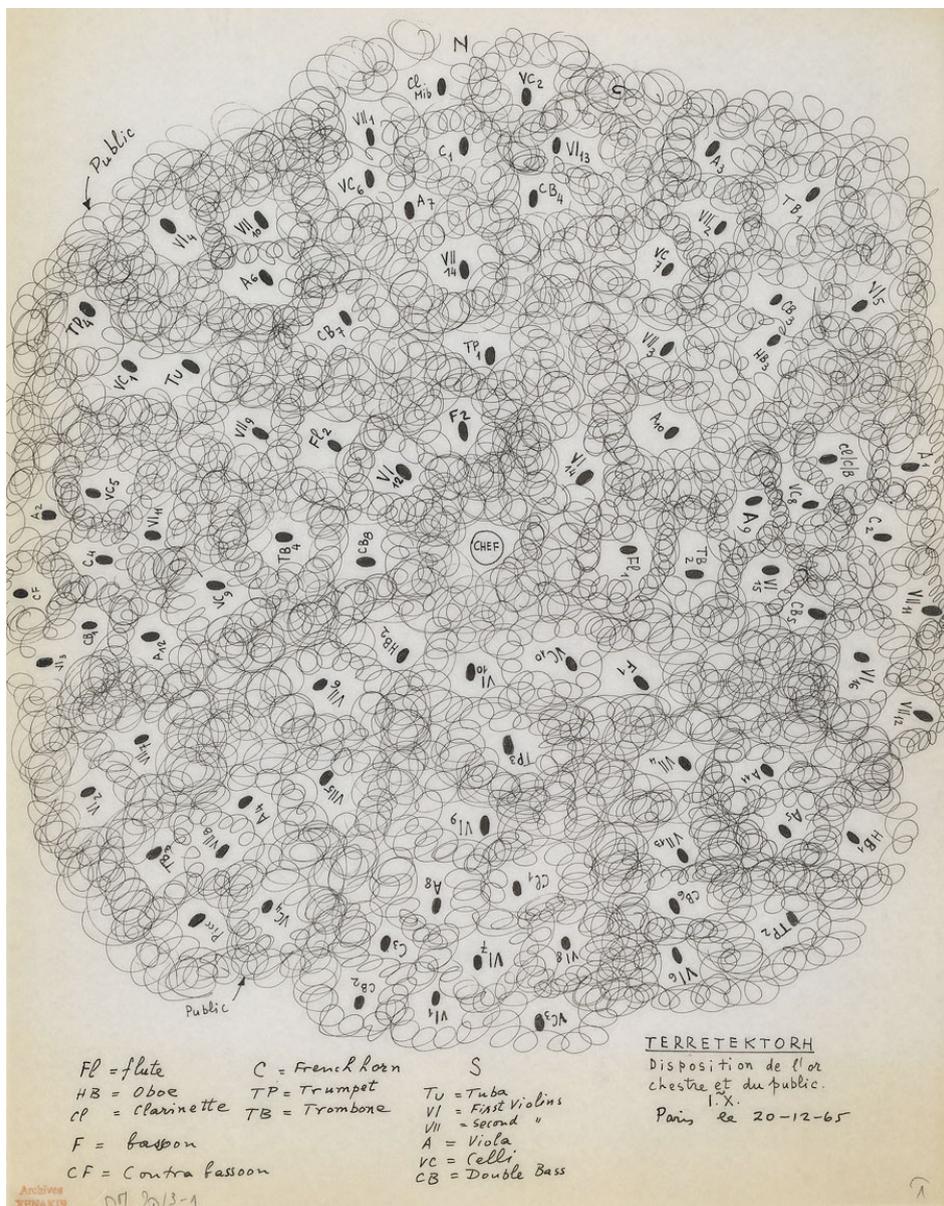
<sup>12</sup> Citation originale : "Sound mass: A perceptually homogeneous and dense auditory unit integrating multiple sound events or components while retaining an impression of multiplicity. Although their acoustical correlates may be highly complex, sound masses are perceptually simple because they resist perceptual segregation in one or more parameters (e.g. pitch, rhythm, timbre)."

De la même manière que les étourneaux, nous sommes au jour le jour confrontés à ces phénomènes d'émergence par masses sonores, à l'écoute de la pluie ou des bruissements de feuilles par exemple. Cette sensation ambivalente entre multiplicité et homogénéité est également caractéristique de ces phénomènes sonores naturels.

Il sera question dans la partie suivante d'explorer la manière dont ces phénomènes ont pu être source d'inspiration et de modélisation, comment les compositeurs ont tenté de transposer par la simulation de leurs modèles, ces effets d'émergence.

## Partie 2. Les phénomènes naturels comme modèle de composition : articulation de textures sonores par modélisations physiques

« L'auditeur sera [...] soit perché sur le sommet d'une montagne au milieu d'une tempête l'envahissant de partout, soit sur un esquif frelé en pleine mer démontée, soit dans un univers pointilliste d'étincelles sonores, se mouvant en nuages compacts ou isolés » (à propos de sa pièce *Terretektorh*, Xenakis, 1969).



Plan de disposition de l'orchestre (représenté par les points noirs) et du public (représentés par les cercles entremêlés) pour le concert de Terretektorh.

Les compositeurs utilisant des masses sonores comme moyen de composition tels que Xenakis, Ligeti ou encore Varèse ont systématiquement utilisé un vocabulaire très diversifié pour décrire leurs œuvres, faisant souvent références à des phénomènes naturels.

“Nuages, tempête, pluie, grondements, gaz, étincelles, etc...” sont autant de concepts hautement métaphoriques fréquemment employés, mais qu’il semble difficile de corréler directement à des attributs sonores ou musicaux.

On ressent par exemple chez Xenakis une certaine fascination pour les phénomènes météorologiques tels que les sons de la pluie ou de la grêle, des tempêtes et orages<sup>13</sup>, mais aussi du son de cigales chantant par centaines. Il évoque dans un texte autobiographique sa jeunesse à Athènes avant l’arrivée de la seconde-guerre mondiale :

*« Dans ma jeunesse, je me destinais à l’archéologie, sans doute parce que je vivais plongé dans la littérature antique [...]. Parallèlement, je m’intéressais beaucoup aux mathématiques et aux sciences, en particulier à l’astronomie [...]. Et j’aimais infiniment la nature. J’allais à bicyclette à Marathon. À l’endroit supposé de la bataille, il y avait un tumulus avec un bas-relief d’Aristoclès, et là je restais longtemps à m’imprégner des bruits de la nature, des cigales, de la mer. Avant d’avoir lu Debussy, qui parle si bien du vent, Debussy que j’ai entendu beaucoup plus tard, j’avais déjà ressenti les mêmes impressions que lui » (Xenakis 2006, p.17)*

Mais si ces inspirations ont guidé certains compositeurs à tenter de reproduire leurs effets par le biais de suggestions métaphoriques, comme il y en a toujours eu dans l’histoire de la musique (chez Debussy par exemple, pour ne citer que lui), il y a eu chez Xenakis une volonté de reproduire ces effets par le biais de modélisations physiques. En effet, à partir des années 60, les découvertes successives de théories scientifiques telles que la théorie de la morphogénèse, la théorie des catastrophes, les théories du chaos, la géométrie fractale, etc. mais aussi des théories de l’information, ont progressivement influencé certains compositeurs dans leur manière d’aborder leurs musique, changeant notamment leurs visions de la forme musicale<sup>14</sup>.

L’utilisation d’éléments sémantiques propres aux phénomènes naturels témoignent de la diversité et de la complexité des sonorités que les compositeurs tentaient de reproduire par la mobilisation de masses sonores.

Dans le cadre de ce mémoire, et de la même manière que ces compositeurs, ce sont les nuées d’étourneaux qui ont été ma source d’inspiration principale. Il ne s’agissait néanmoins pas pour moi de reproduire artificiellement un phénomène naturel mais bien de tenter d’en reproduire les dynamiques pour les insuffler dans une application musicale.

---

<sup>13</sup> Mâkhi, sa fille, raconte un épisode orageux lors de leurs vacances d’été : « L’orage éclate au milieu de la nuit. Bruit d’abord lointain du tonnerre. Nous sommes tous les trois dans la tente, nous comptons les secondes. Illumination de la toile par intermittence. Orange, noir, orange, noir. La pluie commence à entrer à l’intérieur. L’orage maintenant est sur nous. Percussions fracassantes. Mon père est comme fou, il doit sortir, il doit aller voir. Je le suis. Très vite, nous sommes trempés, il court vers le sommet de la montagne. Il m’a oubliée. Il court comme pour entrer à l’intérieur du ciel rempli d’éclairs et de sons ». (M. Xenakis, 2002 : 40).

<sup>14</sup> Pour une recherche plus approfondie, consulter “Entre naturalisme sonore et synthèse en temps réel, images et formes expressives dans la musique contemporaine” de Márta Grabócz (2013).

Dans quelle mesure pouvons-nous utiliser le modèle des nuées pour composer le phénomène d'émergence ?

En restant toujours dans le cadre du paradigme des masses sonores, il s'agira alors de comprendre par quels moyens compositionnels il serait possible de simuler leurs effets, à la fois par l'étude des matériaux sonores mobilisés, mais aussi par la manière dont ils sont agencés au sein d'une composition.

Les travaux de Xenakis, entre la pensée granulaire et la pensée systémique stochastique, constituent les socles théoriques de mon développement.

## I. Composer avec le son.... composer le son, du grain à la texture, paradigme granulaire

*"Il faudrait imaginer un son complexe comme un feu d'artifice de toutes les couleurs dans lequel chaque point lumineux apparaîtrait et disparaîtrait instantanément sur le ciel noir. Mais dans ce feu il y aurait tellement de points lumineux et ils y seraient ainsi organisés que leur succession rapide fourmillante créerait des formes, des volutes à déroulement lent ou au contraire des explosions brèves incendiaires de tout le ciel. Une ligne lumineuse serait constituée par une multitude suffisante de points apparaissant et disparaissant instantanément." (Xenakis 1963, p.61 chap. II Musique Stochastique Markovienne - Partie théorique)*

### A. La pensée granulaire

Dans le chapitre 2 de Musiques Formelles (1963), Iannis Xenakis expose une nouvelle théorie, une proposition singulière énonçant une nouvelle manière de décrire le phénomène sonore, différente de celle de l'analyse spectrale, omniprésente dans le paysage de la théorie musicale et des sciences du son. En analogie directe avec le principe de dualité onde-corpuscule, il introduit une théorie granulaire en considérant la propriété corpusculaire - et non plus ondulatoire - du phénomène sonore. Il s'agit d'énoncer la possibilité de décrire tout son comme une décomposition de suite de "grains" élémentaires, échantillons sonores extrêmement courts. Cette théorie est comme le pendant temporel de la décomposition harmonique de l'analyse de Fourier, selon laquelle il est possible de décomposer tout son en une somme de sinus :

*"Tout son est une intégration de grains, de particules élémentaires sonores, de quanta sonores. Chacun de ces grains élémentaires a une triple nature : la durée, la fréquence et l'intensité. Tout son, toute variation sonore même continue est conçue comme un assemblage de grains élémentaires suffisamment nombreux et disposés dans le temps d'une façon adéquate. » (Xenakis 1963, p.61)*

Ce "paradigme granulaire", comme l'appelle Makis Solomos (2006), a de nombreux antécédents théoriques et musicaux, en plus de la théorie onde-corpuscule<sup>15</sup>. Mais c'est véritablement le physicien Dennis Gabor qui avec son article Acoustical Quanta and the Theory of Hearing (1947) transpose de manière effective cette théorie corpusculaire dans le domaine du son. Il développe l'idée que *"tout son peut être décomposé en une famille de fonctions obtenues par des changements temporels et fréquentiels d'une seule particule gaussienne. En d'autres termes, tout son peut être décomposé en une combinaison appropriée de milliers de grains élémentaires"*<sup>16</sup>.

On peut opérer un rapprochement entre cette description corpusculaire, qui décrit l'émergence d'un flux continu par juxtaposition de grains sonores, et le principe de fusion par agencement séquentiel d'unités sonores élémentaires, comme nous l'avons étudié dans la Partie 1.

Mais le véritable pouvoir de cette théorie réside dans sa réciproque : si tout son est décomposable en une suite de grains, alors il serait possible de recomposer tout type de sonorités par un agencement particulier de grains.

Ainsi, une nouvelle voie s'ouvrirait pour Xenakis et les compositeur·trices qui lui ont succédé, voie menant à la *"création de matériaux sonores nouveaux et à des formes nouvelles"*<sup>17</sup>.

*"Qu'est-ce qu'un son, peut-on se demander avec la conception granulaire ? Est-ce un grain ? Est-ce un amas de grains ? Est-ce une partie d'une œuvre granulaire ? Ou même, est-ce l'œuvre entière ? De la microstructure à la macrostructure de l'œuvre, du micro- au macrotemps, du son au sens physique (matériau) à l'œuvre entière (forme), naît un continuum. Le granulaire nous invite à penser une composition musicale comme une entité totalement construite à partir de grains. C'est pourquoi sa logique, tout en étant liée à une technologie (un type de synthèse sonore), est plus qu'une technologie : elle repose sur une manière de penser la musique : on pourrait parler d'un paradigme ou d'une sensibilité granulaire. Ce paradigme ne se limite donc pas à un penchant pour des sonorités de type granulaire : il s'intéresse à la continuité entre micro- et macrotemps, qui définit la musique comme une totalité composable de bout en bout. » (Makis Solomos, 2013, "De la Musique au Son", chap 5. Composer le son, p.257)*

## B. Micro-sons et grain

### La pensée granulaire dans la musique

---

<sup>15</sup> Pour une recherche approfondie, consulter "The granular connection" de Makis Solomos (2006), ainsi que le chapitre 2 de "Microsound" de Curtis Roads (2001).

<sup>16</sup> traduit par Makis Solomos (2013, p.256), citation issue de Roads (2001), p.57 : *"In Gabor's conception, any sound can be decomposed into a family of functions obtained by time and frequency shifts of a single Gaussian particle. Another way of saying this is that any sound can be decomposed into an appropriate combination of thousands of elementary grains"*.

<sup>17</sup> Mention de Xenakis dans l'introduction du chapitre 2 de Musiques Formelles (1963), précédant l'introduction de sa théorie granulaire

*« Le musicien travaille avec des discontinuités pour construire des continuités. Comme Robinson dans son île, il cherche des façons d'échapper à un temps amorphe, inarticulé, en creusant des marques, des inscriptions sur un support - des éléments discrets, des symboles : des « manettes » qu'il utilise pour atteindre un certain niveau opératoire et morphologique. »*  
(Vaggione, 2010, p.45)

En analogie avec l'art de la mosaïque, les grains peuvent être assimilés aux *tesselles*, fragments de pierres, d'émail ou de céramique, ayant chacun leur propre forme, texture, couleur, et qui, juxtaposés les uns aux uns selon une certaine disposition, permettent de faire émerger un motif, une figure holistique à grande échelle.

Ces jeux de décomposition de la matière, puis de recombinaison sont propres au style musical issu de cette "sensibilité granulaire" : la micro-composition.

Une des premières tentatives de mise en œuvre de ces modes de composition remonte à la pièce *Concret PH* composé Xenakis en 1958. Il s'agit d'une pièce faisant seulement usage d'un matériau sonore "concret" : un enregistrement de sons de braises sur bande magnétique. Xenakis compose cette pièce en agencant différentes parties de l'enregistrement par un procédé de montage, des morceaux de bandes mises bout à bout : il découpe, au sens littéral du terme, le son pour le recomposer selon ses désirs. Mais la micro-composition, c'est-à-dire la composition par la manipulation de sons extrêmement courts et appartenant au micro-temps, devra attendre l'arrivée quelques dizaines d'années plus tard de l'ordinateur et des technologies numériques afin de pouvoir se déployer pleinement. Grâce aux logiciels de montage, il devient alors possible d'articuler finement les agencements de grains à une échelle bas-niveau, permettant de composer des sonorités inaccessibles jusqu'alors.

Dans *Consort for Convolved Violin* composé par Horacio Vaggione, l'une des figures majeures de ce courant, toute la composition est construite à partir d'enregistrements de violons. En extirpant des échantillons très courts de ceux-ci, il construit sa pièce en réorganisant ces micro-sons selon plusieurs échelles de représentation.

Au sein de logiciels de montages dédiés, tels que IRIN (Caires 2004), ces grains ont la possibilité d'être manipulés à la manière de notes, permettant de composer des pièces entières en arrangeant des stratifications polyphoniques, suivant des structurations hiérarchiques propres à la composition musicale.

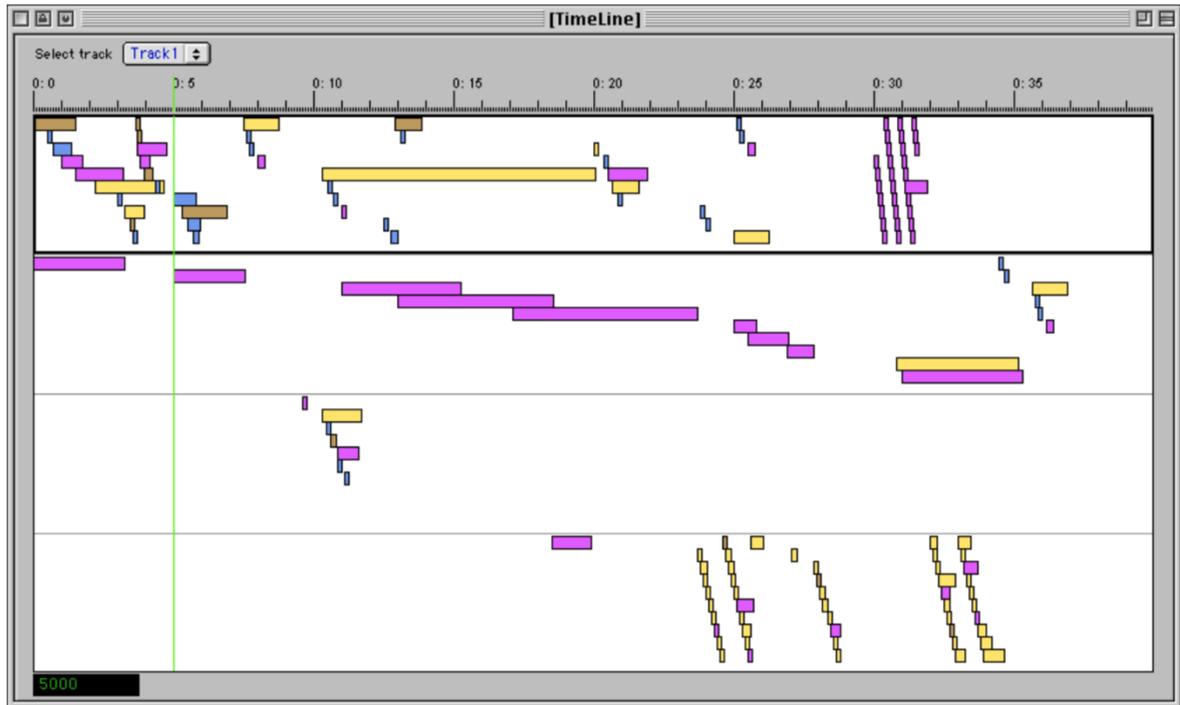


Figure 7. Visualisation selon le mode « BuffersView »

Logiciel de montage IRIN, dédié au micro-montage, développé par Carlos Caires

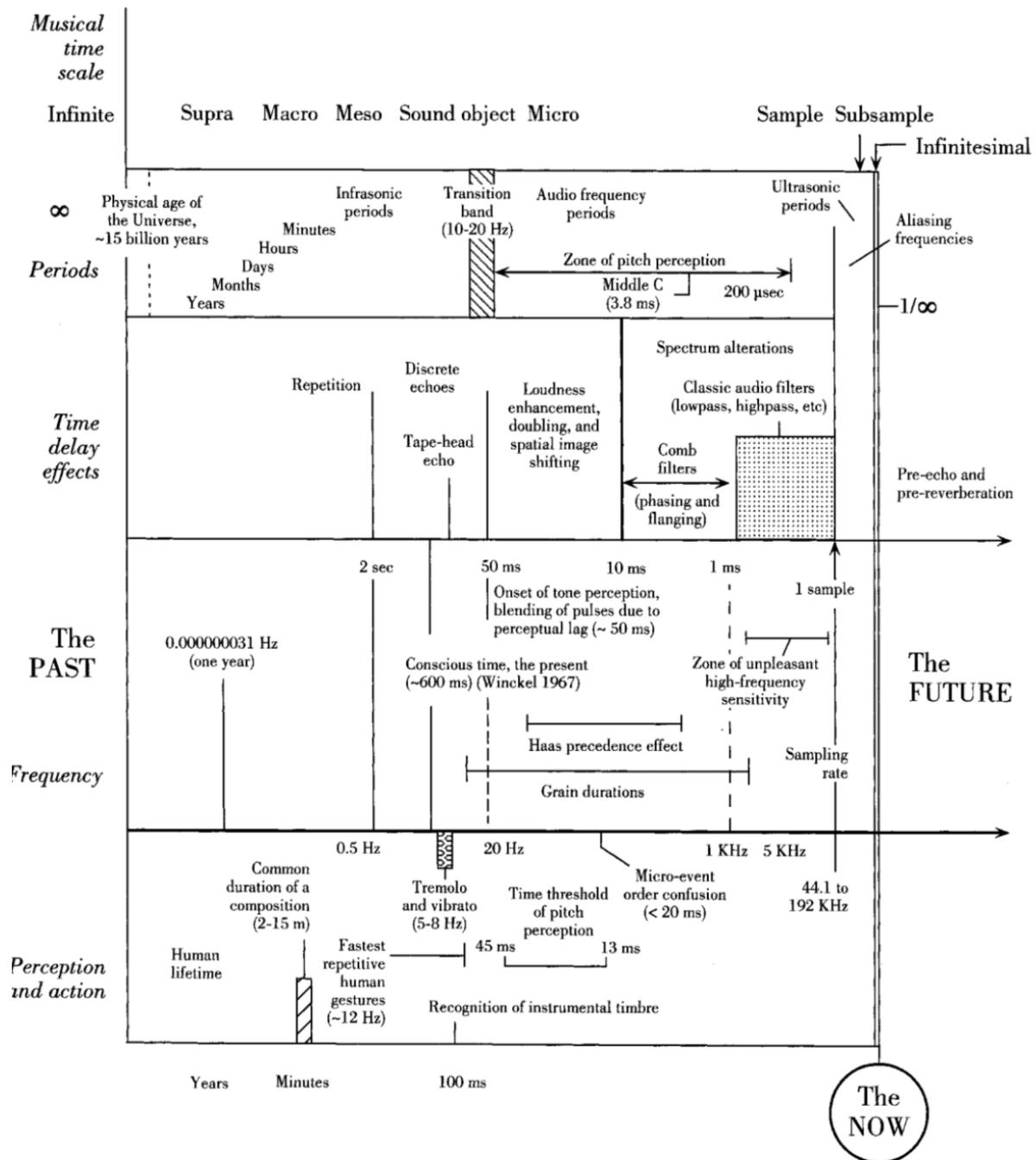
L'intérêt de l'approche granulaire pour Vaggione, réside dans ce transfert catégoriel du niveau du signal (du grain) vers le niveau symbolique (de la note) :

*“L'intérêt de l'approche granulaire, pour la composition musicale, consiste donc dans le traitement symbolique d'éléments présents à l'échelle du micro-temps. Les grains, pris dans ce sens, constituent des éléments ductiles avec lesquels on peut travailler des entités morphologiques, en les agglutinant et en les projetant partout dans le composable, à toutes les échelles possibles. En partant d'une approche du sonore marqué par la discontinuité et la fragmentation, nous créons pour ainsi dire directement des objets musicaux, des objets composés” (Horacio Vaggione, 2006: “Symboles, signaux, opérations musicales”)*

Il s'agit donc pour Horacio Vaggione d'aller chercher des sons *en dessous des notes*, des sons fugaces, éphémères, qui échappent parfois à notre perception : les micro-sons.

### Caractérisation du micro-temps

Dans le chapitre 1 de son livre Microsound (2001), Curtis Roads établit une décomposition du temps selon 9 échelles, utilisées pour articuler des structures sonores musicales. Cette stratification a pu être construite par l'organisation et les hiérarchisations traditionnelles utilisées pour construire des compositions musicales, mais également par des études psychoacoustiques, notamment lorsque l'on descend à l'échelle du microtemps. Par exemple, une composition musicale entière, se mesurant en minutes, heures ou parfois dans certains cas extrêmes en jours, sera perçue dans sa forme globale *macroscopique*. Au sein de cette composition s'articulent des phrases de quelques secondes ou minutes, à l'échelle *mésoscopique*, qui contiendront une organisation d'*objets sonores* (traditionnellement des notes), se mesurant en une portion de secondes ou secondes. Toute la musique classique occidentale s'est servie de cette dernière échelle comme base unitaire de toute composition. Il s'agit de la plus petite unité notable sur une partition, le symbole élémentaire d'un événement pouvant être identifié traditionnellement comme musique. L'échelle microscopique, se mesurant en millisecondes jusqu'en centaines de millisecondes, se trouve donc en dessous de celle des objets sonores.



**Figure 1.1** The time domain, segmented into periods, time delay effects, frequencies, and perception and action. Note that time intervals are not drawn to scale.

*Le temps segmenté en différentes échelles (tiré de Roads 2001)*

Si les micro-sons échappent parfois, par leur nature éphémère, à notre perception, nous y sommes en réalité confrontés au quotidien : du clic de souris d'ordinateur au son d'une brindille qui casse, du claquement de doigt à un bruit de couvert, du clic de stylo au son de l'interrupteur d'éclairage, etc. la liste peut s'étendre indéfiniment.

Ces micro-phénomènes sonores sont caractérisés par leur nature transitoire, percussive, et se manifestent dans le mode réel par la collision ou le frottement d'un objet sur un autre, ou par la fragmentation de celui-ci. Ils sont donc caractérisés par leur nature fugace, à énergie non entretenue, disparaissant aussitôt qu'ils sont apparus.

S'ils se manifestent par le biais d'actions mécaniques dans le monde réel, ils peuvent être synthétisés dans le domaine numérique par la fragmentation du matériau sonore lui-même, en venant sélectionner un morceau particulier d'un son qui n'est pas nécessairement un micro-son.

C'est ce que fait Horacio Vaggione dans sa même pièce Consort for Convolved Violin (2011), qui segmente la morphologie d'une note pour en extraire seulement ses qualités transitoires, telles que l'attaque, caractérisée par le bruit de l'écrin sur la corde. Ainsi, la manipulation de microsons permet un "*affranchissement par rapport aux causalités instrumentales*"<sup>18</sup> (Vaggione, 2010), permettant de créer de nouvelles morphologies à partir de sons réels, décorréliées de leur réalité matérielle : l'attaque d'un son peut ne pas être suivi de sa résonance.

Bien sûr, la division de ces échelles temporelles dépend du contexte dans lequel elles se déploient. Dans le cadre de mon projet de recherche, il me paraissait plus intéressant de définir le phénomène micro-sonore non pas en fonction d'une restriction de durée, mais plutôt selon la manière dont il est perçu. Étant donné que j'exploite, pour *le synthétiseur de nuées* (voir Partie 3), des sons réels sans les altérer dans leurs durées, je considérerais dans ce mémoire, comme microson, tout son caractérisé perceptivement par sa nature intrinsèquement éphémère. Évidemment, il s'agit ici de prendre en considération la capacité à reconnaître l'unité sonore manipulée, le rapport source-cause est donc ici capital.

En effet, l'écoute d'un son percussif "concret" appelle une matérialité physique sous-jacente : l'action mécanique par laquelle elle est produite. Que ce soit la frappe d'une touche sur un clavier d'ordinateur, la pose d'un couvert sur la table, ou que ce soit quelque chose qui frappe, claque, tape, éclate, casse, rappe, frotte, caresse, etc... le micro-son convoque un imaginaire qui évoque une unique action mécanique indécomposable, qui "d'un seul coup" fait surface pour disparaître aussitôt. Ainsi, un micro-son pourra, dans le cadre de ce mémoire, parfois durer quelques centaines de millisecondes s'il respecte ces conditions.

---

<sup>18</sup> Citation originale : "*La musique électroacoustique nous donne accès à des morphologies musicales qui ne peuvent pas être obtenues par d'autres moyens. Étant des entités temporelles très feuilletées, étant donné surtout les possibilités offertes par un affranchissement par rapport aux causalités instrumentales, les morphologies électroacoustiques se trouvent pour ainsi dire directement ancrées dans le microtemps: les contractions et dilatations des matières, les granulations, les convolutions entre diverses microéchelles, sont des choses qui font partie de ce qui est à composer*" (Vaggione, 2010)

Mais, si ces micro-phénomènes sonores échappent parfois à notre attention, c'est parce qu'ils surgissent la plupart du temps en masse dans le monde réel. En effet, réunis par milliers, ils sont caractérisés par des sonorités aussi variées que des bruits de pas sur du gravier, un versement de lentilles ou grains de riz dans un bol, de la pluie comme de la grêle, le gargouillements d'un ruisseau, etc.. Toutes ces sonorités sont communément regroupées sous le terme de *textures*, et sont caractérisées par leur caractère granulaire. Il s'agira de trouver une définition appropriée à ce terme polysémique, afin de pouvoir identifier ce qui fait le lien entre tous ces sons.

### Du grain à la texture

Parmi tous les exemples que nous avons cités, la texture sonore semble être caractérisée par le fait qu'elle n'est pas "lisse". En effet, elle *"laisse entendre sa texture granuleuse : l'oreille reconnaît en quelque sorte qu'[elle] est composée au sens littéral du terme (composée de plusieurs grains)"*. (Solomos 2013, p.256).

D'après MacKay *"il existe un certain nombre de facteurs convergents qui donnent naissance au phénomène auditif particulier que l'on appelle généralement "texture". La qualité fondamentale du percept textural en musique est qu'il s'agit essentiellement d'une image globale stable, composée d'une multitude d'événements constitutifs, de la même manière que l'image visuelle d'un tapis est le produit d'un grand nombre de poils ou de fils."* (MacKay 1984, p.171).

Il est par ailleurs intéressant de noter que l'étymologie du mot provient justement du latin *textura* : "l'action de tisser", c'est-à-dire l'assemblage d'un ensemble de fils distincts selon un motif régulier, qui une fois entrelacés, forment un tissu.

Il ajoute : *"Bien que nous soyons conscients de la micro-structure de l'image sonore, il est à bien des égards plus facile et plus attrayant de concentrer son attention sur le percept global. Malgré la prédominance des caractéristiques globales de la texture, la micro-structure peut attirer une attention fugace et momentanée par le biais d'incohérences imprévisibles qui perturbent temporairement la régularité de la texture. La perception texturale semble donc impliquer une interaction entre l'attention auditive, qui est diffusée sur une masse sonore étendue et verticalement complexe, et la conscience des événements microstructuraux isolés qui composent la masse sonore."*<sup>19</sup>

Cette dichotomie perceptive en deux niveaux, entre microstructure et macrostructure, caractérise respectivement la sensation de multiplicité et d'homogénéité des textures sonores. La qualité "granulée" correspondrait donc à notre capacité de percevoir un

---

<sup>19</sup>Citation originale : *"There are a number of convergent factors which give rise to the particular auditory phenomenon which is generally understood as "texture". The fundamental quality of the textural percept in music is that it is essentially a stable global image, comprised of a multitude of component events much in the same way as the visual image of a rug is a product of a large number of hairs or threads. Although we are aware of the micro-structure of the sound image, it is in many important ways easier and more inviting to focus attention on the global percept. Despite the dominance of the global features of the texture, the micro-structure can attract fleeting and momentary attention through any unpredictable inconsistencies which temporarily disturb the regularity of the texture. Textural perception therefore seems to involve an interaction between the auditory attention which is diffused over an expansive and vertically complex sound mass and the awareness of the isolated microstructural component events which make up the sound mass."* (MacKay 1984, p.171)

agencement d'éléments microscopiques au sein d'une image sonore statistiquement stable. Cette définition rejoint à bien des égards celle des masses sonores énoncée à la fin de la Partie 1, au détail près qu'elle correspond à celle du principe de regroupement par agencement séquentiel d'unités sonores.

Cette définition est également transposable dans le champ de la vision, ainsi que dans le domaine haptique : on utilise aussi bien le terme texture pour désigner une "image avec du grain", que de la sensation d'une surface granulée. Il est donc probable que l'origine de l'utilisation de ces concepts dans le domaine sonore provienne de l'expérience d'écoute de sonorités issues du frottement d'un élément physique sur une surface granulée. Le son du roulement d'une valise sur une surface goudronnée en est un exemple typique.

### Perception de la texture : le coefficient de remplissage

Comme j'ai eu l'occasion de le pointer dans la Partie 1, le passage du grain à la texture se fait par un changement quantitatif, en augmentant progressivement le paramètre de densité. A l'approche d'une averse, les sons des gouttes de pluie se détachent en frappant unes à unes le toit d'un abri. Quelques secondes plus tard, après que leur densité ait augmenté, chacune d'entre elles se fond progressivement dans la masse, laissant place à la texture pluviale globale. A partir de quel instant percevons-nous ce changement ? Comment percevons-nous la consistance de ces nuages de micro-sons lorsqu'ils sont formés ?

Curtis Roads nous donne certains éléments de réponse grâce au coefficient de remplissage (Fill Factor) (2001, p.105-6). Il s'agit d'un calcul issu du produit entre la durée du grain et sa densité (nombre de grains par secondes), permettant d'avoir une indication de la consistance d'un nuage de grains. Par exemple, avec un coefficient de remplissage d'une valeur dépassant 1.0 (prenons 20 grains d'une durée de 100 ms chacun, équivalant à un coefficient à égal à :  $20 * 0.1 = 2.0$ ), le nuage sera perçu comme "plein", tandis qu'avec un coefficient inférieur à 0.5 (prenons 10 grains de 40 ms :  $10 * 0.04 = 0.4$ ), nous percevons un nuage poreux et fluctuant.

Ce coefficient agit comme un indicateur plus que comme une valeur déterminante, car dans le cas de déclenchements de grains aléatoires, comme dans l'installation de Zimoun ou dans le Poème Symphonique de Ligeti, les micro-sons peuvent ne pas nécessairement se suivre les uns à la suite des autres, créant à certains moment quelques sensations de "trou". Mais afin de "créer ce que nous entendons comme un nuage consistant, une bonne règle de base est de régler la densité par seconde du nuage à au moins  $2/D$ <sup>20</sup>, avec D la durée du grain.

Prenons l'exemple du Poème Symphonique pour illustrer ces phénomènes. J'ai estimé la durée d'un clic de métronome à environ 10 ms. Lorsque les 100 métronomes sont activés au début de la pièce, et si nous évaluons un tempo moyen de 120 battements par minutes, alors

---

<sup>20</sup> citation originale : "To create what we hear as a solid cloud, a good rule of thumb is to set the density per second of the cloud to at least  $2/D$ ", ibid, p.106

cela nous donne une densité moyenne de  $120 / 60 * 100 = 200$  grains par seconde. Avec ces valeurs, le coefficient de consistance sera alors de  $200 * 0.01 = 2.0$ . Nous sommes donc dès le début de la pièce à la limite d'une perception consistante du nuage. Cet ordre de grandeur nous permet de nous rendre compte que la sensation de la continuité du nuage de tic-tac est en réalité déjà compromise au commencement de l'oeuvre, alors que Ligeti pensait que le chiffre symbolique de 100 métronomes lui aurait largement suffi à atteindre cet effet<sup>21</sup>.

Bien sûr, ce coefficient de remplissage est un indicateur limité, car il ne prend en compte que le facteur de fission temporelle. Dans le cas de la pièce de Ligeti, il aurait également fallu prendre en compte le point d'écoute afin d'évaluer le facteur de fission spatiale, ainsi que le rapport champ direct/champ diffus avec la réverbération de la pièce, faisant fluctuer notre sensation de fusion ou de fission en fonction de notre place dans le public (dans ce cas de figure, la sensation de fission augmente à mesure que l'on s'approche de l'orchestre percussif).

### Le timbre de la texture

Nous n'avons étudié jusqu'ici que des textures sonores dans lesquelles les grains étaient similaires et étaient répartis dans le temps de manière aléatoire. Cependant, la qualité timbrale du nuage dépend largement du timbre des grains eux-mêmes (caractérisation que nous avons appelé « essentialiste » dans la Partie 1 ), mais aussi et surtout de la manière dont ils sont agencés entre eux (caractérisation "comportementale"). Comme le souligne Wallace T. Berry dans *Structural Functions in Music* (1976), *"on pourrait dire que la nature des interactions et des interrelations au sein du tissu musical, en dehors de la densité calculable, constitue l'aspect qualitatif, par opposition à l'aspect purement quantitatif, de la texture. [...] Les changements dans l'indépendance et l'interdépendance relatives entre les composantes d'une texture musicale donnée constituent certains des facteurs les plus décisifs (et les plus subtils) dans la morphologie expressive de la structure."*<sup>22</sup> (Berry 1976, p.184-85).

Afin de parvenir à articuler les morphologies des textures sonores, il conviendra donc de trouver les outils permettant le contrôle et la manipulation à la fois du timbre des grains ainsi que de la manière dont ils sont agencés dynamiquement dans le temps.

---

<sup>21</sup> Rappelons nous de sa citation originale : *"So I thought, I'll produce a rhythmic grid that is so thick at the beginning that it seems practically continuous: that would signify blurring and disorder. To do this I needed a sufficiently large number of metronomes - 100 was just an estimate."* (Ligeti 1997, p.12)

<sup>22</sup> citation originale : *"The nature of interactions and interrelations within the musical fabric, apart from calculable density, might be said to constitute the qualitative, as distinct from the purely quantitative, aspect of texture. [...] Changes in relative independence and interdependence among concurrent components in a given musical texture will be seen to constitute some of the most decisive (and subtle) factors in the expressive shaping of structure."*

## Théorie de l'émergence sonologique

C'est dans la lignée de cette réflexion que le compositeur et théoricien Agostino Di Scipio pose le constat d'une fracture de la dichotomie entre le niveau de la manipulation des son (niveau macroscopique, qui relève de la composition ou du montage des objets sonores) et le niveau de la composition du son lui-même (niveau microscopique, à l'intérieur des sonorités, qui relève du *sound design*) (Di Scipio 1994b). Ces deux échelles toujours restées disjointes mais complémentaires se fondent l'une dans l'autre grâce à la manipulation du microtemps avec les outils numériques : il s'agit de *composer le son avec d'autres sons*. La manière dont sont organisés et choisis les micro-sons à une échelle bas-niveau conditionnerait donc l'émergence d'une sonorité à un niveau supérieur. Comme le souligne Di Scipio, *"la question qui se pose au compositeur-trice (et à la théoricien-ne de la musique) est la suivante : comment (ou quel type de) relations microtemporelles entre les quanta acoustiques peuvent-elles donner naissance à une macrostructure, un objet ou un processus audible présentant certaines caractéristiques [sous-entendu nouvelles, d'une autre nature]"*<sup>23</sup>. La tâche qui incombe à la compositeur-trice sera donc de chercher des manières de systématiser l'organisation de ces sons par l'implémentation d'algorithmes. Chacune d'entre elles constitue alors ce que Di Scipio désigne comme **une théorie de l'émergence sonologique** (Di Scipio 1994a).

Il s'agira alors d'explorer les différentes techniques existantes de synthèse de textures sonores, avant de pouvoir exposer ma proposition d'implémentation de l'algorithme de nuées.

## Synthèse de textures sonores

La synthèse de textures sonores est très souvent associée à la synthèse de sonorités issues de phénomènes naturels (pluie, feu, vent, etc.), et elle trouve donc de nombreuses applications dans des domaines aussi variés que le cinéma, les jeux vidéo, la création multimédia ou encore les arts sonores. Il existe de nombreuses techniques de synthèse de textures au sein desquelles nous pouvons dissocier deux grandes familles : d'une part, la synthèse par modèles physiques, qui synthétise le phénomène sonore dans son entièreté en étudiant les modèles d'émergence bas-niveau de sons du monde réel, souvent par des modèles excitation-résonance, et d'autre part, la synthèse par granulation, qui agence des grains pré-existants dans le temps<sup>24</sup>. Si la prolifération de ces techniques témoigne du caractère non-uniforme des textures sonores<sup>25</sup>, elles restent de manière générale très limitées dans leur possibilité d'articulation expressive.

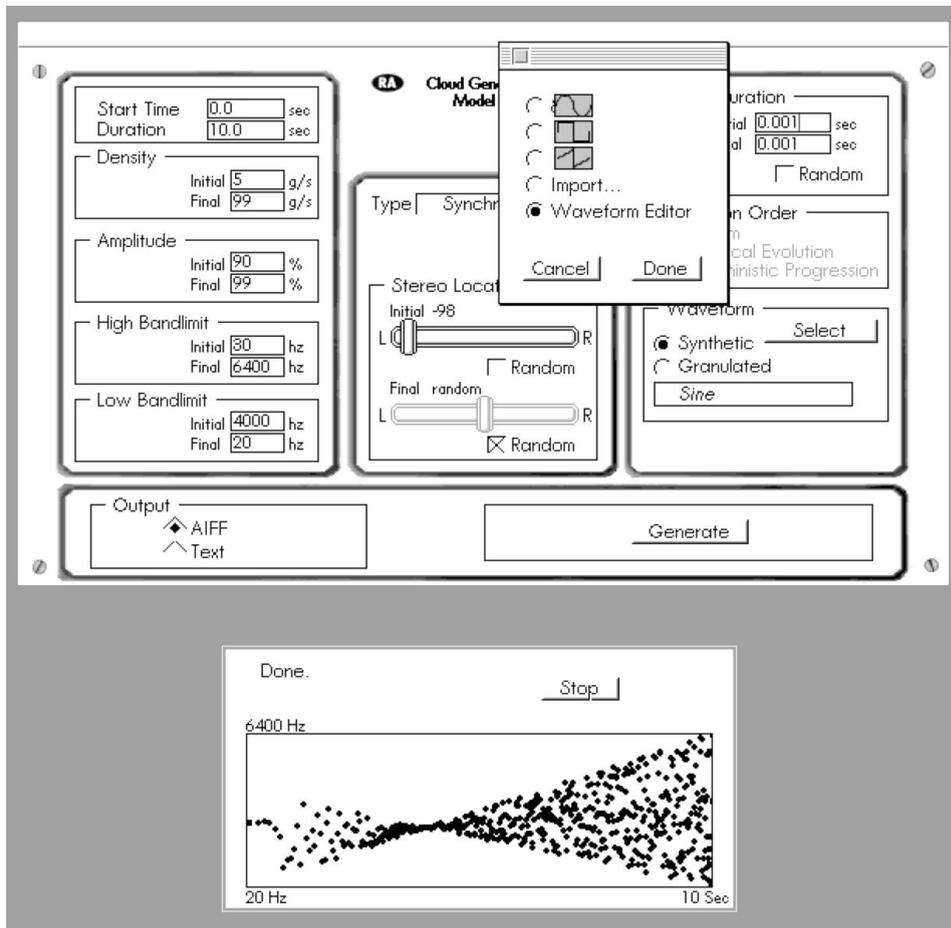
---

<sup>23</sup> citation originale : *The question for the composer (and music theorist) becomes: how (or what kind of) microtemporal relationships among acoustic quanta may bring forth a macrostructure, an audible object or process of certain characteristics. In a sense, any answer to this question would require from the composer an ability to implement his/her own theory of sonological emergence.* » (Agostino Di Scipio, 1994)

<sup>24</sup> Un état de l'art complet des différentes techniques de synthèse de texture est consultable dans State of the Art Texture Synthesis par Diemo Schwarz (2011).

<sup>25</sup> Dans le cas des techniques de modélisation physique, chaque typologie de texture demanderait une implémentation spécifique dédiée (Schwarz 2011, p.7).

Par exemple, il n'existe que très peu d'outils permettant la génération de textures par manipulation de nuage de grains. On retrouve seulement le Cloud Generator développé par Curtis Roads en 1995 au CEMaMu aux côtés de Iannis Xenakis et de ses équipes. Cet instrument reste néanmoins très limité dans les types de sonorités générées, car la répartition des grains dans le temps se fait de manière soit aléatoire, soit synchrone, suivant une certaine densité générale.



*Interface du Cloud Generator développé par Curtis Roads*

Finalement, si le modèle granulaire semble être le meilleur moyen d'atteindre des sonorités aussi riches que les textures sonores de systèmes naturels<sup>26</sup>, il n'existe à ce jour que très peu de programmes permettant une articulation fine de morphologies de nuages de grains.

L'articulation de masses de micro-sons appelle nécessairement un système de manipulation particulier, car il s'agit de manipuler un ensemble d'entités en même temps par la formulation d'intentions globales. Xenakis formule, à la suite de son introduction à la pensée granulaire, une autre théorie qui relève cette fois-ci de la composition de ces grains : la

<sup>26</sup> Agostino Di Scipio explique que les stratégies de design sonore par micro-composition parviennent à se rapprocher de la richesse et de la complexité de sonorités issues de phénomènes naturels : "The sonic complexity of these sounds arises from continually changing relations of phase in the signal, something that occurs at a local scale of time (transient phenomena) but that profoundly affects the global perceptual properties of the musical structure: le grain de la matière sonore, in Pierre Schaeffer's words, a most relevant aspect of timbre located somewhere between the phase-level in the signal (audio-time) and the scale of the amplitude envelope shape (control-time)." (1994a, p.139)

composition par lois stochastiques. Il s'agira alors dans la suite de ce développement, de questionner la pertinence de la transposition de ces modèles directement inspirés de phénomènes naturels dans le champ du sonore.

## II. Manipulation de masses sonores par modélisation de systèmes inspirés de phénomènes naturels : de la stochastique aux systèmes auto-organisés

*« Regardons les volutes d'une fumée de cigarette dans une pièce calme. Nous savons que la fumée est composée de molécules et leurs mouvements désordonnés suivent statistiquement des règles qui nous sont sensibilisées comme des volumes ou des surfaces mouvantes terriblement mobiles et riches. [...] Identifions les sons ponctuels, par exemple : pizz. [pizzicati], aux molécules ; nous obtenons une transformation homomorphe du domaine physique au domaine sonore. Le mouvement individuel des sons ne compte plus. L'effet massal et son évolution prennent tout un sens nouveau, le seul valable. » (Xenakis, 1958, « Les trois paraboles », in I. Xenakis, 1971 : 16-19)*

Dans l'introduction de la Partie 2, j'ai évoqué l'influence des systèmes naturels dans l'imaginaire des compositeurs, de l'inspiration poétique des sonorités qui émanent de celles-ci. Mais lorsque Xenakis introduit son paradigme granulaire pour établir de nouveaux modes de composition, il n'avait pas l'intention de se servir de ces théories pour les utiliser de manière traditionnelle, en inscrivant des notes sur sa partition selon son inspiration.

Xenakis était davantage intéressé par la simulation de leurs lois d'organisation, d'en imiter les modes d'émergence. *“Ce qui l'intéresse, ce n'est pas la reproduction d'un phénomène naturel en tant que tel, mais le transfert vers la musique des lois qui le constituent”* (Solomos, 2019).

Ainsi, il intégrera dans ses compositions les principes mathématiques de la suite de Fibonacci, des mouvements browniens, des automates cellulaires, etc. Mais c'est avec l'appui de statistiques et des probabilités, ou plus exactement la *stochastique*, qu'il développera la majorité de ses pièces, car elles constituent selon lui un modèle universel emprunté à la nature :

*“Mais d'autres voies conduisent aussi au même carrefour stochastique. Tout d'abord des événements naturels tels que les chocs de la grêle ou de la pluie sur des surfaces dures ou encore le chant des cigales dans un champ en plein été. Ces événements sonores globaux sont faits de milliers de sons isolés, dont la multitude crée un événement sonore nouveau sur un plan d'ensemble. Or cet événement d'ensemble est articulé et forme une plastique temporelle qui suit, elle aussi, des lois aléatoires, stochastiques. Si donc on veut modeler un grand amas de notes ponctuelles telles que des pizzicati de cordes, il faut connaître ces lois mathématiques, qui ne sont d'ailleurs ni plus ni moins qu'une expression dense et serrée d'une chaîne de raisonnements logiques.” (Xenakis 1963, p.19)*

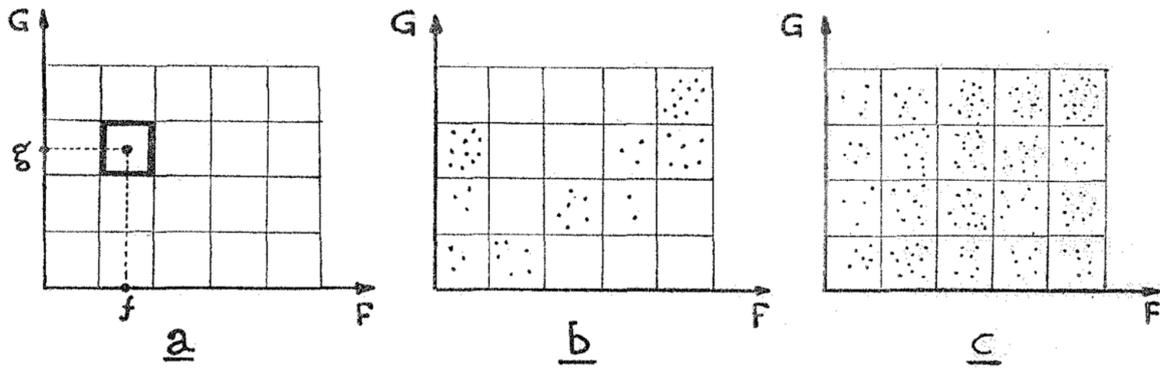
Il théorise et expérimente ces hypothèses stochastiques dans *Musiques Formelles* (1963), c'est-à-dire le calcul de probabilités par des procédés aléatoires et statistiques, conjointement avec ses hypothèses granulaires.

C'est dans le sillage de cette pensée, qui mobilise des procédés d'écriture systémiques, que je proposerais dans une seconde partie l'étude d'un autre modèle inspiré de phénomènes naturels : un algorithme de modélisation de nuées d'oiseaux.

#### A. Stochastique, entre ordre et chaos

Qu'est-ce que permet la stochastique que ne permettent pas les moyens traditionnels de composition ? Nous devons pour répondre à cette question considérer la nature des sons qui sont à manipuler. En appliquant des transformations homomorphes entre le domaine des gaz à celui des sons (en l'occurrence des pizzicati de cordes), Xenakis doit former des nuages de grains à partir d'une myriade de particules sonores. Afin de ne pas avoir à déduire lui-même les valeurs de fréquence et d'intensité de chaque particule, il systématise par des procédés mathématiques une génération de valeurs pseudo-aléatoires qu'il corrèle ensuite aux attributs sonores.

Voici par exemple 3 tableaux illustrant ce procédé de composition. On peut voir dans le tableau A la génération d'un grain unique, caractérisé par sa fréquence (en abscisse) et son intensité (en ordonnée). On constate une augmentation de la densité des grains dans les tableaux B puis C, aboutissant pour ce dernier à une sonorité ressemblant à *“un bruit blanc à faible densité qui peut donc être perçu comme un chatolement sonore occupant tout l'espace audible”* (ibid p.70).



Différentes densité de grains projetées dans un espace temps-intensité (f, G) (tiré de Xenakis 1963)

La construction de la composition s'effectue par la succession de ces différents tableaux, dont la distribution des grains s'effectue selon le degré d'ordre ou de désordre que le compositeur aura décidé.<sup>27</sup> Cette évolution entropique de trames sera utilisée dans Analogique A + B, une composition qui déploie ces nuages de grains par alternance entre un effectif de cordes et des sons électroacoustiques (grains élémentaires sinusoïdaux).

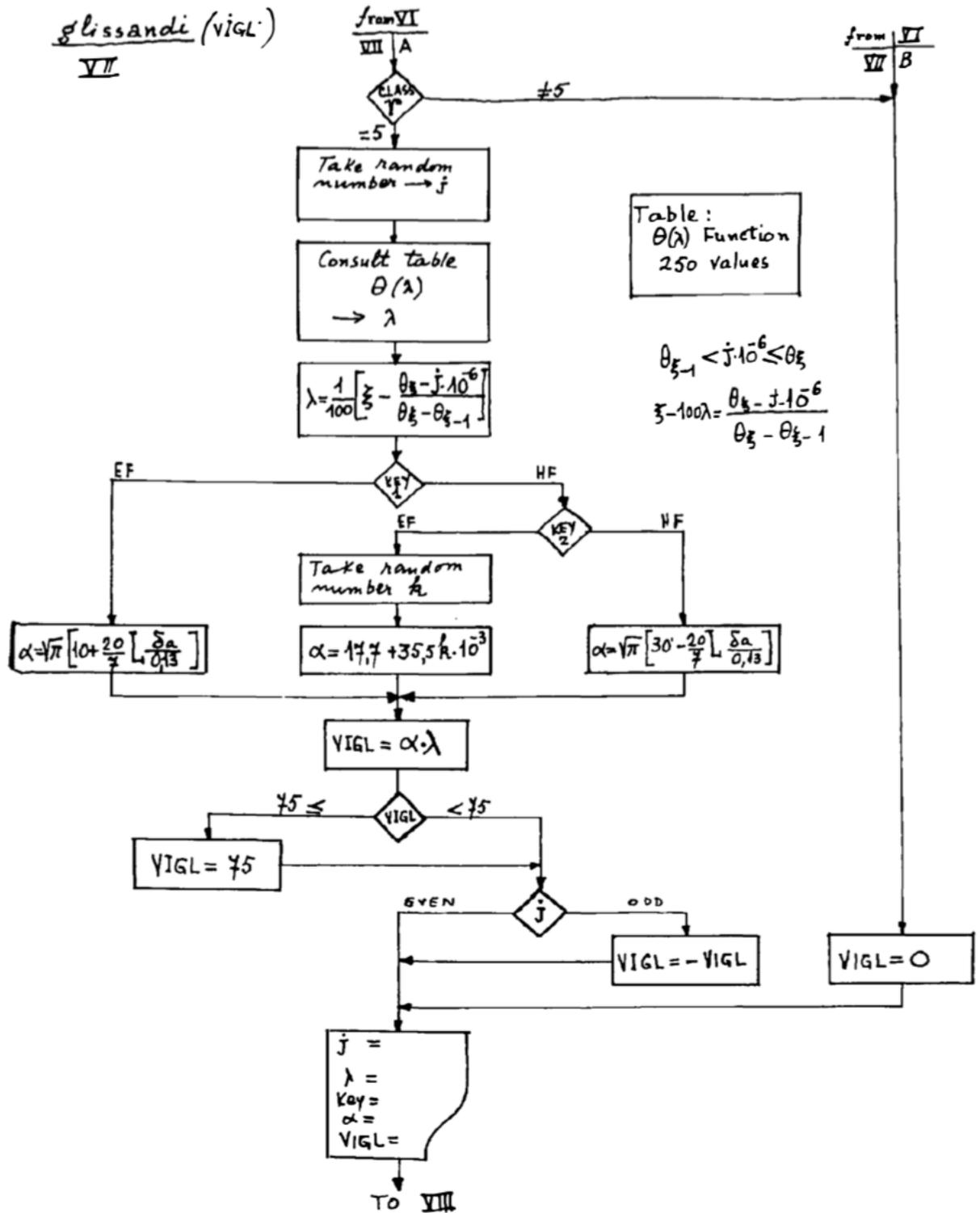
Ces constructions selon des jeux d'ordre et de chaos, telles une « esthétique dionysiaque » comme l'appelle Makis Solomos (2019), au caractère déchainé et imprévisible, sont à l'image des expériences qu'il a vécu lors de violentes manifestations en Grèce avant son arrivée en France :

*« Tout le monde a observé les phénomènes sonores d'une grande foule politisée de dizaines ou de centaines de milliers de personnes. Le fleuve humain scande un mot d'ordre en rythme unanime. Puis un autre mot d'ordre est lancé en tête de la manifestation et se propage à la queue en remplaçant le premier. Une onde de transition part ainsi de la tête à la queue. La clameur emplit la ville, la force inhibitrice de la voix et du rythme est culminante. C'est un événement hautement puissant et beau dans sa férocité. Puis le choc des manifestants et de l'ennemi se produit. Le rythme parfait du dernier mot d'ordre se rompt en un amas énorme de cris chaotiques qui, lui aussi, se propage à la queue. Imaginons de plus des crépitements de dizaines de mitrailleuses et les sifflements des balles qui ajoutent leur ponctuation à ce désordre total. Puis, rapidement, la foule est dispersée et, à l'enfer sonore et visuel, succède un calme détonant, plein de désespoir, de mort et de poussière. Les lois statistiques de ces événements vidés de leur contenu politique ou moral sont celles des cigales ou de la pluie. Ce sont des lois du passage de l'ordre parfait au désordre total d'une manière continue ou explosive. Ce sont des lois stochastiques » (Xenakis 1963, p19)*

Ainsi, grâce aux lois stochastiques, Xenakis a la possibilité de régler l'évolution d'un état ordonné à un état désordonné. Il a également la possibilité de contrôler le passage du nuage d'un état discontinu à un état continu, par le biais de la densité. Il parvient ainsi à modéliser des masses de sons comme une foule par des formulations d'intentions à grande échelle

<sup>27</sup> Pour une explication plus détaillée des calculs mobilisés et des choix du compositeur, consulter la sous-partie "Etude de l'ataxie au niveau des trames (nuage de grains)" p.84 de *Musiques Formelles* (1963). Un résumé de l'ensemble du procédé compositionnel est résumé dans "Stochastics and Granular Sound in Xenakis' Electroacoustic Music" écrit par Agostino Di Scipio (2012).

telles que « désordre général » ou encore « ordre partiel ». Ces traductions d'intentions de haut-niveau aux milliers de composantes bas-niveau (chaque composante étant indépendante l'une de l'autre) n'est possible que par l'utilisation de procédés systémiques. Xenakis adopte donc ici une démarche cybernétique : il s'agit de faire l'analyse d'un système complexe en identifiant ses mécanismes, en décomposant les relations fonctionnelles, et en étudiant leurs interactions, en vue d'en extraire des modèles. Une fois ces modèles algorithmiques établis, s'offre à lui le choix de la traduction des données générées au domaine musical. Cet algorithme de composition constitue donc, avant même l'arrivée de la programmation sur ordinateur, une théorie de l'émergence sonologique.



Extrait du diagramme correspondant à l'algorithme d'Achorripsis (tiré de Xenakis 1963)

## Limites de la stochastique dans la formation de sonorités émergentes

Cependant, même si les lois stochastiques permettent à Xenakis de manipuler dans une certaine mesure l'entière du système, elles comportent certaines limites dans la formation de sonorités de deuxième ordre.

Pour beaucoup d'auditeurs, l'écoute d'Analogique A+B est en réalité assez décevante vis-à-vis de toutes les implications théoriques en jeu (Di Scipio 1997). Cela peut s'expliquer par plusieurs facteurs. Comme il le souligne lui-même, les limites de jeux de l'orchestre à corde ne permettent pas se détacher de la nature timbrale des pizz pour se laisser porter vers d'autres sonorités, et les nuages de grains sinusoïdaux générés de manière électroacoustique ont été synthétisés dans un contexte technologique assez peu avancé, qui ne permettait pas en l'occurrence de manipuler numériquement des échantillons appartenant au micro-temps<sup>28</sup>.

D'autre part, la distribution stochastique des sons générés ne peut être perçue autrement que comme aléatoire, puisque le réseau de relations entre les composantes du système ne forme pas une image cohérente, discernable dans le cadre d'une syntaxe musicale. Comme le pointe Di Scipio, *"la limite des mécanismes de Xenakis réside dans le fait qu'ils permettent difficilement de faire émerger un flux évolutif au sein de la matière sonore et de déterminer l'émergence de sonorités de second ordre."*<sup>29</sup>

De plus, contrairement aux systèmes naturels, la musique stochastique est un système fermé, insensible aux événements extérieurs. Elle ne dépend que des conditions initiales, et tend seulement vers deux états antagoniques : l'ordre ou le chaos.

Di Scipio affirme que la spécificité des propriétés émergentes réside dans le fait qu'elles *"semblent exiger davantage un ordre chaotique vivant qu'un désordre statiquement organisé"*<sup>30</sup>.

Autrement dit, on pourrait alors se demander quel type de système permettrait à la fois d'organiser une masse dans sa globalité, selon des intentions générales de haut-niveau (que l'on pourrait relier à la sensation d'ordre et d'homogénéité), mais qui laisserait également la place à l'émergence de complexités locales, au sein desquelles il serait possible d'entendre

---

<sup>28</sup> "La réalisation « Analogique A » remplace les grains élémentaires sinusoïdaux par des nuages très ordonnés de grains élémentaires qui restituent le timbre des cordes. Une réalisation avec des instruments classiques ne saurait en aucun cas, en raison des limites humaines de jeux, donner des trames dont le timbre puisse être de nature différente de celui des cordes. L'hypothèse donc d'une sonorité de deuxième ordre ne pourrait dans ces conditions se trouver ni confirmée ni infirmée. Par contre une réalisation avec des moyens électromagnétiques puissants tels que des cerveaux électroniques, machines analogiques, etc., devrait permettre la démonstration d'une sonorité de premier ordre, à base de grains élémentaires sinusoïdaux, ou du type de Gabor". (Xenakis 1963, p.122)

<sup>29</sup> citation originale : "In short: the limit of Xenakis' mechanisms lies in the fact that they hardly bring forth an evolutive flow in sound matter and determine the emergence of 2nd-order sonorities." Di Scipio 1997, p.176

<sup>30</sup> citation originale : "In all systemic situations, emergent properties seem to require more a lively chaotic order than a statically organized disorder [...]." (Di Scipio 2001, p.83)

les interactions entre les unités élémentaires (que l'on pourrait assimiler à un grouillement interne chaotique, donnant l'impression d'une multiplicité).

Cette description peut s'apparenter à celle d'un écosystème. Il sera donc question d'étudier la pertinence de la transposition d'un modèle écosystémique pour l'articulation de masses sonores, en partant de l'hypothèse que les comportements issus de la modélisation seront à l'image des sonorités générées, autrement dit, que *"l'écologie du mécanisme de génération sonore reflète l'écologie du phénomène sonore"*<sup>31</sup> (Di Scipio 2012, p.7)<sup>32</sup>.

Sur ce principe, j'avais émis l'hypothèse dans l'introduction de ce mémoire d'un rapprochement possible entre le phénomène des nuées d'étourneaux et ces nuages de micro-sons. Je propose donc à présent l'étude d'un algorithme de modélisation de nuées, sur lequel tout le reste de mes recherches s'appuiera: l'algorithme *boids*.

## B. Systèmes auto-organisés comme orchestrateurs des nuages de grains : le cas de l'algorithme *boids*

### Boid et vie artificielle

En 1986, l'informaticien Craig Reynolds implémente le premier algorithme de modélisation de nuées qui marquera durablement le champ de la recherche en vie artificielle. Utilisé surtout comme une technique d'animation graphique, cet algorithme appelé *boids* (contraction de *bird* et du suffixe *-oid*) a notamment contribué à modéliser des vols collectifs d'oiseaux, des bancs de poissons ou encore des effets de foule au cinéma comme dans les jeux vidéo. A cette époque, des progrès remarquables sont observés dans le champ de la vie artificielle, science alliant à la fois informatique et biologie pour la modélisation de systèmes artificiels s'inspirant de systèmes vivants. On peut citer des modélisations majeures telles que les automates cellulaires, le jeu de la vie de John Conway ou encore la fourmi numérique de Christopher Langton.

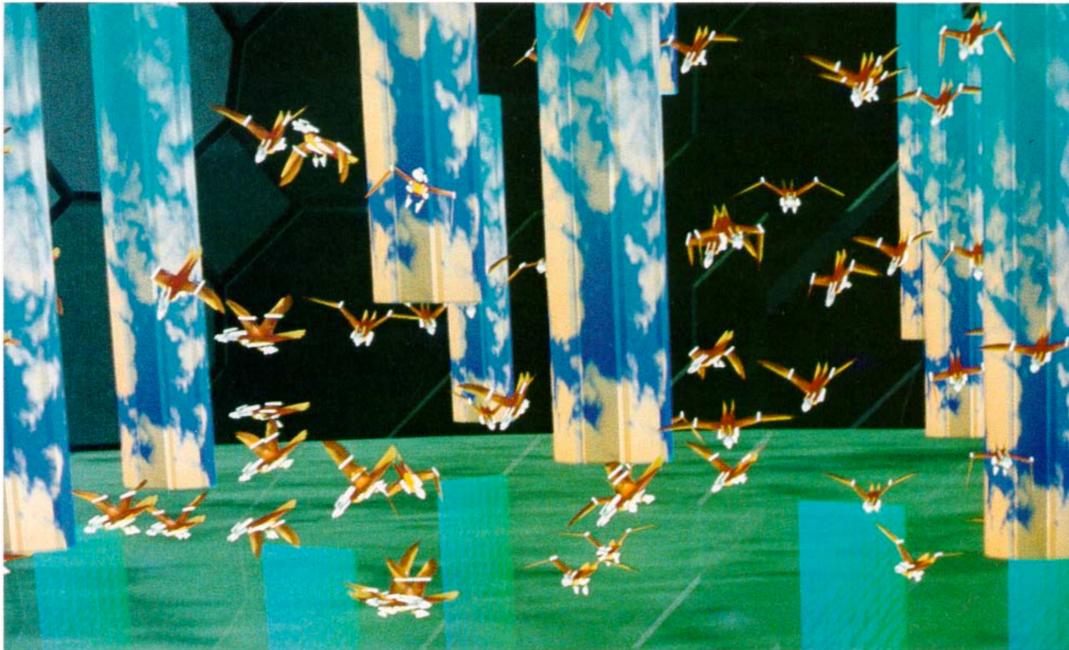
L'algorithme *boids* est encore largement exploité aujourd'hui car il repose sur des principes de programmation très simples et que les dynamiques comportementales qui en émergent restent très convaincantes. Mon intérêt s'est porté sur l'étude de cet algorithme car il a

---

<sup>31</sup> citation originale : *"It becomes necessary for the user to carefully consider the perceptual correlates of the interconnection among the component elements in the sound-generating mechanism. The user is led to focus on structural and audible relationships among components in the sound phenomenon, i.e., on the variety of causal forces of and in the sound event – that is: on the context of (and in) the sound as perceived. In that sense, the approach exhibits aspects of ecological validity. [...] In a sense, the ecology of the sound generating mechanism reflects the ecology of the sound phenomenon."*

<sup>32</sup> Face à ce constat, et dans la ligne de la pensée systémique de Xenakis, Di Scipio propose l'implémentation d'un système ouvert, "parlant" et "écoutant" : le projet Écosystème Audible (Di Scipio 2020). Au sein de ce projet, tous les sons synthétisés et projetés dans l'espace proviennent de sons captés durant la performance. Utilisant des procédés de feedback (tels que les larsen), cette installation est caractérisée par un haut degré d'interconnexion et d'interdépendance entre les agents humains (performeurs) et "non-humains" (équipement et environnement). La singularité de cette performance à l'œuvre continuellement réside dans "son caractère "vivant", c'est-à-dire sur une forme d'*agentivité écosystémique* émergeant d'un dispositif performatif" (ibid). Le dispositif s'auto-régule par lui-même tel un système auto-organisé, captant et réinjectant les sons qu'il perçoit, devenant un véritable écosystème sonore.

depuis aussi été utilisé dans de nombreux projets associés à la recherche musicale, qu'ils soient d'ordre compositionnel, de synthèse sonore ou de spatialisation.



Capture d'écran du court-métrage "Breaking the Ice" utilisant l'algorithme boids pour modéliser à la fois un vol collectif d'oiseau et un banc de poisson<sup>33</sup>.

### Description de l'algorithme

Le principe de l'algorithme est simple : il s'agit de donner une certaine forme d'autonomie à un ensemble d'agents pour qu'ils puissent se mouvoir et s'organiser collectivement, sans l'intervention d'éléments extérieurs. On les appelle des systèmes *multi-agents auto-organisés*.

Chaque boid a un "champ de vision" limité dans lequel il est capable de percevoir ses voisins. Lorsque l'un d'entre eux rentre dans le rayon de perception d'un autre, ils adaptent leurs trajectoires en fonction de leur comportement respectifs. Ces mouvements autonomes apparaissent selon la formulation de trois lois :

- cohésion : les boids vont se rapprocher entre eux lorsqu'ils entrent respectivement dans leurs rayons de perception, permettant de maintenir une cohésion de groupe : il s'agit donc d'une force attractive.
- séparation : les boids vont s'éloigner entre eux s'ils sont trop proches afin de maintenir une certaine distanciation sociale et éviter ainsi les collisions : il s'agit donc d'une force répulsive.
- alignement : les boids vont s'aligner sur la vitesse et la direction de leurs voisins, créant des dynamiques de navigation de groupe

---

<sup>33</sup> Le film est consultable ici : <https://www.youtube.com/watch?v=pbFEQv259yw>

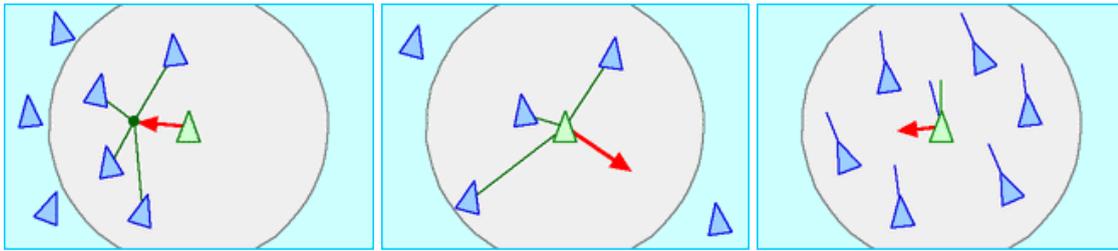


Illustration des principes de forces de l'algorithme boids : cohésion, séparation et alignement

Comment alors contrôler un tel système si ces agents sont autonomes ? Il conviendra d'implémenter certaines variantes, des forces extérieures au système pour venir contraindre leur contour. Car à ces trois lois peuvent s'ajouter d'autres fonctions afin d'augmenter la complexité du système, en fonction du contexte dans lequel on les utilise (par exemple des fonctions de fuite vis-à-vis d'un prédateur, ou d'attraction vers des points de ressource). Les différentes forces extérieures implémentées dans le cadre de ce projet de recherche seront détaillées dans la Partie 3.

La particularité de ces systèmes grégaires dans le monde réel réside dans le fait que leur complexité ne dépend pas de leur taille<sup>34</sup>. En effet, comme les étourneaux, ou les poissons, ont une perception limitée de leur environnement (puisqu'ils bloquent mutuellement leurs champs de vision par leur accumulation), ils interagissent seulement avec leurs plus proches voisins, sans avoir conscience de la taille globale de l'ensemble du groupe.

Ainsi, le système peut comporter seulement quelques dizaines d'agents comme il peut en comporter des milliers. Cette propriété est cruciale dans la modélisation de ces systèmes grégaires, car ils permettent l'apparition de dynamiques locales. Celles-ci se manifestent perceptivement par l'émergence d'une sensation de multiplicité, grâce à aux légers mouvements individualisés vis-à-vis de l'ensemble homogène.

Au sein de l'interface boids, les agents maintiennent une certaine distance spatiale entre eux grâce aux lois de séparation et de cohésion, entraînant une densité de regroupement plus ou moins grande en fonction de leur coefficient. La loi d'alignement, comme son nom l'indique, permet aux boids de s'aligner entre eux, ou autrement dit de **réduire l'écart des valeurs de vitesse et de direction qui les sépare**.

En imitant ses voisins, chaque boid essaie grâce à cette loi de se fondre dans le groupe. Dès l'instant où il rejoint le groupe, le boid perd sa singularité puisque ses valeurs de vitesse et de direction seront semblables à celles de ses voisins. Ainsi, ce sont la différence (ou la distance) entre ces différentes valeurs (position, vitesse, direction, etc.) qui entraîneront des sensations de regroupement ou de dispersion.

En réalité, comme nous l'avons évoqué à l'issue de la Partie 1, tout est une question de *distance perceptive* lorsqu'il s'agit d'induire des sensations de fusion ou de fission. Elles peuvent prendre la forme de distances spatiales comme c'est le cas avec l'algorithme boids,

<sup>34</sup> Ce n'est cependant pas le cas dans la modélisation graphique, puisque l'algorithme doit évaluer chaque distance qui sépare les boids entre eux. Le complexité de l'algorithme sera donc de  $O(n^2)$ .

mais peuvent également se manifester par des distances perceptives sonores (niveau sonore, hauteur, spatialisation, timbre, etc.).

Ainsi, nous pourrions imaginer les boids navigant dans un espace autre que le domaine spatial, par lequel il est communément représenté. En fin de compte, l'algorithme boids peut s'apparenter à un algorithme de modélisation d'un système d'entités au comportement grégaire tout court, sans nécessairement mentionner le fait qu'elles doivent être représentées au sein d'un repère spatial à deux ou trois dimensions : les boids peuvent ainsi évoluer au sein de n'importe quel espace à N dimensions. Tout ce qui compte désormais, c'est l'évolution du comportement des boids en fonction des distances qu'ils entretiennent entre eux. Ainsi, il est en théorie possible de s'abstraire de l'utilisation des boids dans un domaine de représentation spatiale : celle-ci devient seulement un moyen pratique pour l'observation de leurs évolutions et interactions.

On appelle l'opération de traduction de ces évolutions d'un espace donné vers le domaine sonore **sonification** (Kramer 1993).

Il s'agira donc dans le cadre de ce projet de recherche, de trouver des stratégies de sonification des données qui peuvent être extraites de l'algorithme boids. En identifiant chaque boid à une entité sonore, j'émetts l'hypothèse qu'il sera possible de manipuler un système de masses sonores, de la même manière que l'on manipule le système boids graphique. Ainsi, par la traduction des effets des comportements des boids dans le domaine du son, **il serait ainsi possible de faire émerger des nuées sonores.**

De nombreux projets de recherche musicale ont été menés depuis, s'inspirant de ces modèles dits *multi-agents* afin d'explorer de nouveaux modes de composition<sup>35</sup>.

Comment alors se servir de l'algorithme boids pour articuler des matériaux sonores ?  
Comment traduire leurs comportements en son ?

Il s'agira dans la partie suivante d'étudier les projets exploitant cet algorithme à des fins de composition, afin d'examiner les stratégies de corrélation entre celui-ci et le modèle de synthèse sonore.

---

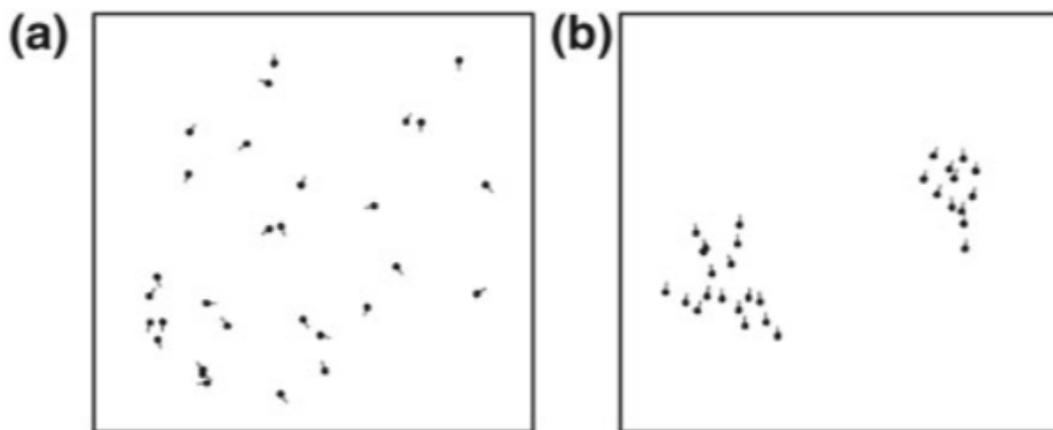
<sup>35</sup> On constate, surtout à partir des années 90, un engouement croissant pour ces systèmes évolutionnistes ou multi-agents dans le champ musical. Leurs modélisations deviennent accessibles par la démocratisation des ordinateurs personnels et donnent naissance à des styles musicaux tels que l'evolutionary computer music (Miranda & Biles 2007)

### C. Boids comme instrument musical

De nombreux projets ont émergé à partir des années 2000 dans l'exploitation de l'algorithme boids à des fins musicales, pour de la synthèse sonore, d'organisation temporelle d'objets ou de spatialisation (liste non exhaustive : Blackwell & Young 2004, Wilson 2008, Davis & Rebelo 2005, Huepe et al 2014, Eigenfeldt & Pasquier 2011, Arpegis et al 2018, Pearse 2016).

J'ai décidé de mener une courte étude de cas sur le projet Generating Music from Flocking Dynamics (Huepe et al 2014), car il illustre de manière claire l'importance des procédés de traduction entre un système visuel et comportemental et le système sonore. Huepe implémente trois types de corrélations entre l'algorithme boids à des paramètres de synthèse sonore. L'étude de ces différentes stratégies de traduction me permettront d'identifier lesquelles sont les plus à même de fonctionner pour mon propre algorithme, détaillé dans la Partie 3.

Pour les trois cas, Huepe fait passer le système boids d'un état chaotique (les trois lois d'auto-organisation ne sont pas actives) à un état ordonné (lois actives), et évalue la pertinence des stratégies de corrélation par la perception du passage d'un état à l'autre.

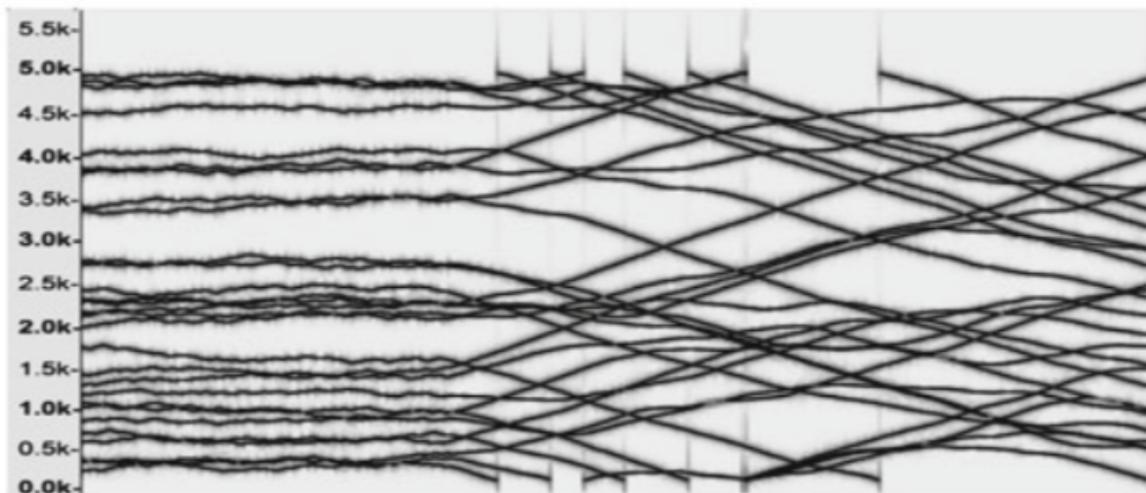


Passage d'un état stable (a) à un état instable (b) (tiré de Huepe et al. 2014)

L'approche directe :

Pour la première stratégie de corrélation, Huepe associe chaque boid à un oscillateur émettant un signal sinusoïdal. Il établit une translation directe entre les paramètres de position (x, y) de chaque boid et les attributs d'amplitude et de fréquence de chaque oscillateur. Lorsque l'on regarde les spectrogrammes, on arrive bien à imaginer le type de sonorités générées : il s'agit de grands clusters tels que nous l'avons vu avec les glissandi de Xenakis. Tous les changements de fréquences et d'amplitude témoignent donc de déplacements de boids, permettant d'identifier la dynamique individuelle de chaque agent (notamment par la fluctuation des fréquences entendues), mais il est néanmoins impossible de déceler les interactions qu'entretiennent les agents entre eux : nous n'avons que des traductions de valeurs absolues de déplacement.

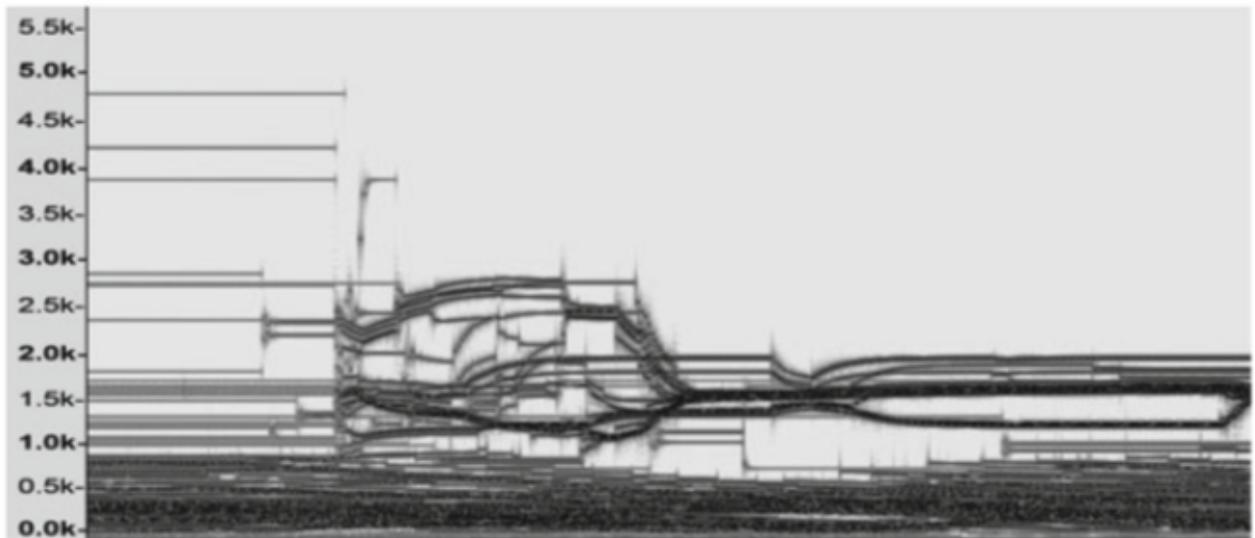
Ainsi, bien que les mouvements de groupe soient bien perceptibles par les effets de glissandi ascendant et descendant (par inférence dynamique), la sonorité générée reste un matériau pauvre, très harmonique et dissonant, difficilement exploitable dans le cadre d'une composition.



*Spectrogramme du passage de l'état stable à l'état instable pour l'approche directe  
(tiré de Huepe et al. 2014)*

### L'approche relative :

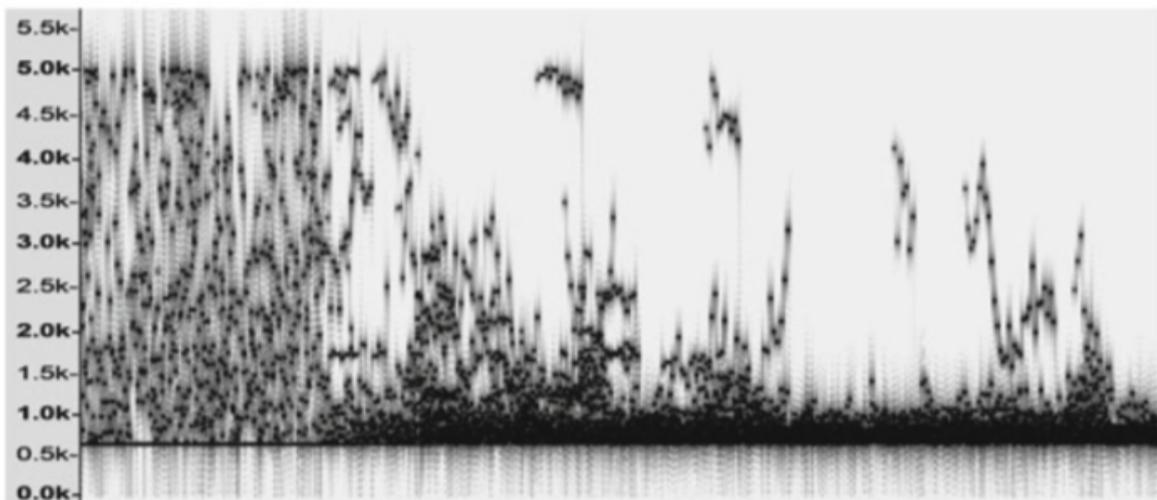
Dans la seconde stratégie, Huepe utilise le même moteur sonore mais, cette fois, il met l'accent sur les interactions qu'entretiennent les agents entre eux. Ainsi, chaque agent changera de fréquence lorsqu'il entrera en interaction avec un autre, et se stabilisera autour d'une fréquence voisine du boïd avec lequel il a interagi. Ainsi, dans la phase chaotique, la sonorité générée est une cacophonie générale, chaque fréquence fluctuant à chaque rencontre, tandis qu'on perçoit dans la phase ordonnée un regroupement stable autour d'un cluster. Un rapprochement dans l'espace résulte en un rapprochement spectral, une sensation de regroupement est identifiable par fusion spectrale, en plus de l'inférence dynamique (fluctuations de fréquences vers un état stable).



*Spectrogramme du passage de l'état stable à l'état instable pour l'approche relative*  
*(tiré de Huepe et al. 2014)*

### L'approche par modèles physiques :

Dans la troisième stratégie, il utilise un moteur sonore différent, générant des sonorités bruitées de type "friction" qui se rapprocherait selon lui de sonorités que l'on pourrait *"intuitivement associer à la réaction physique d'un groupe de particules mobiles en interaction."*<sup>36</sup> Chaque agent produit alors un son de "frottement" lorsqu'il rencontre un autre agent. A l'image de l'effet Doppler, la fréquence fondamentale du son émis sera dépendante de la vitesse relative des deux agents entrant en collision. Ainsi, deux agents qui se rencontrent via des directions opposées émettront un son aigu, tandis que deux agents naviguant dans le même sens produisent des sons graves. Dans la phase chaotique, nous pouvons constater un résultat sonore cacophonique, résultant d'une multitude de fréquences différentes. Lorsque le système passe progressivement à la phase ordonnée, on constate une polarisation autour d'un cluster de fréquences graves. On voit également à quelques endroits de la phase ordonnée des motifs "saillants" : ils correspondent à une dissociation d'un petit groupe d'agent vis-à-vis du gros groupe.



*Spectrogramme du passage de l'état stable à l'état instable pour l'approche par modèles physiques (tiré de Huepe et al. 2014)*

Quelles conclusions pouvons-nous tirer de ces analyses ?

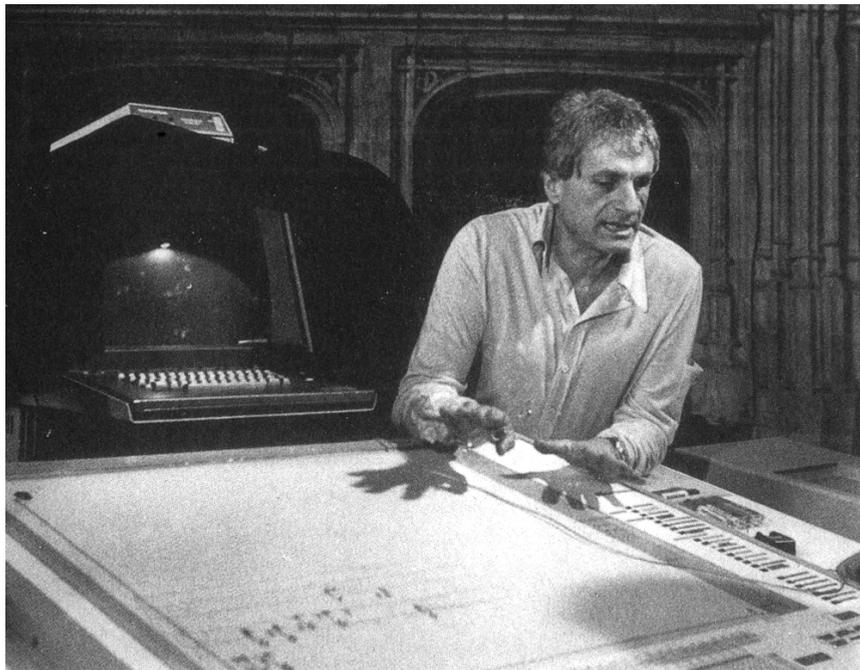
Dans un premier temps, nous pouvons distinguer la première tentative de sonification des deux autres par le fait que le processus de traduction s'opère de manière directe, de données visuelles aux données sonores. Dans ce cas, il s'agit de transposer de manière analogue, sans transformations, les valeurs absolues d'un domaine de représentation à un autre (position dans l'espace (x, y) aux valeurs spectro-énergétiques (f, A)). Cette technique est

---

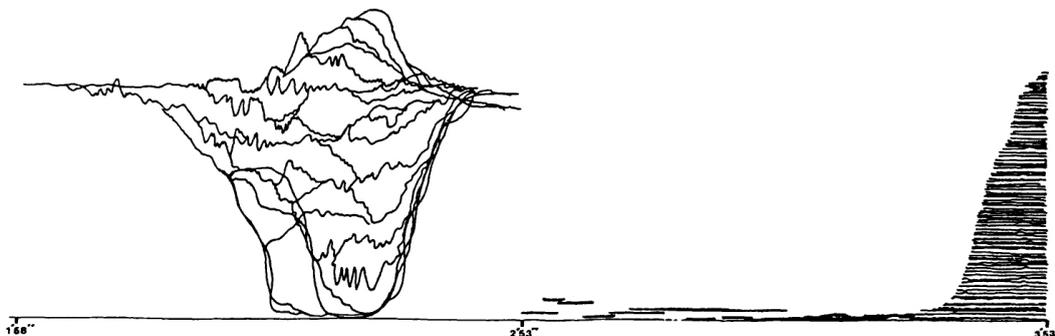
<sup>36</sup> citation originale : "In this approach, the aim is to implement an algorithm that captures elements of what we could intuitively associate to the physical response of a group of interacting moving particles." (p.167)

très peu utilisée car elle ne prend pas en considération les potentielles transpositions d'effets perceptifs d'un domaine à un autre. Elle peut néanmoins servir à "monitorer" dans le domaine du son une certaine évolution de données : le matériau musical généré devient alors une "illustration sonore" d'un phénomène, par exemple visuel. Les applications de cette sonification directe dans le domaine musical restent donc très limitées.

On retrouve néanmoins ces techniques de sonification directe à des fins compositionnelles avec la synthèse sonore graphique dont Xenakis a été le précurseur. Grâce au développement du système UPIC (Unité Polyagogique Informatique), créé au CEMaMu, il était alors possible d'écrire directement les trajectoires de fréquences d'oscillateurs par l'utilisation d'un stylet sur une tablette graphique.



Iannis Xenakis se tenant devant l'UPIC



*Extrait de la partition graphique de Mycènes Alpha - Xenakis (1978)*

Les lignes parallèles résultent en de grands clusters, qui ressemblent à bien des égards à ceux générés dans la tentative de sonification directe de l'algorithme boids. Concernant ce type de synthèse, il est important de préciser que la forme des dessins graphiques ne rend pas compte de la résultante sonore (à moins que l'on soit habitué à la lecture de spectrogrammes).

Dans les deux autres tentatives de corrélations, il s'agissait davantage de traduire les effets des interactions des boids, puisque la sonification se fait de manière relative et locale. En ce sens, ce modèle se rapproche plus de la manière dont nous percevons le phénomène visuel : nous percevons avant tout des variations de comportement lorsque les boids interagissent entre eux, ce sont donc ces interactions qu'il s'agira de rendre audible.

D'autre part, nous pouvons distinguer l'approche par modèles physiques des deux autres parce qu'elle adopte un modèle de synthèse de sons non entretenus, tels des grains, contrairement aux sons entretenus issus des oscillateurs sinusoïdaux. Ce dernier modèle traduit de manière intuitive la manière dont pourrait se manifester le phénomène sonore dans une situation réelle. Ces collisions sont comme de petites percussions : l'aspect sensible et tangible de cette sonification par modèle physique correspond à un modèle à suivre pour la suite de mes recherches. Le modèle granulaire s'inscrit donc dans la lignée de cette réflexion.

Finalement, nous pouvons constater que les différents procédés de sonification modulent très largement les types de matériaux sonores générés, ainsi que de la manière dont ils sont perçus. Il ne suffit donc pas d'avoir l'interface boids pour garantir une manipulation cohérente de nuages de grains, le cœur du travail se situe alors dans la corrélation des effets induits par leur comportement pour les traduire au son.

En somme, il s'agira de faire des corrélations métaphoriques entre le système source (boids) et le système cible (nuage de grains) : plutôt que des translations directes de données, il s'agira d'opérer des traductions de fonctions, ou d'effets perceptifs. John Holland énonce dans Emergence : from chaos to order (1999) trois postulats lorsqu'une corrélation métaphorique a lieu :

1. Il existe un système source contenant un ensemble de faits et de régularités.
2. Il existe un système cible avec des régularités et peut-être des faits qui sont difficiles à percevoir ou à interpréter.
3. Il existe une traduction de la source à la cible suggérée par un transfert d'inférences du système sources aux inférences du système cible.

C'est précisément ce transfert d'inférence qu'il s'agira de parfaire lors de la conception du programme du *synthétiseur des nuées*, afin de rendre tangibles les dynamiques complexes intrinsèques à l'algorithme boids.

## PARTIE 3. Le bruissement des nuées : implémentation du synthétiseur de nuées et monstration publique



*Flyers-affiches du projet réalisés par Eloïse Vo au sein du lieu d'exposition de la Nef*  
*En arrière plan se trouve le dispositif de diffusion sonore*

A l'issue de ces recherches théoriques, il s'agissait maintenant pour moi de confronter mes hypothèses à la pratique, à la fois par l'implémentation de l'algorithme du *synthétiseur de nuées*, mais aussi par la monstration publique du projet dans sa globalité, sous la forme d'une installation-concert : *Le bruissement des nuées*. Ces deux phases de travail seront décrites en deux temps, suivi d'un retour d'expérience.

## I. Implémentation de l'algorithme de l'orchestrateur de nuées

L'étude de cas du projet de recherche Generating Music by Flocking Dynamics (Huepe et al 2014) a permis de mettre en évidence le rôle crucial de la sonification entre l'algorithme boids et le moteur de synthèse sonore pour percevoir des « nuées sonores ». Dans ce projet, se sont des moteurs de synthèse par oscillateurs ou par modèles physiques qui ont été utilisés, permettant de modeler leurs paramètres (fréquence, intensité) en fonction d'attributs issus de l'algorithme boids (position, interaction).

Dans le cadre de ce mémoire, plutôt que de les synthétiser, je souhaitais pouvoir générer des sonorités à partir d'autres sons tirés d'un corpus, à la manière de mosaïques, comme j'ai eu l'occasion de l'exposer dans la Partie 2.I. il s'agit de construire des sonorités par l'agencement de micro-sons, sélectionnés dans le corpus en fonction de leurs attributs (hauteur, timbre, intensité). Comment cela peut-il se faire ?

Il existe une technique de synthèse sonore appelée synthèse concaténative par corpus qui permet de créer des sonorités par la concaténation d'autres sons en fonction de leurs attributs grâce à des descripteurs audio.

Je reviendrais dans un premier temps à l'algorithme boids, et à la manière dont il peut être articulé en vue d'un jeu instrumental, avant de présenter les principes de cette technique de synthèse par le prisme de l'analyse de l'instrument CataRT. Je m'attarderai ensuite sur l'analyse de descripteurs audio dans une troisième partie. Enfin, j'exposerai mes choix de sonification puis présenterai l'architecture générale de l'algorithme.

### A. Boids comme instrument musical expressif

#### Contrôle du système

L'algorithme boids fonctionne par l'utilisation de 3 lois fondamentales comme nous l'avons vu précédemment : séparation, cohésion et alignement. Ces trois forces, intrinsèques au comportement de chacun des boids, peuvent être modifiées par la manipulation de leurs coefficients respectifs. Ainsi il sera possible de changer la dynamique d'un système en altérant leurs « volonté » : en augmentant le coefficient de séparation par exemple, chacun des boids aura « envie » d'adopter une distanciation moyenne plus grande, résultant en un effet d'écartement.

Cependant, le maniement de ces coefficient seuls limitent les possibilités expressives du système, et il est apparu nécessaire d'implémenter un certain nombre de forces extérieures afin de garantir une plus grande maniabilité de celui-ci.

J'ai alors décidé d'implémenter quatre forces supplémentaires :

- Friction, me permettant de contenir les vitesses du groupe dans son ensemble, semblable à un effet de ralenti comme si les boids étaient plongés dans une matière visqueuse;
- Centre de gravité, me permettant d'attirer les boids vers un point donné, ici la position moyenne de l'ensemble des boids;
- Vortex, me permettant de créer un mouvement d'ensemble circulaire autour de la position moyenne de l'ensemble des boids;
- Mouvement aléatoire, me permettant de générer un comportement chaotique.

C'est uniquement en venant jouer sur les coefficients de ces forces que l'on va parvenir à manipuler le système dans son entièreté.

On remarquera qu'avec ces nouvelles forces, le comportement des boids imitant des oiseaux n'est plus reconnaissable. Il ne s'agissait pas de toutes façons d'implémenter un algorithme modélisant précisément un vol collectif d'étourneaux ou un banc de poisson, mais de se servir de leur particularité grégaire pour ensuite transformer le système en un instrument expressif.

Afin de les manipuler dans l'optique d'une performance instrumentale, j'ai utilisé à la fois l'interface expressive Touché pour contrôler les quatre forces extérieures, et un simple contrôleur avec potentiomètres linéaires pour contrôler les trois forces intrinsèques.

### Données extraites du système

Afin de pouvoir faire les corrélations, il faut extraire des données significatives relatives aux comportements des boids.

Nous avons vu avec l'étude de Huepe (2014) que les données statiques des positions absolues n'étaient pas réellement exploitables, et qu'il fallait davantage aller chercher les dynamiques locales et relatives des boids.

Les mouvements de chaque agent sont décrits par leur vecteur vitesse. C'est donc uniquement à partir des données de ce vecteur que nous allons travailler.

D'une part, la norme du vecteur vitesse, ainsi que sa variation dans le temps (autrement dit son accélération) seront des données importantes à exploiter.

D'autre part, les données relatives à l'angulation sont également très significatives. Par exemple, l'angulation relative de chacun des boids par rapport à son entourage traduit à quel point les agents sont alignés entre eux : s'ils sont côte à côte et vont dans la même direction, alors la différence d'angulation sera faible, s'ils vont dans une direction opposée (autrement dit, prêts à entrer en collision ou s'écartant les uns des autres dans un environnement proche) alors la différence d'angulation sera maximale.

Enfin, la densité de regroupement spatial, malgré son caractère statique, semble également être une donnée intéressante. Elle peut être extraite par le nombre de voisins que chaque boîd "voit" dans son champ de perception. Plus il "voit" de voisins, plus la densité de regroupement sera grande.

→ des schémas explicatifs de ces différentes données sont à ajouter en vue d'une prochaine version

## B. Synthèse concaténative par corpus

Avec l'apparition de bases de données sonores de plus en plus larges et facilement accessibles, ainsi que des descriptions toujours plus pertinentes de leurs contenus, la synthèse concaténative par corpus a été et continue d'être de plus en plus utilisée pour la composition, la synthèse d'instruments de haut-niveau et l'exploration interactive de corpus sonores (Schwarz 2007). Développée à l'IRCAM par l'équipe ISMM (Interaction Son Musique Mouvement) à partir des années 2000, elle reste néanmoins une technique qui peut s'avérer complexe à prendre main, rendant difficile son exploitation à grande échelle. Mais dans le cadre d'utilisations spécifiques, comme c'est le cas dans ce projet, elle est indispensable.

Cette technique est construite à partir du modèle de la synthèse granulaire, en l'augmentant par la possibilité de sélectionner des grains spécifiques en fonction de leurs caractéristiques sonores, grâce aux descripteurs audio. Ces derniers sont des outils d'analyse permettant d'assigner une valeur à une méta-donnée d'un signal audio désignant un caractère sonore spécifique (niveau sonore, hauteur moyenne, ratio bruité/harmonique, etc.).

Ainsi, les descripteurs permettent un classement de large bases de données sonores en fonction de différents attributs timbraux (McAdams 2013). Pour le descripteur « hauteur » par exemple, une valeur de méta-donnée faible sera attribuée à un signal grave, et vice-versa. Illustrons une utilisation possible de cette technique par l'exemple de la synthèse de la voix. Il est possible de synthétiser, dans une certaine mesure de vraisemblance, une voix parlée en concaténant une suite de phonèmes selon un ordre spécifique.

Dans un premier temps, il s'agira d'analyser un corpus de phonèmes (le système source) qui vont être utilisés comme unités sonores élémentaires, telles des tesselles de mosaïques. Ceux-ci vont être classés selon certains critères, en fonction des valeurs de métadonnées qui leur sont assignées par les descripteurs audio. Ils peuvent par exemple être classés selon leur « brillance » : un /s/ aura une valeur plus haute qu'un /o/.

Dans un second temps, il s'agira d'analyser le signal cible afin d'établir la suite de phonèmes correspondant à la phrase ou au mot à synthétiser.

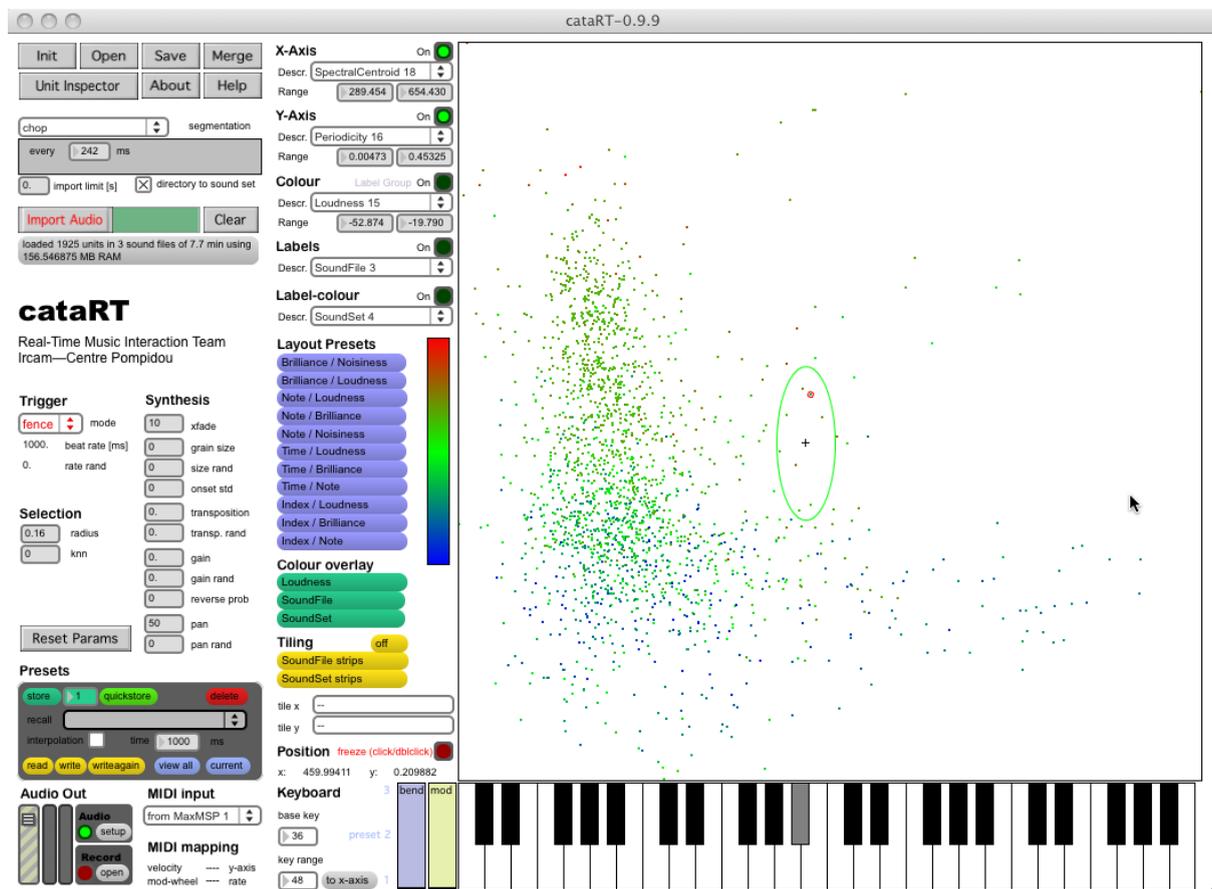
Il s'agira enfin de recréer le signal cible avec le matériau sonore du système source, en sélectionnant les phonèmes du corpus telles des briques pour les assembler en un mot. Cela est fait par la mise en correspondance des valeurs de métadonnées qui leur sont associées. Pour remplacer un phonème donné du système cible, l'algorithme va chercher

dans la base de données (système source) les échantillons dont les métadonnées se rapprochent le plus du phonème ciblé.

Il est ainsi possible de synthétiser une voix virtuelle masculine (système source : corpus de phonèmes masculins) par l'analyse d'une voix féminine (système cible : signal de voix féminine). Si le son synthétisé est obtenu par l'analyse d'un signal cible, comme c'est le cas dans cet exemple, alors on parle d'*audio mosaicing*.

De la même manière, il s'agira dans mon projet de corrélérer les données extraites de boids (système cible) à certains caractères audio à partir d'un corpus de son (système source). Cette traduction de données, ou sonification, sera étudiée dans les parties suivantes.

Il existe un instrument particulièrement intéressant qui permet la synthèse concaténative en temps-réel, par la navigation au sein d'un espace de nuage de grains : il s'agit de cataRT, développé à l'IRCAM par Diemo Schwarz.



Interface de l'instrument virtuel cataRT, module standalone

L'interface est composée d'un espace dans lequel sont projetés un ensemble de grains provenant d'un corpus donné, et classés selon des valeurs de méta-données établies par l'analyse de descripteurs audio. L'utilisateur-trice va pouvoir déclencher ces grains par la navigation du curseur sur ces derniers. Ces unités micro-sonores pourront ainsi être *concaténées* ensemble au sein d'un flux continu.

Sur la capture d'écran, on peut voir que les grains sont classés sur l'axe des abscisse (x) en fonction de la valeur associée à leur barycentre spectral (*Spectral Centroid*), et sur l'axe des ordonnées (y) en fonction de la valeur associée à leur caractère bruité (*Periodicity*, ou *Noisiness*), mais également par leur couleur en fonction de la valeur associée à leur niveau sonore (*Loudness*). Ainsi, un grain au caractère aigu, bruité et fort sera projeté dans le coin en bas à droite de l'interface, et aura une couleur rouge.

L'utilisateur·trice a le choix de sélectionner les descripteurs qu'il voudra faire figurer en abscisse, ordonnée, ou couleur au sein de l'interface.

Le corpus est segmenté en petits grains en fonction d'une taille définie avant d'être analysé par un ensemble de descripteurs, et vont ensuite être projetés dans l'espace de représentation.

Cette interface est particulièrement intéressante car elle permet d'organiser un espace timbral. Elle agit à la fois comme un espace de représentation et un espace d'interaction. Elle permet une exploration des sons de manière interactive, en composant des trajectoires dans l'espace, et ainsi de créer des évolutions sonores inédites sans pour autant altérer la nature timbrale de la source originelle. Cette préservation est particulièrement importante pour garder la "signature" sonore du corpus sélectionné et pour obtenir des textures convaincantes. Cela permet la génération d'une variété infinie de textures, renouvelée par chaque changement de corpus.

Pour la création du *synthétiseur de nuées*, on pourrait par exemple dans un premier temps identifier le curseur à la position d'un boid. Il faudrait alors démultiplier autant de fois le moteur de synthèse sonore qu'il y a de boids, permettant des corrélations indépendantes.

Cependant, les possibilités de jeu restent tout de même très limitées par la navigation du curseur au sein de l'interface. La corrélation entre les mouvements du curseur et la génération de sons par concaténation de grains peut être assimilée à une sonification de données directes. De la même manière que l'exemple de l'approche directe de Huepe, au sein de laquelle la position des boids était corrélée directement à la fréquence et l'intensité de chaque oscillateur, les effets relatifs et dynamiques des mouvements du curseur ne peuvent pas être traduits en son.

Il s'agira alors d'imaginer une extraction de données de boids traduisant des effets dynamiques (vitesses, accélération, etc..) plutôt que des positions absolues. CataRT reste néanmoins un outil très efficace afin de « monitorer » les valeurs attribuées par les descripteurs audio dans un espace de représentation tridimensionnel (x, y, couleur).

Il s'agira maintenant d'exposer le choix de mon corpus sonore avant de présenter une proposition de sonification.

### Choix du corpus de sons non-voisés

Après avoir expérimenté différents types de corpus au cours de l'élaboration de ce projet, le choix du corpus final pour la phase de monstration publique s'est exclusivement porté sur des consonnes non-voisées de ma propre voix.

Un son non-voisé est un son issu de la parole n'impliquant pas la vibration de cordes vocales. Les consonnes dites « sourdes » telles que /p/, /t/, /k/, /f/, /s/ sont non-voisées contrairement aux /v/, /b/, /d/, etc.

Lorsque j'expérimentais différentes versions de mon algorithme, il était important pour moi de pouvoir tester rapidement les effets du *synthétiseur de nuées* avec divers types de sons. L'utilisation de ma propre voix pour la génération de matériau brut s'est avérée être très bénéfique car elle me permettait de sculpter des morphologies de micro-sons extrêmement diverses très rapidement, seulement par l'utilisation d'un micro.

J'ai donc enregistré en moyenne 80 micro-sons par consonnes, en variant les timbres (plus ou moins fort, plus ou moins aigü), ce qui me faisait au total un corpus d'environ 640 grains.

De la même manière que les compositeur·trices tels Horacio Vaggione travaillant la microcomposition en venant capter les grains en dessous des notes, sur les bruits d'attaque ou de respirations d'instruments, il s'agissait pour moi de venir dérober des fragments de consonnes en dessous des mots, dans leur état le plus brut, afin de les métamorphoser en d'autres sonorités.

Car, l'émergence de sonorités de second-ordre ne peut apparaître que par la perception d'un changement d'identité des sons. Il était donc indispensable d'utiliser un corpus d'unités sonores qui soit facilement reconnaissable : pour percevoir une transformation, il faut savoir identifier un état de départ et un état d'arrivée.

Le choix de me restreindre à de la parole non-voisée, plutôt que voisée, tenait au fait qu'il était plus facile d'atteindre l'émergence de textures par l'utilisation de sons bruités et percussifs, plutôt que d'utiliser par exemple des voyelles, qui impliquent une hauteur définie. Il était beaucoup plus difficile de s'abstraire de sons voisés, car ils rappellent inmanquablement une entité humaine, un caractère « humanoïde » artificiel.

Aussi, il était important pour moi de générer mon propre corpus plutôt que d'utiliser des signaux pré-existants, car il était alors possible d'enregistrer des sons qui pouvaient être identifiés comme des sons élémentaires. Ces micro-sons issus de sources uniques, fugaces et ponctuels, sont associés comme je l'ai mentionné dans la Partie 2.I à une action mécanique sous-jacente éphémère et indécomposable, émettant un son unique qui tend à disparaître dès l'instant où il est apparu. Les sonorités issues de corpus préexistants ont des morphologies déjà composées (elles sont la plupart du temps composées de plusieurs sons) : les conditions d'émergences de sonorités du second-ordre auraient donc été déjà compromises.

Comment alors organiser l'espace timbral de ce corpus de bribes de murmures ? Il s'agira maintenant d'analyser les descripteurs qui ont été utilisés pour ce projet.

### C. Organiser l'espace timbral du corpus sonore par descripteurs audio

L'analyse par descripteurs audio permet une classification des sons par leurs caractéristiques timbrales. Etant donné que le terme polysémique timbral ne veut en réalité pas dire grand chose en soi, car il renvoie à une multitude d'attributs perceptifs tels que le caractère brillant, percussif, doux, rugueux, bruité, etc. d'un son, il s'agira d'étudier les différents types de descripteurs permettant une distinction cohérente de notre corpus.

Commençons par identifier les différents timbres de notre corpus. Heureusement, les paroles non-voisées peuvent être facilement identifiables par leur consonne. Ainsi, mon corpus final se distingue par l'utilisation de 8 timbres différents : /s/, /sh/, /f/, /h expirés/, /t/, /k/, /ks/, /ps/.

J'avais eu l'occasion de tester d'autres consonnes telles que /p/, /r/, /ts/, /h inspirés/ etc. mais je les ai finalement écartés pour différentes raisons : soit les descripteurs ne parvenaient pas à les catégoriser de manière effective, soit les sons étaient trop peu identifiables comme des sons issus de bruits de bouche, soit le manque de temps ne m'a pas permis de finement régler les paramètres de descripteurs, ou d'en chercher davantage pour une meilleure analyse.

Quels sont les descripteurs qui permettent un classement cohérent de ces sons ?

Années après années, un panel très large de descripteurs ont été développés par l'IRCAM, et ont été rendus accessibles au travers de librairies de programmation, tels que le projet Zsa.Descriptors (Malt & Jourdan 2008), le projet Timbre Toolbox (Peeters et al 2011), ou encore les ircamdescriptors des plug-ins Pipo (Schnell et al 2017) inclus dans la librairie MuBu (Schnell et al 2009) et sur lesquels je me suis appuyé dans mon projet de recherche.

S'il en existe de très nombreux, les descripteurs peuvent se distinguer selon 5 catégories : temporelle (caractérisation de l'enveloppe dynamique comme l'attaque, la décroissance, etc.), spectrale (centroïde, étalement, asymétrie, facteur de crête, etc.), spectromorphologique (variation spectrale), harmonique (fréquence fondamentale, ratio harmonique/bruité, inharmonicité, etc.), perceptive (niveau sonore, brillance, etc.) (Peeters 2004).

Pour ce projet, j'ai décidé d'utiliser les descripteurs audio issus de la bibliothèque IrcamDescriptor. Nous n'allons pas dans le cadre de ce mémoire nous étendre sur l'analyse d'un panel de descripteurs mais allons seulement nous restreindre à ceux qui ont été choisis<sup>37</sup>.

Le protocole de sélection a été le suivant : après l'analyse du corpus par un large panel de descripteurs, il s'agissait d'évaluer la pertinence de chacun d'entre eux par l'écoute de la

---

<sup>37</sup> Le détail des formules mathématiques

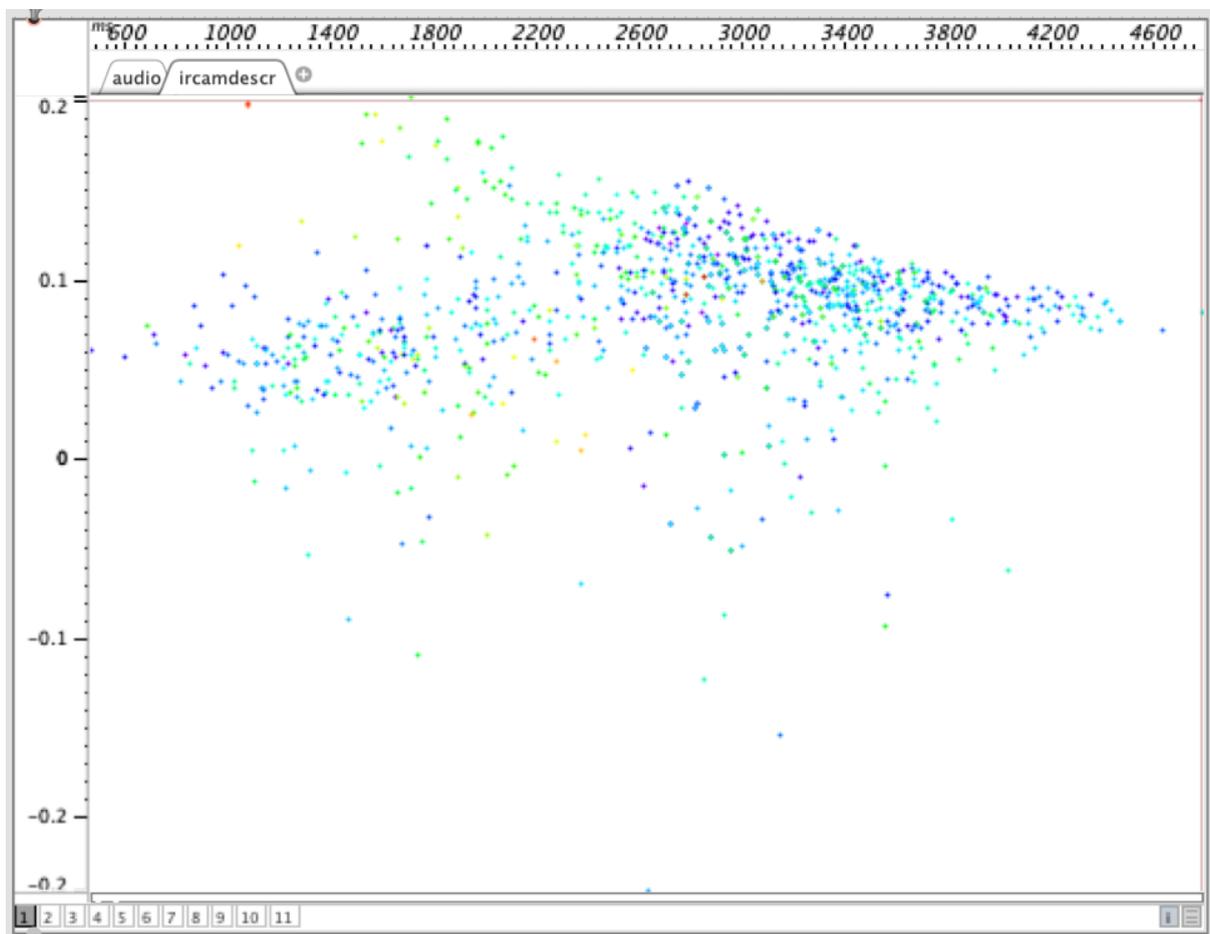
manière dont ils ont été classés. Ainsi, en partant de la valeur minimale de la métadonnée et allant progressivement jusqu'à sa valeur maximale, j'écoutais les sons les uns à la suite des autres afin d'y déceler une potentielle progression signifiante.

A titre d'exemple, le descripteur *Spectral Centroid* (barycentre spectral) me permettait de classer les micro-sons en fonction de leur « brillance ». Ainsi, un /x/, très brillant, a une valeur élevée, contrairement à un /h/.

Un autre descripteur qui m'a été très utile n'est autre que celui du niveau sonore (*Loudness*). Un /ps/, intégrant une plosive, sera alors classée plus forte qu'un /h/.

Enfin, le descripteur *Spectral Decrease* (« décroissance spectrale ») m'a été très utile pour caractériser l'enveloppe dynamique d'un micro-son. Il permet de classer les sons en fonction de la pente de décroissance de leur amplitude spectrale. Plus un son sera bref, plus il aura une pente de décroissance abrupte (à l'image d'une plosive /k/ ou /t/) et donc une valeur faible, contrairement à un /sh/ ou /f/, qui auront une valeur plus grande. Les sons brefs étant perçus plus « percussifs », nous utiliserons ce dernier terme pour les qualifier.

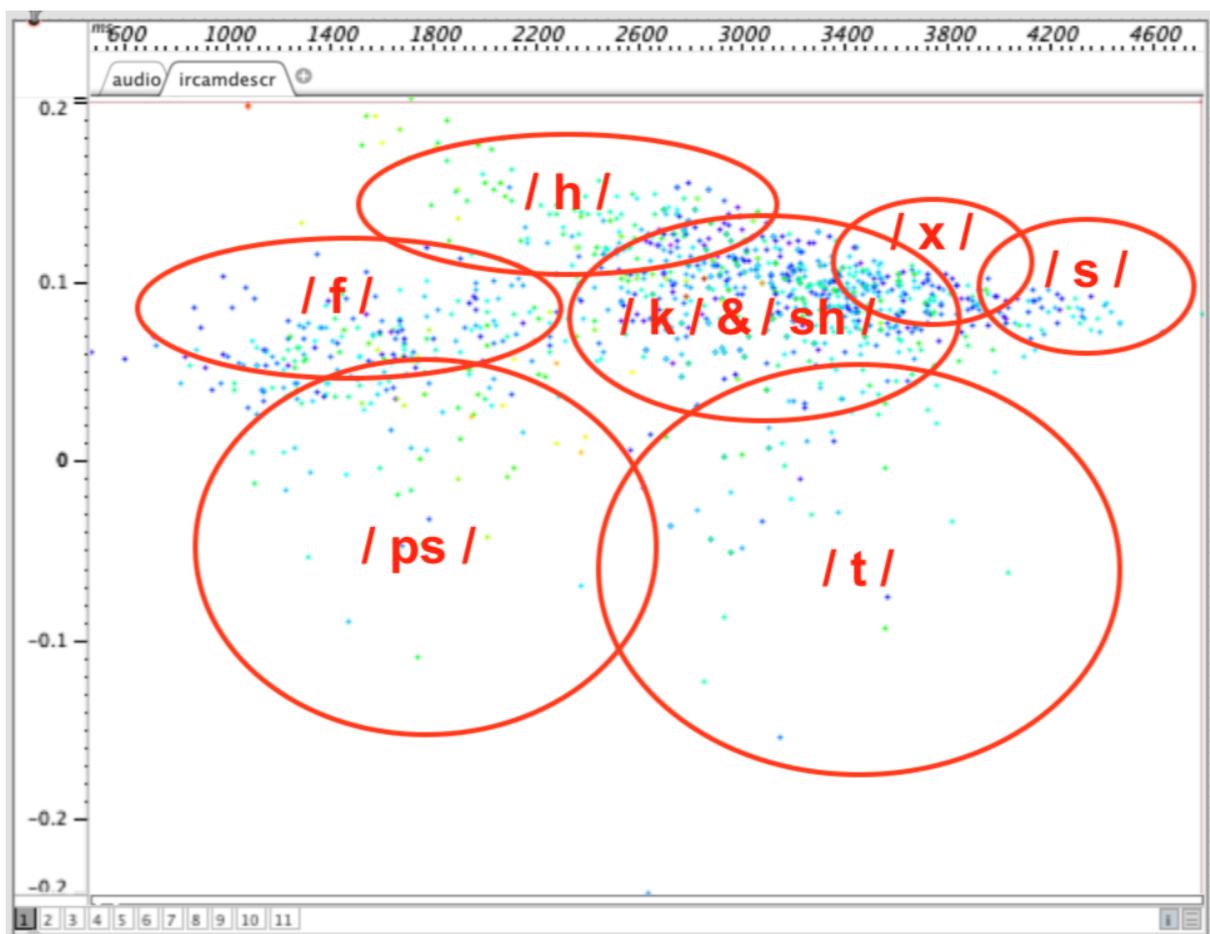
Ainsi, nous avons pu trier le corpus de micro-sons par leurs caractéristiques de niveau sonore, de brillance et de percussivité, nous permettant de les projeter dans un espace de représentation tridimensionnel, à la manière de cataRT.



Interface imubu permettant la représentation des grains selon les métadonnées choisies ici l'espace (x, y, couleur) est associé aux métadonnées (Spectral Centroid, Spectral Decrease, Loudness)

Chaque point représente donc un micro-son, et est caractérisé en abscisse (x) par sa brillance (*Spectral Centroid*), en ordonnée par sa décroissance spectrale (*Spectral Decrease*), et en couleur son niveau sonore (*Loudness*). Les sons au niveau sonore faible sont représentés en violet-bleu foncés, moyen en vert, et fort en jaune-orangé.

En écoutant les micro-sons par la navigation du curseur dans l'espace de représentation, on arrive clairement à identifier des zones dans lesquelles les différentes consonnes sont ancrées. A droite, on arrive à identifier les sons aigus /x/ puis /s/, tandis qu'on aura plutôt tendance à retrouver des /f/ à gauche. On arrive également clairement à identifier les sons percussifs en bas tels que les /ps/ ou /t/ et tandis que les /h/ seront en haut.



Le schéma délimite les zones dans lesquelles les phonèmes qui y sont associés ressortent le plus, mais celles-ci sont en réalité bien plus fluides : différents types de consonnes peuvent se mêler en fonction de la manière dont elles ont été produites. Cela n'est pas

étonnant, étant donné que j'ai tenté d'enregistrer la plus grande variété de sons possible afin de pouvoir bénéficier d'une large palette timbrale. Ainsi un /k/ pourrait être identifié plus percussif qu'un /ps/, ou un /sh/ pourrait être identifié comme plus grave qu'un /f/. Pour conclure, nous avons grâce aux descripteurs audio réussi à organiser un espace timbral cohérent, représentatif de la manière dont nous percevons les différentes consonnes du corpus. Ainsi, il sera possible de venir corrélérer les données des boids aux micro-sons par la sonification.

#### D. Système de sonification

A l'issue de l'étude de cas de Huepe (2014), j'avais émis l'hypothèse d'une sonification par corrélation métaphoriques entre le système source (boids) et le système cible (synthèse sonore).

Comment opérer ces traductions ?

Dans un premier temps, il a été question pour moi de tester un grand nombre de corrélations de manière empirique, guidées par mon intuition. J'étais davantage intéressé par l'effet que procurait l'agitation des entités du système, et l'imaginaire physique qu'elle invoquait, plutôt que la traduction d'un espace de représentation à un autre.

Par exemple, l'agitation frénétique des boids par l'application de forces aléatoires me rappelait l'agitation de grelots, produisant des sons de plus en plus forts à mesure qu'ils s'agitent de manière chaotique. Ou encore, l'association d'un alignement de groupe et d'une vitesse stable me rappelait le roulement de billes sur une surface, comme un raclement ou un ronronnement doux. La sensation de vitesse pouvait également se traduire par un changement de hauteur, une grande vitesse étant associée à un son aigu comme en témoigne le moteur d'une voiture (de la même manière que les corrélations de friction dans l'étude de Huepe), mais aussi à un niveau sonore plus élevé.

Ces corrélations restaient néanmoins très subjectives, provoquant des effets qui n'allaient pas nécessairement être perçues comme pertinentes par d'autres utilisateur·trices. Il était donc important pour moi de tenter de rendre plus "objectives" ces traductions.

J'ai donc essayé dans un second temps d'appliquer des transpositions de principes de regroupement Gestalt aux principes de regroupement de l'analyse sonore.

Par exemple, j'ai expérimenté une corrélation entre la densité spatiale des boids (grande lorsqu'ils sont regroupés) à la densité temporelle des grains (nuage de grain homogène lorsque la densité est grande). Mais finalement, ce type de corrélations théoriques ne fonctionnaient pas de manière systématique. En effet, si les boids se déplaçaient lentement mais étaient regroupés ensemble, ils produisaient tout de même une grande densité de grains. Et au contraire, la densité sonore pouvait être très faible lorsqu'ils se déplaçaient à toute vitesse de manière chaotique.

J'avais davantage tendance à corrélérer une grande densité à une grande agitation, puisqu'elles témoignent d'une plus grande présence d'entités (telle une foule).

Il semblait donc primordial de prendre en compte le caractère dynamique des boids, et de traduire principalement les inférences comportementales, plutôt que des dispositions statiques.

Comment rendre ces corrélations tangibles ?

Dans un troisième temps, il s'agissait de comprendre pourquoi certaines corrélations fonctionnaient alors qu'elles avaient été adoptées intuitivement. Je me suis alors rendu compte que j'avais tendance à associer le système boids à un modèle physique, correspondant à celui d'un instrument percussif. La sonification vue par ce prisme garantissait donc une certaine forme d'objectivation.

Par exemple, nous parvenons à relier la cause d'une génération sonore d'un son percussif à sa source parce que nous la voyons bouger : nous identifions l'action mécanique sous-jacente à l'émission du micro-son.

A partir de ce constat, la génération de sons lorsque les boids étaient à l'arrêt me paraissait par exemple incohérente. Si le système est lent ou ne bouge plus, alors il ne devrait pas émettre de sons.

### Vitesse

Avant d'aller plus loin, nous devons constater que nous avons oublié de mentionner une chose importante : j'avais émis l'idée que chaque boid devait être relié à un son, or nous travaillons ici avec des sons discontinus. A quelle fréquence chaque micro-son devait être déclenché ?

Une solution à ce problème se trouve dans le Poème Symphonique pour 100 métronomes de Ligeti. Si les sons qui émanent de la pièce sont discontinus, il y a bien des entités « continues », c'est-à-dire dont l'existence ne varie pas dans le temps contrairement aux micro-sons, qui émettent ceux-là. J'ai donc identifié dans l'implémentation de l'algorithme chaque boid à un métronome déclencheur de grains, dont le tempo fait partie des paramètres à corrélérer.

La sensation de vitesse a été plus évidente pour moi lorsqu'elle est rattachée à un changement de tempo, car elle correspond de manière analogue à la manière dont nous nous déplaçons. Entre une démarche de marche et de course, c'est d'abord la perception de l'accélération du rythme de notre foulée qui est en jeu plutôt qu'un changement de hauteur vers les aigus par exemple. On parle aussi d'un tempo plus « rapide » lorsqu'on l'augmente.

Mais en altérant le tempo de déclenchement de grain, c'est surtout la densité globale qui est changée. En effet, plus les tempi vont être rapides, plus il y aura de sons déclenchés en un temps donné, et donc plus leur densité sera grande.

Les changements de vitesse traduisent donc, avec cette sonification, des changements d'état « d'agitation ».

### Angulation relative

Une autre corrélation importante devait être celle de la traduction de l'angulation des boids. Elles évoquaient en moi soit un caractère « lisse » lorsque les boids étaient alignés, soit un caractère « percutant » lorsque leurs angulations étaient chaotiques. En effet, les mouvements des boids alignés en groupe me rappelaient le mouvement associé à un frottement, lorsque l'on passe du papier de verre (une surface granulée) sur une surface à poncer : les « grains » du papier vont tous dans le même sens.

J'ai donc décidé de lier le paramètre d'angulation au descripteur décrivant le caractère percussif d'un son : *Spectral Decrease*.

### Accélération

Aussi, l'accélération ou la décélération des boids évoquait chez moi une sensation de dégagement d'énergie plus ou moins grande. Lorsque je leur donnais une grande accélération, je souhaitais qu'ils puissent faire entendre plus leur voix, et au contraire se taire d'un seul coup lorsque je leur faisais ralentir. J'ai donc lié ce paramètre directement au niveau sonore.

### Regroupement spatial

Finalement le regroupement spatial des boids évoquait chez moi une sensation de masse, de lourdeur, comme si, regroupés ensemble, ils formaient une entité plus massive et lourde. Cette sensation était directement liée à la hauteur d'un son, plus grave si l'entité est lourde et inversement, j'ai donc décidé de lier la perception du nombre de voisins de chacun des boids au descripteur décrivant la brillance d'un son : *Spectral Centroid*.

### Bilan

Finalement, toutes ces sonifications ont été menées d'abord par de l'intuition. J'ai pu trouver certaines corrélations avec des facteurs « objectivants », mais il est évident qu'elles restent très subjectives. Si le synthétiseur de nuées était amené à être diffusé à un public donné, il faudrait soit que celui-ci apprenne et s'imprègne des corrélations qui ont été créées par moi-même, soit qu'il puisse les modifier à sa guise.

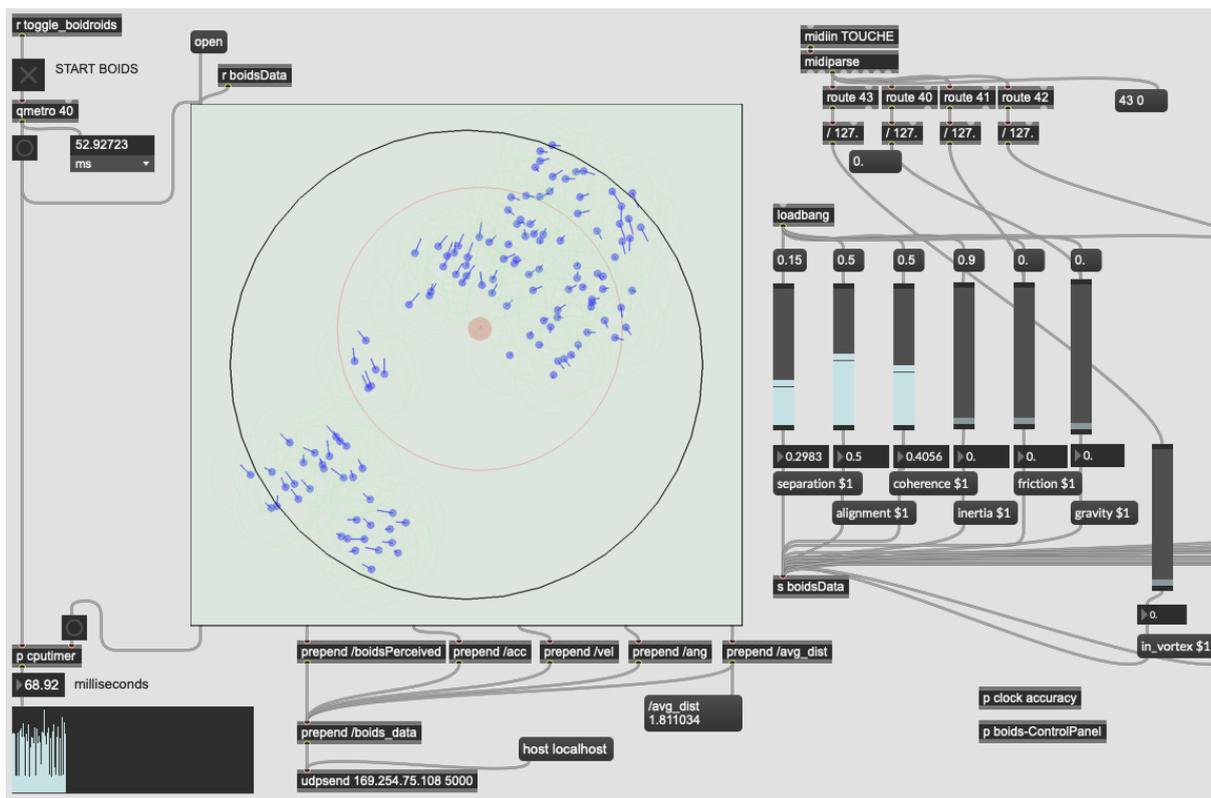
Il s'agira également de ne pas se laisser happer par cette interface graphique, car il ne faut pas chercher à ce que le son généré soit une illustration de ce que l'on voit. Il faudra alors tenter de s'en passer, en tentant à terme un jeu sans interface graphique. Celle-ci ne constitue qu'une représentation d'un système comportemental : l'indicateur d'une bonne sonification pourra être jugée seulement par l'écoute de ce qui émerge de ces nuées sonores.

## E. Structuration globale de l'algorithme

L'ensemble de l'algorithme a été conçu dans l'environnement Max/MSP, de l'interface boids au moteur de synthèse sonore. Il se décompose en 3 grandes parties : l'interface boids augmentée, le moteur de synthèse sonore, et enfin le module de corrélation de données qui fait le lien entre les deux autres parties.

### Interface boids :

L'algorithme boids augmenté a été implémenté en javascript au sein d'un objet *jsui* permettant d'avoir directement une interface utilisateur au sein de l'environnement Max/MSP<sup>38</sup>.



*Capture d'écran de l'interface boids du synthétiseur de nuées*

Les boids sont représentés par des points bleus, et leur vecteur vitesse par la tige qui en sort.

### Moteur de synthèse sonore

L'entièreté du module de synthèse sonore s'appuie sur la librairie MuBu (MultiBuffer) développée par l'équipe ISMM de l'IRCAM. Il s'agit d'une boîte à outils très complète pour l'analyse multimodale du son et du mouvement, la synthèse sonore interactive et l'apprentissage automatique (Schnell et al 2009).

<sup>38</sup> Le code est disponible dans son intégralité en Annexe.

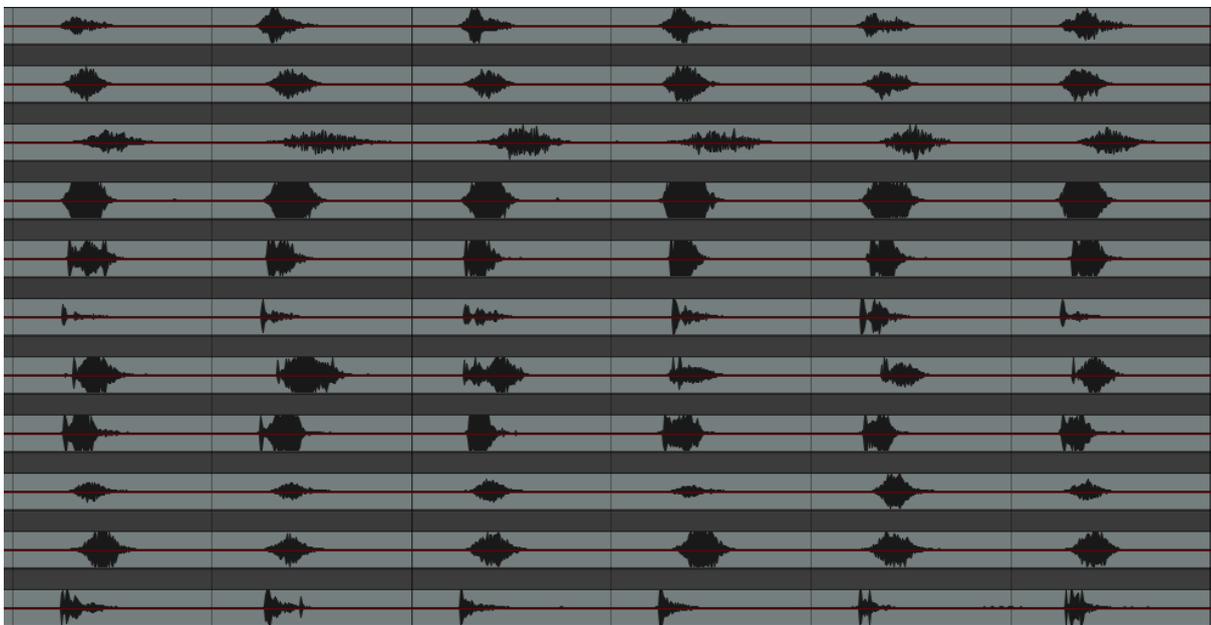
C'est à partir de cette librairie qu'à été implémentée une version Max/MSP de CataRT à partir de laquelle je me suis grandement appuyée pour l'implémentation du synthétiseur. Le moteur de synthèse sonore peut alors se décomposer en 3 modules distincts : le module d'analyse du corpus par les descripteurs audio, le module de sélection des grains et le module de lecture de grains.

#### Module d'analyse mubu.process

Le module d'analyse se segmente en deux parties, d'abord il segmente tout le corpus selon une période de temps donnée (ici 500ms) grâce à la fonction *chop*, découpant les signaux en grains, et dans un second temps analyse tout ces grains grâce à la fonction *ircamdescriptor*.

Concernant la segmentation du corpus, j'ai décidé pour des questions de commodité de monter les sons uns à uns selon une grille, permettant une segmentation systématique des fichiers en une période donnée (500ms). J'avais tenté d'utiliser dans un premier temps un algorithme de segmentation par détection de transitoires, mais certains sons à l'attaque douce (des /sh/, /s/ou encore /h/) étaient mal interprétés par l'algorithme de détection de transitoires.

Les grains sont ensuite identifiés selon leur index : *marker index*, correspondant à leur ordre au sein d'un signal, et *buffer index*, correspondant aux différents signaux. Par exemple le 54ème grain du signal numéro 4 (le signal correspondant aux sons /h/ par exemple) sera identifié comme suit : *marker 54, buffer 4*.



Formes d'onde du corpus de micro-sons non-voisés, montés selon une grille (500ms)

Ensuite, les descripteurs donnent une valeur de métadonnée à chaque grain.

	[time]	LoudnessMean	PerceptualSpectr..	PerceptualSpectr.
0	0	0.5506874	1445.557	0.0481314
1	500	0.6944656	1069.345	0.05317093
2	1000	0.6432462	1145.896	0.03764953
3	1500	0.8504802	1390.583	0.04356433
4	2000	0.6389763	1515.919	0.04497605
5	2500	0.6130506	1086.791	0.02995565
6	3000	0.7212488	1432.232	0.05096607
7	3500	0.764554	1781.162	0.07971352
8	4000	0.976641	2257.905	0.06066529
9	4500	1.196419	2015.09	0.07511117
10	5000	0.9159482	2892.608	0.0951701
11	5500	0.6260154	2536.48	0.08875078
12	6000	0.736457	1970.455	0.07574798
13	6500	0.7461289	2537.525	0.08804561
14	7000	0.5731134	1431.158	0.06557307
15	7500	0.7418224	1308.096	0.0585564

Vue "matrix" de l'objet imubu permettant de voir la valeur de métadonnées associées à chaque grain

### Module de corrélations

Ce module peut être décrit très simplement : il s'agit seulement de transformer les valeurs issues de métadonnées de boids (vitesse, angulation relative, etc.) en valeurs significantes de métadonnées sonores (barycentre spectral, niveau sonore, etc.), par un objet *scale* (simple opération de *scaling* linéaire).

### Module de sélection : mubu.knn

Après avoir été traduites, les métadonnées sonores vont être assignées à chaque « voix » sonore qui leur correspond. Si il y a 100 boids, alors il y aura 100 voix implémentées. L'objet `mubu.knn` (*k-nearest neighbors* (k plus proches voisins)) va alors sélectionner le grain qui se rapproche le plus des métadonnées ciblées par boids, comme nous l'avons décrit dans la Partie 3.I.B. Les grains ainsi sélectionnés sont identifiés selon leur index.

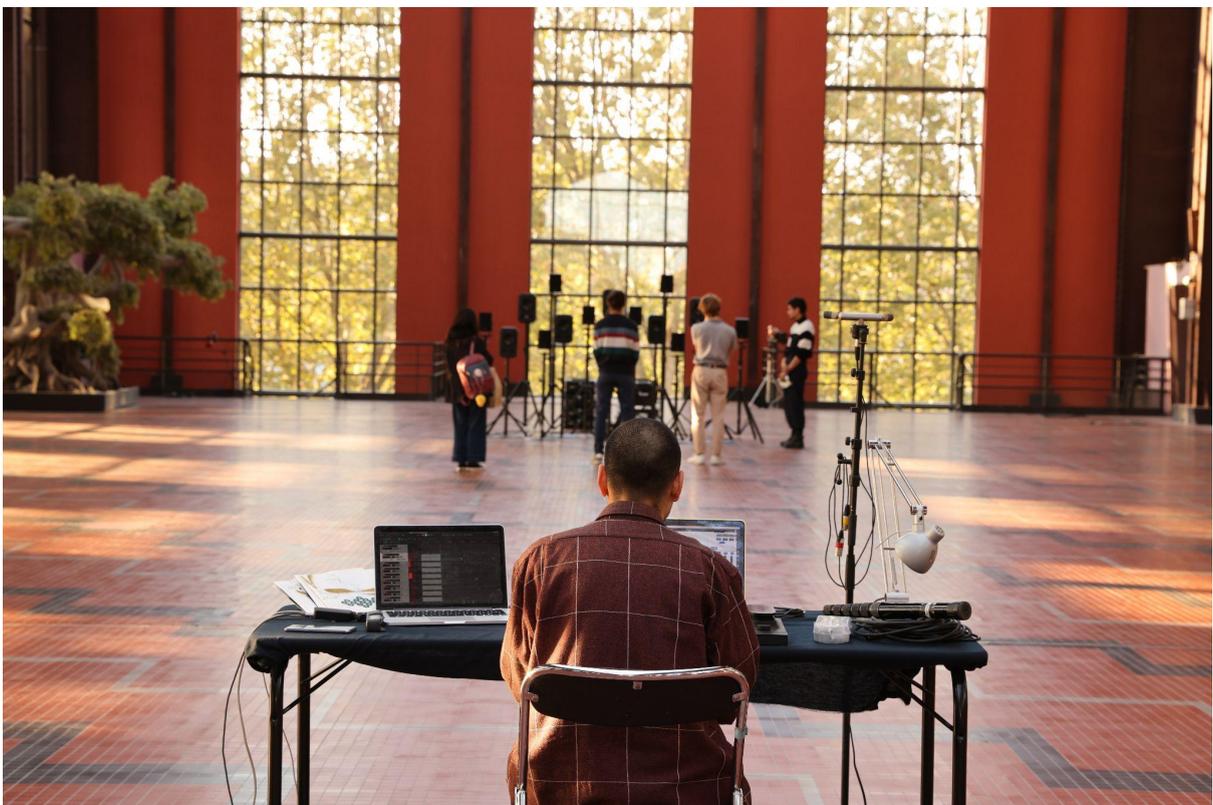
#### Module de lecture : `mubu.concat~`

Le dernier module correspond au synthétiseur concaténatif, permettant la lecture d'un grain selon l'index indiqué par le sélecteur `knn`.

Il n'y a donc, durant tout ce processus, eu aucun traitement de signal pour la synthèse de ces nuées sonores.

Maintenant que nous avons analysé la structure de l'algorithme du synthétiseur de nuées, il s'agira maintenant d'étudier les moyens de spatialiser ces masses sonores dans un espace acoustique. A cette occasion, il est maintenant temps de présenter la monstration publique de ma Partie Pratique de Mémoire (PPM).

## II. *Le bruissement des nuées* : monstration publique



© Matthieu Vachon

*Le bruissement des nuées* a été présenté du mercredi 13 au samedi 16 octobre 2021 dans la Nef de la Cité du Cinéma. Il s'agissait d'un événement commun avec la monstration de la PPM de Théophile Rey, l'installation *Les Vibrotopes*, présentée dans un espace adjacent. La monstration consistait en une série de concerts improvisés en direct, durant chacun environ 15 minutes. Durant ces concerts, je me situais sur la table « régie » sur laquelle je performais, à une vingtaine de mètres du l'ensemble de haut-parleurs, délimitant symboliquement l'entrée de l'espace de déambulation. J'invitais le public à vagabonder dans un espace de déambulation qui se trouvait devant le dispositif de 16 enceintes frontales, en les incitant à expérimenter différents points d'écoute en fonction de leurs envies. A l'issue du concert, je leur exposais un résumé de mon projet de recherche et leur dévoilais l'interface de contrôle boids. S'ils le voulaient, ils étaient ensuite invités à répondre à un questionnaire.

### Choix du dispositif de diffusion sonore

Lorsque je travaillais sur l'implémentation de l'algorithme de synthèse de nuées, j'étais contraint dans mes essais à une écoute au casque, et il devenait de plus en plus difficile d'envisager une restitution sonore par une simple diffusion d'un flux multicanal, par exemple à travers une diffusion stéréophonique.

Il ne s'agissait pas non plus pour moi de rendre compte du mouvement de groupe par la spatialisation de la trajectoire des nuées au sein d'un espace de diffusion multicanal de type WFS, HOA, etc. Je voulais trouver d'autres moyens de rendre compte du surgissement de l'effet de masse dans un espace acoustique.

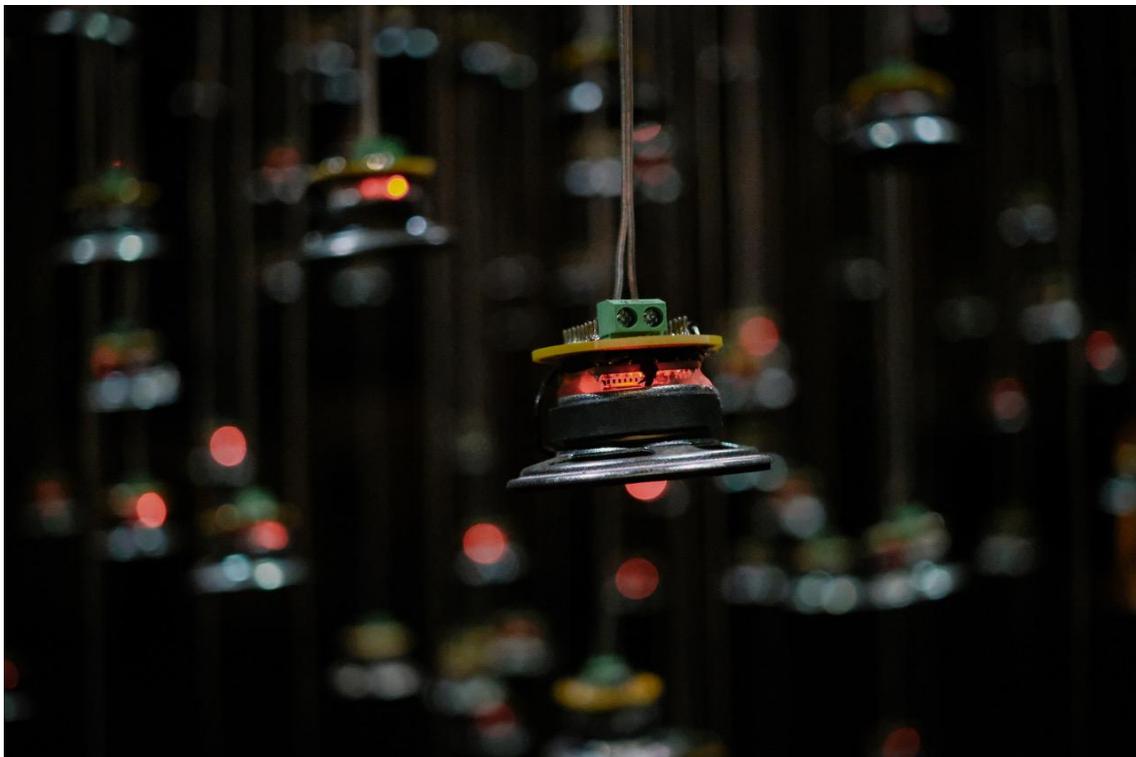
Dans son installation Speaking of Membranes<sup>39</sup> (2014), l'artiste Marianthi Papalexandri Alexandri convoque 136 membranes de hauts-parleurs suspendus par des fils au sein d'une armature cubique. Le principe de fonctionnement de cette installation est très similaire à celle de Zimoun que nous avons étudié jusqu'ici : l'action de moteurs se trouvant au bout des fils fait bouger de manière saccadée chaque membrane de manière aléatoire, parvenant à reproduire les sonorités d'une épaisse averse par l'agrégat de micro-impacts.



Marianthi Papalexandri Alexandri - Speaking of Membranes (2014).  
La chapelle de Saint-Joseph de Montfort / Meu, France

<sup>39</sup> captation visionnable via ce lien : <https://vimeo.com/117043997>

De la même manière, au travers de son installation Essaim<sup>40</sup> (2021), Felix Blume parvient à recréer la sonorité d'un essaim d'abeilles par la diffusion de bourdonnements individuels à travers 250 d'enceintes miniatures suspendues au plafond.



*Felix Blume - Essaim (2021)  
Ateliers Jeanne Barret, Marseille*

<sup>40</sup> captation visionnable via ce lien : <https://felixblume.com/essaim/>

L'une comme l'autre, ces deux installations parviennent à générer des effets de masse par l'utilisation d'une centaine de sources indépendantes. Elles simulent des phénomènes naturels en imitant leur modes d'émergence dans un espace acoustique. Nous parvenons alors à *ressentir* l'effet de masse par la perception d'une multitude de sources sonores, provenant de multiples localisations dans l'espace.

Je cherchais alors des moyens de permettre aux nuées *d'envahir* l'espace de la Nef, dans lequel elles étaient amenées à être projetées. La première étape consistait à trouver un dispositif permettant de générer le plus grand nombre de sources sonores possibles. Avec les moyens mis à disposition par l'École nationale supérieure Louis-Lumière, j'ai pu constituer un système de seize enceintes dont huit Amadeus PMX5 mk IV et huit Tannoy 600A.



*Installation-concert Le bruissement des nuées*

Il s'agissait de projeter les grains de la manière la plus directe, un son assigné à une seule enceinte, sans spatialisation virtuelle. Ce choix a été décidé pour qu'il ne puisse pas y avoir de sources fantômes, que l'auditeur·trice puisse identifier une source d'émission physique, et avoir la possibilité de venir se rapprocher de celles-ci pour l'écouter plus attentivement. Chaque voix était donc projetée au sein du dispositif à travers un unique haut-parleur, qui lui était assigné d'office au sein de l'algorithme.

Ce dispositif de spatialisation multi-mono s'inscrit dans une tradition acousmatique, à l'image de l'acousmonium, conceptualisé en 1974 par François Bayle au sein de l'INA-GRM. L'acousmonium déploie un ensemble de hauts-parleurs aux qualités distinctives, sélectionnées pour leur spécificité au sein du groupe, tels de véritables instruments au sein d'un orchestre. Intégré à l'ensemble, chacun d'entre eux aura donc un rôle spécifique, permettant de retransmettre des sons selon des "couleurs" particulières.



*François Bayle et l'acousmonium, Auditorium de Radio France  
dans le cadre d'Akousma x Présences Electroniques, octobre 2021*

Dans mon cas, les voix diffusées dans chaque enceinte étaient, telles que je l'avais implémenté, impossible à différencier, car la nature de leur contenu était exactement la même. Les enceintes étaient donc interchangeables entre elles, n'ayant pas une singularité particulière. Mais je ne cherchais pas non plus une restitution fidèle des sons au travers des enceintes. Ainsi, le choix des enceintes Tannoy et Amadeus m'étaient imposées pour des raisons logistiques, et me permettaient d'avoir deux "couleurs" de diffusion distinctes. Mais il était également envisageable d'utiliser des supports de diffusions différents, et hétérogènes entre eux. Une fois que les micro-sons se fondent la masse, leur singularité, leur "couleur" individuelle importe finalement assez peu. C'est lorsqu'on est amené à les écouter séparément, lorsque l'on perçoit une fission, qu'il est intéressant de déceler des différences. L'intérêt d'un tel dispositif réside dans le fait qu'il peut s'adapter aux conditions techniques et logistiques dans lesquelles il est amené à être exposé.

En prenant en compte toutes ces considérations, il était alors question de penser la manière dont l'espace de projection pouvait intégrer les enjeux de la perception des masses, à la fois par le dispositif de diffusion multi-mono lui-même comme je l'ai mentionné, mais également par l'espace d'écoute.

Dans le cadre de ces cycles de mini-concerts, il n'était pas question d'assigner un point d'écoute fixe au public, en les faisant s'asseoir, puisque la perception du phénomène d'émergence des masses dépendait grandement de leur position. Ils étaient donc libres de déambuler comme ils le souhaitaient au sein de ce terrain d'écoute, leur permettant de faire l'expérience de la modulation de la perception des phénomènes de masse par leur déplacement.

La perception du phénomène d'émergence dépendant grandement du contexte d'écoute, il s'agissait de prendre en compte le critère de validité écologique de l'installation (Di Scipio 2012).

Au travers de son installation sonore Forty Part Motet<sup>41</sup> (2001), Janet Cardiff propose une réinterprétation du motet *Spem In Alium* pour 40 voix indépendantes de Thomas Tallis. Chaque voix a été enregistrée indépendamment afin de pouvoir être diffusée dans un unique haut-parleur. Ainsi, convoqués dans le même espace acoustique, le chœur virtuel se déploie à travers l'ensemble de 40 hauts-parleurs.

Ce procédé très basique d'enregistrement et de restitution permet à chaque visiteur de vagabonder librement dans l'espace d'exposition, venant tantôt s'approcher des enceintes afin d'isoler une voix, tantôt s'en éloigner afin d'écouter l'ensemble du groupe. Comme nous l'avons évoqué dans la Partie 1.II, le déplacement du point d'écoute permet alors de moduler la sensation de fusion ou de fission de la perception des masses.



*Janet Cardiff - Forty Part Motet (2001)*

J'ai souhaité reproduire ce même procédé pour Le bruissement des nuées en proposant un espace de déambulation devant le (et autour du) dispositif dans lequel l'auditeur-trice était invité à circuler librement selon ses envies.

---

<sup>41</sup> documentation consultable via ce lien : <https://www.moma.org/collection/works/87291>

Les enceintes étant des entités neutres, le public pouvant sans crainte s'approcher par curiosité au plus près de la source d'émission.

L'intérêt d'un tel dispositif réside dans le fait même qu'il n'y ait pas de point d'écoute privilégié : c'est le déplacement de celui-ci au sein de l'espace de déambulation qui altère la perception des sons projetés. Autrement dit, c'est l'auditeur qui compose par le déplacement de son point d'écoute sa propre expérience du phénomène d'émergence.



*© Matthieu Vachon*

## Adaptabilité du dispositif de diffusion sonore



Le choix de la Nef comme lieu d'exposition a été déterminant pour cette monstration publique. Cet espace immense était particulièrement propice à la projection des nuées sonores, grâce à son temps de réverbération très long. En effet, l'éloignement du point d'écoute vis-à-vis du dispositif de hauts-parleurs accentue la sensation de fusion des sons grâce au champ diffus de la Nef. Le brouillage créé par la réverbération permettait aux micro-sons de davantage s'intégrer au sein du flux, laissant entendre une sonorité homogène à une distance assez proche des enceintes. En effet, la distance critique entre champ direct et champ diffus devait se trouver à seulement quelques mètres du dispositif. Il s'agissait alors de trouver le bon équilibre entre la distance inter-enceintes et la distance critique du point de fusion/fission spatiale perçu par le public. Par exemple, si elles avaient été disposées trop proches les unes des autres, la distance critique du point de fusion/fission spatiale aurait été plus proche des enceintes. Il aurait alors fallu s'en rapprocher davantage pour percevoir leur singularité.

Ainsi, ce dispositif de diffusion doit s'adapter dans le lieu au sein duquel il est exposé. Lors de la phase de prototypage de l'installation, j'avais pu investir certains espaces de l'Ecole Louis-Lumière tels que le studio musique ou la salle 10 du premier étage, des espaces à l'acoustique très mate. En plus de cette contrainte, ces espaces ne me permettaient pas d'avoir un recul de plus de 7 mètres environ entre la régie (mon point

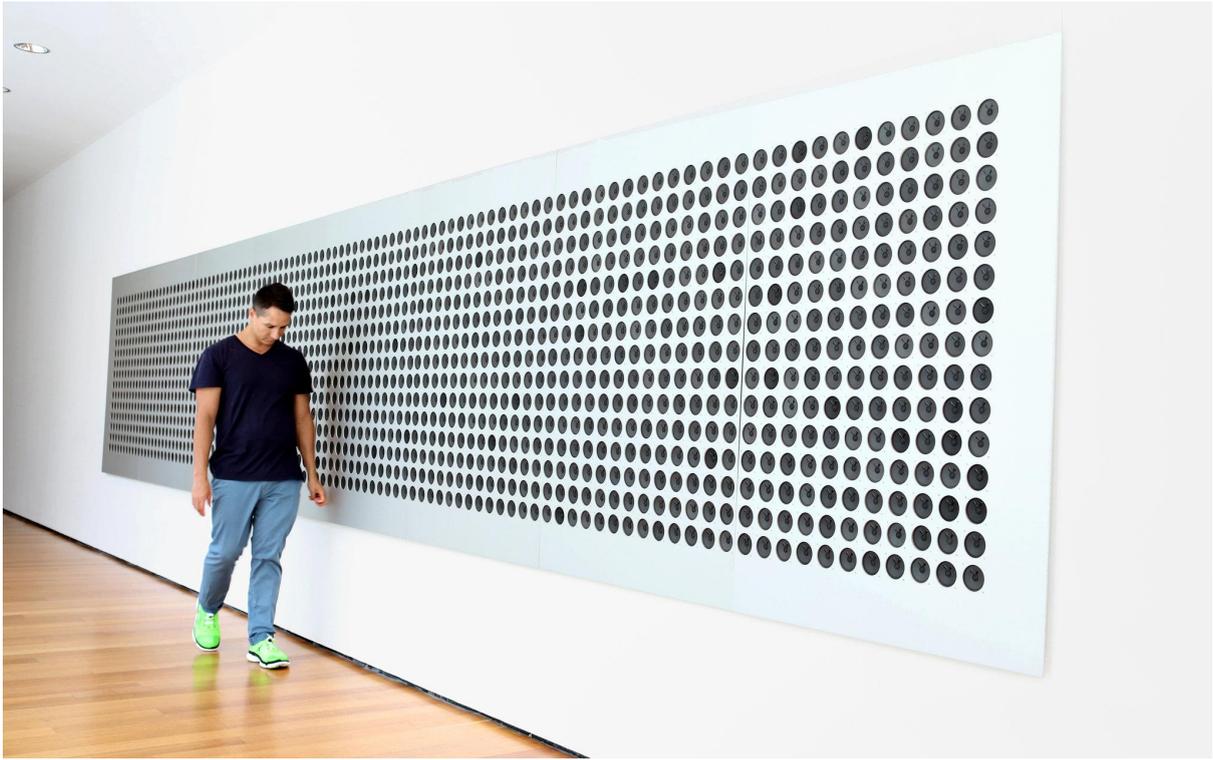
d'écoute) et les enceintes, ce qui limitait grandement les sensations de fusion des sons pour l'émergence de sonorités globales. Il était donc nécessaire de diminuer l'espace inter-enceintes pour favoriser la sensation de fusion par regroupement spatial. Il fallait alors attendre d'investir les espaces de la Nef, et d'avoir un recul suffisant avec le dispositif pour pleinement apprécier les effets escomptés.



*Prototype du dispositif de projection de l'installation, octobre 2021*  
*Les enceintes sont beaucoup plus resserrées que lorsqu'elles ont été installées dans la Nef*

On pourrait alors imaginer, pour chaque lieu investi, une configuration d'enceintes originale.

De la même manière que le mur de 1500 enceintes de Tristan Perich avec son Microtonal Wall (2011), il était primordial pour moi que le dispositif de projection sonore soit frontal. Dans son installation murale, l'artiste sonore diffuse dans chacune des enceintes un son continu élémentaire, ayant toutes une hauteur différente. En s'éloignant du mur, une sonorité du second-ordre émerge, ressemblant à un bourdonnement chaotique. Cette sensation de fusion par "synthèse additive spatiale" nécessite, comme nous l'avons vu dans la Partie 1.II, que le point d'écoute se trouve à équidistance des sources sonores, garantissant un aplanissement des distances perceptives : il sera plus difficile de dissocier les singularités des entités lorsqu'elles se retrouvent sur le même plan, favorisant l'émergence d'une masse globale.



*Tristan Perich - Microtonal Wall (2011)*  
*Museum of Modern Art (MoMA), New York City*

Par ailleurs, de la même manière que nous pourrions observer une nuée d'étourneaux, il s'agissait aussi pour moi d'opérer un rapport de vis-à-vis envers l'entité émergente à laquelle nous faisons face. Il est possible d'observer, ou d'entendre la forme dans sa globalité car son contour est contenu dans un espace délimité : nous pouvons l'identifier comme une entité homogène localisable dans l'espace.

L'exposition de l'installation dans ce grand espace me permettait également d'avoir un recul suffisant pour permettre la perception de nombreux niveaux de profondeur d'écoute. En opérant des vas et viens, l'auditeur·trice avait ainsi la possibilité de se trouver en dehors de la nuée, adoptant une écoute contemplative, dont la sonorité globale était submergée par la réverbération de la Nef, mais avait également la possibilité d'être "dans la nuée" en plongeant sa tête à proximité des sources, adoptant une écoute analytique d'étincelles sonores.

Il n'était donc pas envisageable d'adopter une disposition d'enceintes éclatée, par exemple circulaire, à la manière de Janet Cardiff, car le caractère discontinu des sons émis (très saillants) et la distance trop importante entre les enceintes n'auraient pas permis l'émergence d'une sensation d'homogénéité globale. Une plus grande densité de sources aurait néanmoins permis une étendue spatiale d'enceintes plus importante.

Le grand espace de déambulation garantissait également une grande liberté et fluidité de circulation. Même s'il n'y avait pas de restriction de jauge pour pouvoir assister aux concerts, il n'y a jamais eu plus d'une quinzaine de personnes souhaitant y assister en même temps. Chaque personne pouvait, tour à tour, venir se rapprocher du dispositif et profiter pleinement de l'expérience d'écoute sans jamais être dérangée par des obstructions partielles d'autres personnes.

### Les concerts improvisés

L'entièreté des concerts étaient entièrement improvisés, et étaient toujours interprétés par moi-même en direct. Il était prévu que ces concerts soient programmés selon des horaires donnés, mais en réalité, la plupart du temps je donnais des concerts en fonction de vagues successives de l'arrivée du public. Il m'arrivait donc de donner presque parfois une petite dizaine de concerts par jour. Le fait d'éprouver le synthétiseur des nuées sur ces 4 jours m'ont permis d'acquérir certaines techniques de jeu, des motifs récurrents d'états du système qui me servaient de point d'accroche, et que j'ai identifié selon différentes catégories :

- état compact et calme, presque à l'arrêt, produisant de légers murmures (mélange de "k", "ss" légers) au sein desquels les boids étaient regroupés ensemble, par lequel il est possible "d'éclater" d'un seul coup, produisant un puissant "ppssa !", ou "ha!" surgissant au son, de faire surgir des coups d'éclats
- état désordonné, chaotique, produisant des sonorités percussives foisonnantes
- les mouvements linéaires de groupes étaient perceptibles par des "sh"
- mouvements rotatifs, permettant l'émergence de sonorités "lisses" telles que des "hh"; mouvements alignés, envoyant à des sons de roulement de pierre, de frottements circulaires,
- mouvements groupés supersoniques, produisant des effets de passage forts et fugaces

Ces concerts improvisés me permettaient d'adapter mon jeu en fonction du point d'écoute du public. Ainsi, j'invitais parfois les auditeurs·trices à venir s'en approcher en abaissant le niveau sonore général, ou au contraire à s'en éloigner lorsque le système commençait à s'emballer de manière chaotique.

### III. Retour d'expérience

Au terme de l'étude de la perception des masses sonores et des recherches théoriques associées, il était temps pour moi, par la création du synthétiseur de nuées ainsi que par le biais de concerts, de confronter par la pratique les hypothèses annoncées au cours de ce mémoire. Il s'agissait donc pour moi de recueillir les avis des auditeur-trices ayant fait l'expérience de cette phase de monstration. A l'issue d'une courte présentation de mes travaux au terme de chaque concert, je les invitais à répondre à un petit questionnaire. J'exposerais donc ainsi une synthèse de la trentaine de réponses que j'ai pu recueillir, afin d'en tirer certaines conclusions.<sup>42</sup>

### La déambulation comme modulation de la perception de l'émergence des masses

Dans un premier temps, il s'agissait de savoir si la déambulation avait permis aux auditeur-trices de percevoir des différences de sensations notables vis-à-vis des sonorités générées, en fonction de leur point d'écoute :

*Le rapprochement et l'éloignement des hauts-parleurs ont-ils changé la perception de ce que vous avez écouté ? Si oui, comment ?*

Une très large majorité de réponses a répondu par l'affirmative.

Un certain nombre de personnes ont souligné l'importance de la réverbération dans ces processus de perception, permettant d'agir comme un "liant", faisant fusionner les entités éparses en une masse homogène. D'autres ont évoqué le fait que le rapprochement leur avait permis d'identifier la nature de la source sonore, les bruits de bouche. Un très grand nombre d'entre eux ont ainsi évoqué la dissociation entre une perception globale et une perception localisée, plus précise, qui permet de "mieux distinguer les sons".

L'hypothèse de l'utilisation de l'espace de déambulation comme un terrain de modulation de sensation de fusion ou de fission de l'émergence des masses est donc valide dans le cadre de cette monstration.

### Perception de mouvements de masse

Il s'agissait ensuite de savoir si les auditeur-trices avaient réussi à percevoir des mouvements de masse. La réponse à cette question était hautement importante car elle permet d'affirmer si les inférences comportementales de l'interface boids avaient correctement réussi à être transmises aux masses sonores. En définitive, il s'agissait de savoir si les *nuages de grains* avaient pu se transformer en *nuées de grains*.

*Avez-vous ressenti des mouvements de masse, de groupe de sons ?*

Encore une fois, c'est une écrasante majorité de réponses positives qui ont été recueillies. Dans un contexte où la majorité des utilisations de l'algorithme boids se concentrent sur la manipulation de trajectoires pour de la spatialisation de sources, la suggestion d'une

---

<sup>42</sup> Toutes les réponses recueillies pour ce questionnaire ont été mis en annexe

sensation de mouvement seulement par la manipulation dynamique de micro-sonorités dans le temps constituait pour moi un véritable défi.

### Sonorités évoquées

Dans Defining Transformations (1993), Denis Smalley évoque le fait que l'imaginaire convoqué par le compositeur·trice lorsqu'elle crée une pièce électroacoustique appelait très souvent des références qui diffèrent de celles de l'auditeur·trice. Contrairement à la musique traditionnelle, faisant l'usage de notes, il n'y a pas réellement de référence ou norme sonore en musique électroacoustique. Il était donc évident que le public allait percevoir des sonorités qui dépasseraient ce que j'aurais pu m'imaginer. J'étais donc très curieux de savoir quel type d'imaginaire fragments de murmures assemblés pouvaient convoquer.

Aussi, la question de l'évocation de sonorités différentes de celles renvoyant à la nature du corpus était hautement importante pour valider l'hypothèse de l'émergence de sonorités du second-ordre. Autrement dit, il s'agissait d'évaluer la capacité du synthétiseur de nuées, ainsi que de l'installation dans son ensemble, à transformer l'identité des sons qu'il manipule, seulement par le biais d'un agencement temporel de masses sonores.

*Que vous ont évoqué les sons que vous avez entendus ?*

Les réponses ont été extrêmement variées, et je distingue parmi elles trois catégories d'imaginaires associés.

D'une part, comme nous pouvions nous en douter après avoir mis en évidence dans la Partie 2 leur lien intrinsèque avec les masses sonores, les sonorités les plus évoquées appartiennent la famille de phénomènes naturels (non-vivants), ainsi que des transports : *"épisodes météorologiques", "bruits de la nature", "ruissellement", "vagues", "orage", "train", "bruissement de feuilles sèches", "passage de métro", "feux d'artifice", "vent qui souffle dans la rue, change de direction et qui se mêle à la pluie et la grêle", "ouragans", "multitude de bulles (de savon) qui éclatent"*.

D'autre part, on retrouve des imaginaires associés aux sons qu'émettent des entités vivantes: *"meute (chiens, loups)", "créatures du grenier dans Totoro", "fourmillements d'insectes", "foule", "brouhaha", "vie organique"*.

Enfin, on retrouve des imaginaires associés au monde fantastique ou aux pratiques occultes: *"rituel", "chant", "respiration d'un monstre", "démon", "exorcisme", "le fond d'une grotte", "sciences fiction", "chuchotements", "quelque chose de sombre"*.

Si dans les deux premières catégories d'imaginaire, nous retrouvons un vocabulaire relevant typiquement de celui convoqué pour décrire les masses, par l'imitation de leurs dynamiques, j'é mets l'hypothèse que c'est davantage la nature même du corpus sonore (les fragments de chuchotements) qui ont induits un imaginaire associé aux pratiques occultes.

Ce déluge d'évocations témoigne de la richesse et de la diversité des sonorités que permet d'évoquer la manipulation des masses. Ainsi, avec l'utilisation de seulement 8 consonnes non-voisées, sans même l'application de traitements, il est possible de faire émerger une très large diversité de sonorités.

Ainsi, il est possible de suggérer, en partant de la manipulation de la microstructure de bribes de murmures, *“des images auditives de sources sonores virtuellement concrètes, c'est-à-dire quelque chose que l'auditeur peut imaginer comme le son d'un phénomène acoustique réel mais qu'il ne peut pas tout à fait identifier”*<sup>43</sup>. (Di Scipio 1994, p.144)

### Perception d'une entité “vivante”

Après les premières discussions que j'ai pu avoir avec le public au lancement de la monstration, la mention de la perception d'une entité “vivante” est plusieurs fois revenue. Plus qu'une simple installation inerte, on m'avait plusieurs fois indiqué qu'elle leur évoquait “quelque chose de vivant”, comme une entité organique qui se meut. J'ai donc décidé d'ajouter une question au formulaire, afin de savoir si ce sentiment était partagé par la majorité du public.

*Avez-vous ressenti des présences d'une ou plusieurs entités “vivantes” ? Si oui, décrivez.*

Encore une fois, c'est une écrasante majorité d'affirmations positives.  
Comment expliquer l'évocation d'une telle sensation ?

D'une part, j'é mets l'hypothèse que la nature des sons issus du corpus ont pu largement contribuer à induire ce sentiment. En effet, l'écoute de sons issus de la parole, même de courtes consonnes, peuvent très facilement nous évoquer un sentiment de présence humaine. Certaines personnes m'ont d'ailleurs confié qu'elles avaient réussi à reconnaître ma voix à travers les haut-parleurs.

D'autre part, l'évocation de présences vivantes peut s'expliquer par le fait que l'algorithme de modélisation boïds qui sert “d'orchestrateur” est lui-même un système de vie artificielle, simulant des entités vivantes. La traduction de ces inférences comportementales dans le domaine sonore, par la sonification, pourrait donc être à même d'induire ce sentiment. En ce sens, l'écologie du mécanisme de génération sonore reflète bien l'écologie du phénomène sonore.

Enfin, cela peut également s'expliquer par le fait que, derrière cet algorithme, il y avait des inductions d'intentions par mon interprétation en direct. En effet, l'adaptation de mon jeu lors de la déambulation du public a probablement beaucoup importé dans l'évocation de ces sentiments. On aurait pu croire que l'entité était sentiente par le simple fait qu'elle pouvait interagir avec la présence d'un public.

Il était donc intéressant de constater que le public n'interagissait pas de la même manière avec l'installation en fonction de ce que ce sentiment allait suggérer chez eux.

On retrouve dans les réponses à cette question beaucoup de fois, par exemple, la mention de l'évocation d'un “monstre”, d'une “créature” ou encore d'un “fantôme”.

---

<sup>43</sup> citation originale : “[...] the richness of the sonic structure and the gestural discontinuities coming out of the micro-level process, however, give rise to auditory images of virtually concrete sound sources, i.e. something that the listener can imagine as the sound of a real acoustic phenomenon but cannot quite identify.”

Certains adoptaient donc un comportement craintif, se tenant à l'écart du dispositif pour ne pas se faire surprendre, tandis que d'autres assimilaient l'ensemble d'enceintes comme un simple dispositif de spatialisation inerte.

Nous pouvons expliquer ces comportements par le caractère intimidant, au comportement imprévisible et parfois sauvage du surgissement des masses, induites par mon propre jeu.

Comme nous ne sommes pas à même de *voir ce qui se trame derrière les enceintes*, un sentiment de crainte vis-à-vis de cette entité virtuelle pouvait alors émerger.

Chaque son provenant d'une source physique unique, il était aussi possible d'identifier chaque enceinte comme une voix indépendante, partie constituante d'une "créature" à plusieurs têtes, autrement dit une masse sonore identifiée comme un organisme unique.

Ainsi, l'installation elle-même devenait une **entité incarnée**.

### Biais perceptifs induits par l'interface graphique boids

Enfin, une dernière question me permettait de clore ce volet d'hypothèses. Après chaque concert, j'invitais le public à venir voir les "coulisses" de l'instrument que je jouais en direct, notamment en dévoilant l'interface de l'algorithme boids. J'ai mentionné dans la Partie 3.I.D que l'interface graphique pouvait potentiellement largement biaiser l'expérience sonore, car elle risquait seulement de devenir l'illustration sonore d'un support graphique. Je souhaitais alors interroger le public sur les potentiels biais perceptifs qu'il pouvait générer.

*Avez-vous observé l'interface graphique utilisée pour la manipulation des sons ? Si oui, votre ressenti sur les sonorités générées a-t-il changé après l'avoir observé ?*

Cette fois-ci, les avis étaient bien plus contrastés. Il était évident que l'interface modifiait les sensations d'écoute pour la plupart des personnes du public, car ils arrivaient ainsi à établir un lien source-cause qui leur permettait de comprendre d'où émanaient ces sons.

Pour certains, il était donc important de pouvoir voir le phénomène en même temps que de l'entendre. Pour d'autres, même si l'interface leur permettait de comprendre la manière dont était structuré l'instrument, leur sensations d'écoute restaient inchangées.

Dans tous les cas, lorsque je dévoilais l'interface, le public était systématiquement happé et émerveillé par les mouvements des boids. L'aspect spectaculaire et hypnotisant de la découverte de l'interface s'explique tout simplement par le fait que nous ne sommes pas habitués à manipuler ce type de systèmes.

Dans le cadre de l'implémentation du synthétiseur de nuées, il s'agira alors d'atteindre un degré de maîtrise de l'instrument suffisant afin de pouvoir s'en affranchir.

Comme je l'ai mentionné dans la Partie 2.II.B, l'interface graphique ne représente en réalité qu'une manière de monitorer graphiquement les données comportementales des entités boids. Pour l'instant, même après avoir joué un certain nombre de fois avec, il est encore difficile de m'en passer.

Dans tous les cas, il n'est pas question de révéler, ou de se servir de l'interface graphique pour qu'elle soit rendue visible au public. Les nuées sonores se suffisent à elles-mêmes, libérées d'une écoute "explicative", ou "illustrative" d'un support visuel, qui contraindrait par ailleurs largement l'évocation de sonorités diverses.

### Le synthétiseur de nuées comme instrument

Une chose est sûre, un apprentissage de la manipulation de l'instrument est indispensable afin de pouvoir en maîtriser les contours. Cet apprentissage est double : il s'agit d'une part d'apprendre comment manipuler les mouvements des boîis et de la manière dont ils peuvent réagir, autrement dit de maîtriser le système comportemental dans son entièreté; et d'autre part il s'agira de comprendre la manière dont sont corrélés les mouvements aux sons générés, en fonction des corrélations qui ont été pré-établies. Ces procédés de sonification peuvent cependant s'avérer contre-intuitifs pour la personne qui l'utilise. Il s'agira alors de changer les variables en jeu pour qu'il puisse correspondre au mieux à la manière dont l'utilisateur·trice perçoit le phénomène.

Toutes ces problématiques doivent être prises en compte pour un potentiel partage de l'instrument à d'autres utilisateur·trices.`

## Conclusions et perspectives

Les recherches menées au cours de ce mémoire nous ont permis de comprendre les mécanismes de la perception des masses sonores ainsi que leur mode d'émergence. Grâce

à l'étude d'un corpus de pièces de musique contemporaine, de Xenakis à Ligeti, mais également par l'étude d'un corpus d'installations sonores, nous avons pu mettre en évidence les qualités expressives et sensibles des masses sonores. Après avoir souligné le lien intrinsèque entre la perception des masses sonores et des sonorités issues de phénomènes naturels, il s'agissait de sonder les stratégies de composition de textures à partir de l'agencement de micro-sons.

*Comment, à partir d'une goutte, faire émerger la pluie ?*

La proposition d'un modèle de nuées d'étourneaux comme moyen d'articulation morphologique de ces textures s'est alors concrétisé par la création du *synthétiseur de nuées*. Les retours d'expérience recueillis auprès du public lors de la présentation de l'installation-concert, *Le bruissement des nuées* (partie pratique de mémoire), ont permis de confirmer mes premières hypothèses ; mettant en évidence la modulation de la perception des masses sonores par le déplacement du point d'écoute, ou encore de la capacité du *synthétiseur de nuées* à articuler la transformation de l'identité d'une sonorité.

*Qu'est-ce qui émerge d'une nuée de fragments de murmures ?*

Nous avons ainsi pu démontrer que, seulement par le biais d'un agencement temporel particulier d'un corpus restreint de micro-sons (en l'occurrence ici huit types de consonnes non-voisées), et sans avoir recours à du traitement, il était possible de créer des sonorités extrêmement riches et diversifiées; ces textures pouvant évoquer des phénomènes naturels (grêle, ouragan, vagues, orage, ruissellement, etc.) ou bien un sentiment de présence d'une entité « vivante ».

Les domaines de recherche associés à la composition par masses sonores restent néanmoins aujourd'hui assez peu investis vis-à-vis du potentiel qu'elle représente, et les enjeux de leur compréhension par le prisme d'études perceptives commencent à peine à être soulevés<sup>44</sup>.

Les difficultés de conception du *synthétiseur de nuées*, tout comme la particularité du dispositif de projection sonore, ou plus généralement de la nature non-uniforme des masses sonores, peuvent expliquer le manque de recherche et de ressources portant sur la synthèse des masses sonores. L'inspiration de phénomènes naturels pourrait peut-être constituer une voie privilégiée afin d'insuffler un intérêt pour la synthèse de masses<sup>45</sup>.

Le *synthétiseur de nuées*, ayant été à ce stade implémenté en vue du cadre spécifique de la PPM, présente néanmoins certaines limites. Par exemple, les corrélations ont été façonnées spécifiquement pour le corpus utilisé à cette occasion, et leur fonctionnement n'est pas garanti avec d'autres types de micro-sons.

---

<sup>44</sup> Comme en témoignent les études menées par Jason Noble (2018, Noble & McAdams 2020, Noble et al. 2020).

<sup>45</sup> On voit apparaître de plus en plus de modules ou plug-ins s'inspirant des mécanismes issus de phénomènes naturels, par exemple avec les procédés génératifs ou auto-poïétiques, comme en témoigne la collection de plug-in "Inspired by Nature" créée par Dillon Bastan inclus dans le populaire logiciel de création musicale Ableton Live 11.

A partir de cette première expérience, il s'agira d'envisager l'enregistrement d'un plus large corpus, présentant plus de variété, afin de questionner la capacité d'adaptation du système de corrélation, permettant alors une potentielle mise à disposition publique. Celle-ci pourrait par exemple se concrétiser par la création d'un module Max standalone, ou d'une intégration dans un plug-in Max4Live. L'adaptabilité de l'interface matérielle, ainsi que du support de diffusion sonore, seront alors également interrogés.

Il s'agit alors moins de créer un instrument multi-tâche, que de réfléchir à un système adaptatif, dont les enjeux peuvent se déplacer dans divers champs artistiques, que ce soit comme instrument musical ou comme outil de design sonore. L'écriture de ce mémoire a ainsi amorcé une pratique artistique protéiforme, dont la PPM constitue la première cristallisation.

Ainsi, j'espère que *Le bruissement des nuées* pourra investir d'autres espaces, et continuer son évolution, telle une entité initiant un dialogue avec les milieux dans lesquels elle s'immisce. Que ce soit par l'utilisation de corpus sonores spécifiques à ces espaces, de dispositifs de diffusion particuliers, ou par la prise en compte de contraintes spatiales et architecturales, il s'agit d'explorer la notion d'*in situ*, afin de composer de nouveaux terrains de déambulation.

## Annexes

Descripteurs audio utilisés :

→ ces captures d'écran sont tirées de Peeters 2004. Les équations devront être réécrites dans une version ultérieure

### 8.1.1 Total Loudness and specific loudness (cuidado:AudioLoudness): DP.i\_loud\_v

**Description :** (Moore, Glasberg et al. 1997)

The **specific loudness** is the loudness associated to each Bark band. We note  $N'(z)$  the loudness in the  $z^{\text{th}}$  Bark band. The precise expression of the loudness can be found in (Moore, Glasberg et al. 1997). This expression was approximated by neglecting terms of the expression acting only in specific cases (very weak signals) and by expressing it in relative scale (X. Rodet 2001). The specific Loudness can be expressed in a simple form:

$$N'(z) = E(z)^{0.23}$$

The **total loudness** is the sum of individual loudness (Zwicker 1990)

$$N = \sum_1^{nb\_band} N'(z)$$

where  $N'(z)$  is the specific loudness

### 6.1.1 Spectral centroid (mpeg7:AudioSpectrumCentroid) DS.i\_sc\_v

The spectral centroid is the barycenter of the spectrum. It is computed considering the spectrum as a distribution which values are the frequencies and the probabilities to observe these are the normalized amplitude.

$$\mu = \int x \cdot p(x) \, dx$$

where

- $x$  are the observed data:  $x = freq\_v(x)$
- $p(x)$  is the probability to observe  $x$ :  $p(x) = \frac{ampl\_v(x)}{\sum_x ampl\_v(x)}$

### 6.1.6 Spectral decrease (cuidado:AudioSpectrumDecrease)

**Description:** The spectral decrease also represents the amount of decreasing of the spectral amplitude. This formulation comes from perceptual studies, it is supposed to be more correlated to human perception.

**Formulation:**

$$decrease = \frac{1}{\sum_{k=2:K} a} \sum_{k=2:K} \frac{a(k) - a(1)}{k - 1}$$

## Réponses au questionnaire à l'issue du concert

*Que vous ont évoqué les sons que vous avez entendus ?*

- Des mouvements, des épisodes météorologiques (plume, tempete, retour au calme...)

- Sons de films de sciences fiction
- Une sorte de vie organique, plutôt non sympathique.
- Des chuchotements amplifiés - Des bruits de la nature amplifiés
- Rituel, chant, ruissellement, enveloppement, insectes
- Des vagues ou des fourmillements d'insectes
- Asmr / orage/ train
- Un monstre impressionnant mais bienveillant
- Étirement temporel, délitement, le moment où frodon met l'anneau et qu'il délire, un brasier d'âmes, le fond d'une grotte
- Parfois un bruissement de feuilles sèches, parfois un passage de métro, parfois un coup métallique
- Air, vent, eau, ruissellement , chuchotements, papier en mouvement
- Les sons m'ont évoqués des ambiances à la fois naturelles et mécaniques. Une immersion dans un milieu vivant où les éléments vivent à plusieurs rythmes mais toujours ensemble. Il existe à la fois un dialogue des sons entre eux et l'installation en apporte un autre avec le public. Et l'impression que cela pourrait durer des heures, avec des variations nouvelles à chaque fois. J'ai trouvé très intéressant que les bruissements soient portés au même niveau sonore que les sons plus francs.
- Sensation d'être dans une meute (chiens,loups), + chuchotements + feux d'artifice
- Le vent qui souffle dans la rue, change de direction et qui se mêle à la pluie et la grêle.
- La pluie, un bruissement, des craquements
- Le nombre, la masse, le mouvement mais parfois l'individualité avec une certaine proximité des sources comme un chuchotement à l'oreille
- Train, créatures du Grenier dans totoro
- Du brouhaha
- Des ouragans
- Chuchotement, quilles de bowling qui tombent
- Dynamique de foule
- Qqchose d'assez sombre
- Eau, beatbox, rituel satanique, démon, exorcisme
- Une foule, ou une sorte de monstre
- La respiration d'un monstre et l'agitation pleins de petits insectes. Et aussi une multitude de bulles (de savon) qui éclatent
- Des bruits de bouche, des bulles qui explosent, des circulations d'air, de grandes bourrasques

*Avez-vous ressenti des mouvements de masse, de groupe de sons ?*

- Oui, une masse formée de petits sons individuels
- Oui
- Moyennement, mon attention a été plutôt portée sur le déplacement des petits sons entre les différentes enceintes
- oui
- Oui
- Oui plusieurs fois notamment en prenant de la distance

- Oui
- Oui mais parce que j'ai lu la présentation avant...
- Oui
- Oui, on sentait qu'il y avait une direction commune même si chaque son n'évoluait pas en même temps que les autres
- Oui, avec des variations d'intensité, beaucoup de mouvements et des pauses qui accentuent la notion du mouvement. J'ai eu comme des sensations de déséquilibre parfois en fermant les yeux.
- Oui, il existe une réelle cohérence de groupe que l'on ressent très bien dans le mouvement des sons que l'on tente de spatialiser.
- Oui
- oui, d'un bruissement global émergeait souvent des formes de vagues sonores se propageant dans l'installation par exemple.
- Oui
- Oui
- Oui
- Oui
- Oui par moments
- Oui
- Oui
- Chuintements, cris de foule et d'individus
- Oui
- Oui
- Oui, surtout lorsque les sons étaient plus calmes et ordonnés
- Oui, des réunions , des séparations, des déplacements
- Oui très clairement
- Oui

*Avez-vous ressenti des présences d'une ou plusieurs entités "vivantes" ? Si oui, décrivez.*

- j'ai eu la sensation d'être face à quelque chose d'organique, un grouillement de sons
- Oui
- Plusieurs, peut-être pas, mais le ressenti d'une seule mais monstrueuse entité, en mouvement constant.
- oui
- Humains, insectes
- Oui des insectes qui grouillent mais également une multitude d'êtres humains
- Une meute qui se déplace en essayant de faire le moins de bruit possible

- Un monstre à plusieurs têtes :-)
- Sons en ch : souffle d'un démon ou truc du style. Sons plus percussifs : un lieu qui nous parle
- Plutôt une seule entité par moments, lorsque les sons convergent, même si en s'approchant on peut saisir plus les choses séparément. Ca m'a fait penser un peu à la limaille de fer à l'approche d'un aimant, qu'on peut voir comme plein de petits bouts ou comme un ensemble cohérent
- Oui, peut être plus dans l'esprit des insectes parfois, mais aussi des claquements de bec ou d'os
- Oui, j'avais l'impression d'être le plus souvent en forêt, où les enceintes agissaient comme des brindilles animées qui portent le son. On ressentait bien les grouillements, les frémissements, les sauts, les mouvements de course et d'autres plus appliqués. Le vent aussi.
- Oui comme des voix humaines, des animaux
- Oui, globalement la sensation était celle d'un foisonnement d'entités sonores habitant les enceintes et se regroupant parfois dans leur comportement pour un son global, ou un "cri" ensemble" et parfois vivant leur petite vie une chacune dans leur coin.
- Oui, mais plus un groupement de plusieurs entités
- Oui mais avec une répétition de ces entités (plusieurs fois la même sources)
- Oui, comme des voix ou des bruits de petites créatures
- Oui humaines (des fragments de mots)
- Oui, comportement chaotique donc paraît naturel
- Oui comme des fantômes ou des personnes
- Individus ou Groupes se rejoignant , se dispersant où se courant après
- Oui, je l'ai senti dans mon ventre. C'était envahissant
- Démon, monstre
- Oui surtout lorsque l'ordre venait être perturbé
- Oui à la fois une globale assez passionnée et des petites plus tranquilles
- Tantôt des individualités correspondants aux enceintes, comme des petits insectes et tantôt une sorte de masse voire une sorte de créature en mouvement par moments
- Oui, des déplacements d'oiseaux dans le ciel, du souffle humain, des mots susurrés mais incompréhensibles

*Le rapprochement et l'éloignement des hauts-parleurs ont-ils changé la perception de ce que vous avez écouté ? Si oui, comment ?*

- Oui, cela m'a permis d'isoler des sons ou au contraire de les entendre en masse
- Sons de groupe
- Oui, absolument. En proximité, le focus se fait plus sur les "petites bestioles" qui courent rapidement d'une enceinte à l'autre, alors que lorsque l'on s'éloigne, on ressent plus la "vie" de la grosse entité.
- non
- J'ai préféré la proximité car je sentais qu'on me racontait quelque chose

- Oui ! Le fait d'avoir plus ou moins de reverb et de ressentir le tout plutôt comme un bloc ou comme une multiplicité d'éléments individuels créé différents effets de nuées
- Le rapprochement permet de ressentir la sonorisation, l'éloignement permet de ressentir la masse
- Faudrait pouvoir comparer avec une écoute stéréo par exemple. Mais je me doute que ce doit être différent
- La grande reverb aide à tout prendre comme une seule masse, alors que de près on distingue chaque son individuel
- Oui, en s'éloignant un peu on ne peut plus déterminer les choses individuellement, on a plus une impression d'ensemble
- Oui, je reconnaissais plus les bruits de bouches de près
- Oui, j'avais l'impression que le son suivait mes mouvements.
- Pas forcément ressenti ça comme ça, plutôt plus immergé dans un groupe
- Un peu, et j'ai exploré en venant proche des enceintes et en m'éloignant, mais finalement j'ai adopté une distance assez constante ensuite, la (belle) reverb de la nef liant un peu l'ensemble placé trop loin je préférerais ne pas trop m'éloigner pour mieux sentir les mouvements de masses évoqués précédemment, et pas trop près non plus (ou juste à quelques moments par curiosité pour isoler plus une source) pour éviter d'avoir trop fort les effets de "cri" fort quand toutes les enceintes sont actionnées d'un coup en même temps
- Oui, j'avais l'impression d'entendre un bruit global en étant éloigné et d'être sous les intempéries en étant proche.
- En se rapprochant, on distingue mieux chaque son, en s'éloignant, il devient global
- En terme de niveau oui, l'écho est plus présent quand on s'éloigne.
- Le son est plus englobant de près
- Oui, on se sent plus ou moins inclus dans la bulle sonore selon la position dans l'espace
- On ressent plus les individualités quand on est proche
- Oui certains sons ne sonnent pas de la même manière je les perçois plus ou moins 'sec'
- Oui, plus proche les sons attrapent plus le corps
- Oui, on ressent bien l'effet de masse en s'éloignant et la perception plus localisée n se rapprochant
- On passe de l'entité unique aux multiples entités en se rapprochant de l'installation
- Oui, quand on se rapproche on a plus l'impression d'être au milieu d'un phénomène ou on peut percevoir des événements séparés et quand on s'éloigne c'est plus changeant. On perçoit des individus sans réussir à vraiment les distinguer les uns des autres et selon le son produit on perçoit une masse homogène
- Oui, changement dans l'immersion selon le placement, prise d'acoustique liée au lieu lorsqu'on s'éloigne qui rend imprécis mais englobe

*Avez-vous observé l'interface graphique utilisée pour la manipulation des sons ? Si oui, votre ressenti sur les sonorités générées a-t-il changé après l'avoir observé ?*

- Oui
- Oui, j'ai davantage saisi les mouvements que je pouvais entendre.
- Oui, le visuel influant quasi systématiquement sur le ressenti auditif, la perception n'est pas la même. Mais il aurait fallu avoir un écran à proximité du dispositif, et peut-être faire une déambulation de 5 mn sans l'écran, puis la même avec l'écran.
- non
- Pas beaucoup
- Non. J'étais trop occupé à écouter les nuées
- Oui, mon ressenti n'a pas changé
- Oui, perception plus proche de la présentation initiale.
- J'ai pu comprendre pourquoi tels sons se déclenchaient ensemble. Avec cette compréhension s'est un peu perdue la magie car je pouvais comprendre les paramètres en même temps que je voyais l'évolution de la nuée
- Certainement que ça joue dans la façon dont je me repasse rétrospectivement l'expérience dans la tête
- Oui, le ressenti ne change pas mais l'interface apporte beaucoup dans un autre registre. La spatialisation est différente et complète bien la compréhension de l'œuvre.
- Non
- Oui, j'ai d'ailleurs apprécié de pouvoir voir l'interface de contrôle, c'était intéressant de voir comment ces nuées étaient pilotées et surtout les choix de mapping entre descripteur audio et comportement de masse des boids. Ce n'a pour autant pas changé mon ressenti, on comprend l'intérêt du travail et la difficulté à ajuster la bonne relation entre paramètres de contrôle et synthèse sonore; mais l'effet à l'écoute reste le même, ni diminué (le côté "live", piloté en direct se sentant déjà bien même sans connaître l'interface - on a pas la sensation d'une bande son qui passe en tout cas), ni augmenté car on ne peut pas vraiment déduire/imaginer les gestes de contrôle effectués par le musicien juste par l'écoute du résultat.
- Oui, il me semble avoir perçu les sons de façon plus littérale après avoir vu l'interface.
- Pas de changement
- Oui une impression de réalité et de vérité se crée.
- On comprend mieux le fonctionnement du son, et le côté organique du son est renforcé par l'interface
- Oui les phénomènes semblaient plus explicites
- Oui on appréhende mieux les sons individuellement
- Oui
- Une fois que le concert était fini seulement, permet d'avoir une meilleure compréhension du comportement des entités
- Oui, et un peu, oui, ça a changé l'imaginaire associé à l'œuvre en passant d'un ressenti de foule et d'humain à un ressenti de phénomène naturel. Ce qui est assez intéressant
- On n'a pas la sensation de mouvement dans les sons produits là où c'est très marquant dans l'interface. Mais je ne crois pas que ça change ma perception. (je connaissais déjà le sujet de toute façon)

- Oui, mais mon ressenti n'a pas changé sur les sonorités, en revanche ça a modifié une perception de déplacement peut-être, dans le fait que ce soit visuellement restreint à un cercle

## Bibliographie

Perception, Psychoacoustique :

- Bregman, A. S. (1994). *Auditory Scene Analysis : The Perceptual Organization of Sound*. A Bradford Book.
- Douglas, C., Noble, J., & McAdams, S. (2016). *Auditory Scene Analysis and the Perception of Sound Mass in Ligeti's Continuum*. *Music Perception*, 33(3), 287-305. <https://doi.org/10.1525/mp.2016.33.3.287>
- Gabor, D. (1947). *Acoustical Quanta and the Theory of Hearing*. *Nature*, 159(4044), 591-594. <https://doi.org/10.1038/159591a0>

- Grabócz M. (2013). *Entre naturalisme sonore et synthèse en temps réel, images et formes expressives dans la musique contemporaine*. Archives contemporaines.
- Thoresen, L. (2015). *Emergent musical forms: Aural explorations*. University of Western Ontario.
- Noble, J. (2018, avril). *Perceptual and semantic dimensions of sound mass* (Thèse). [https://www.mcgill.ca/mpcl/files/mpcl/noble\\_2018\\_phdthesis.pdf](https://www.mcgill.ca/mpcl/files/mpcl/noble_2018_phdthesis.pdf)
- Noble, J., Thoret, E., Henry, M., & McAdams, S. (2020). *Semantic Dimensions of Sound Mass Music*. *Music Perception*, 38(2), 214-242. <https://doi.org/10.1525/mp.2020.38.2.214>
- Noble, J., & McAdams, S. (2020). *Sound mass, auditory perception, and 'post-tone' music*. *Journal of New Music Research*, 49(3), 231-251. <https://doi.org/10.1080/09298215.2020.1749673>
- Smalley, D. (1993). *Defining Transformations*. *Interface*, 22(4), 279-300. <https://doi.org/10.1080/09298219308570638>
- Wai-man Ip, G., Chiu, C. Y., & Wan, C. (2006). *Birds of a feather and birds flocking together : Physical versus behavioral cues may lead to trait- versus goal-based group perception*. *Journal of Personality and Social Psychology*, 90(3), 368-381. <https://doi.org/10.1037/0022-3514.90.3.368>

#### Musique :

- Berry, W. T. (1976). *Structural Functions in Music*. General Publishing Company.
- Di Scipio, A. (1994a). *Micro-time sonic design and timbre formation*. *Contemporary Music Review*, 10(2), 135-148. <https://doi.org/10.1080/07494469400640371>
- Di Scipio, A. (1994b). *Formal processes of timbre composition challenging the dualistic paradigm of computer music*. *Proc. of the ICMC94*, pp. 202–8. San Francisco: ICMA.
- Di Scipio, A. (1997). *The problem of 2nd-order sonorities in Xenakis' electroacoustic music*. *Organised Sound*, 2(3), 165-178. <https://doi.org/10.1017/S1355771898009029>
- Di Scipio, A. (2002). *The Synthesis of Environmental Sound Textures by Iterated Nonlinear Functions, and its Ecological Relevance to Perceptual Modeling*. *Journal of New Music Research*, 31(2), 109-117. <https://doi.org/10.1076/jnrmr.31.2.109.8090>
- Di Scipio, A. (2012). *Stochastics and Granular Sound in Xenakis' Electroacoustic Music*. In *Proceedings of the international Symposium Xenakis, La musique électroacoustique* (Université Paris 8, Mai 2012).
- Di Scipio, A. (2020). *Qu'est-ce qui est « vivant » dans la performance live electronics ? Une perspective écosystémique des pratiques de création sonore et musicale*. [Thèse de doctorat, Université Paris 8] [https://xoom.virgilio.it/source\\_filemanager/ag/go/agostinodiscipio/adiscipi/Di%20Scipio%20-%20th%C3%A8se%20EDESTA.pdf](https://xoom.virgilio.it/source_filemanager/ag/go/agostinodiscipio/adiscipi/Di%20Scipio%20-%20th%C3%A8se%20EDESTA.pdf)
- Keller, D. & Truax, B. (1998). *Ecologically-based Granular Synthesis*. *Proceedings of the 1998 International Computer Music Conference*. San Francisco: International Computer Music Association.
- Kramer, G. (1993). *Auditory display: Sonification, audification, and auditory interfaces*. Perseus Publishing.

- Ligeti, G. (1997). *Music-Making Machines : Livret pour Mechanical Music* (trad. Annelies McVoy et David Feurzeig), New York, Sony Classical, coll. « György Ligeti », 5<sup>e</sup> éd.
- MacKay, J. (1984). *On the perception of density and stratification in granular sonic textures : An exploratory study*. *Interface*, 13(4), 171-186.  
<https://doi.org/10.1080/09298218408570451>
- Miranda, E. R., & Biles, J. A. (2007). *Evolutionary Computer Music*. Springer.
- Roads, C. (2001). *Microsound*. MIT Press, Cambridge.
- Schwarz D. (2011) *State of the Art Sound Texture Synthesis*. Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx-11), Paris, France, September 19-23.
- Solomos, M. (2006). *The Granular Connection (Xenakis, Vaggione, Di Scipio, ...)*. The Creative and Scientific Legacies of Iannis Xenakis International Symposium. Guelph, Canada.
- Solomos, M. (2013). *De la Musique au Son, L'émergence du son dans la musique des XXe-XXIe siècles*. PU Rennes.
- Souris, A. (1976). *Conditions de la musique et autres écrits*. Éditions de l'Université de Bruxelles, Éditions du CNRS, Paris.
- Vaggione H. (2006). *Symboles, signaux, opérations musicales*. Paris, MSH/Paris Nord, Université Paris VIII.
- Vaggione H. (2010). *Représentations musicales numériques : temporalités, objets, contextes*, dans Soulez, A., & Vaggione, H. (2010). *Manières de faire des sons*. L'Harmattan.
- Xenakis, I. (1963). *Musiques formelles : Nouveaux principes formels de composition musicale*. La revue musicale. Éditions Richard-Masse.
- Xenakis, I. (1969). *Terretektorh - Nomos gamma*. livret
- Xenakis, I. (2006). *Musique de l'architecture*. Editions Parenthèses.

#### Modélisations, boids :

- Apergis, A., Kaliakatsos-Papakostas, M., & Floros, A. (2018). *Sonoids : Interactive Sound Synthesis Driven by Emergent Social Behaviour in the Sonic Domain*. 15th Sound and Music Computing Conference (SMC2018), Limassol, Cyprus, 4–7 July.  
<https://doi.org/10.5281/zenodo.1422680>
- Blackwell, T., & Young, M. (2004). Self-organised music. *Organised Sound*, 9(2), 123-136. <https://doi.org/10.1017/s1355771804000214>
- Davis, T., & Rebelo, P. (2005). *Hearing Emergence : Towards Sound based Self-Organization*. ICMC.
- Huepe, C., Colasso, M., & Cádiz, R. F. (2014). *Generating Music from Flocking Dynamics*. *Controls and Art*, 155-179. [https://doi.org/10.1007/978-3-319-03904-6\\_7](https://doi.org/10.1007/978-3-319-03904-6_7)
- Reynolds, C. W. (1987). *Flocks, herds and schools : A distributed behavioral model*. *ACM SIGGRAPH Computer Graphics*, 21(4), 25-34.  
<https://doi.org/10.1145/37402.37406>