



HAL
open science

Vers un modèle pour la sonification de données scientifiques

Maxime Poret

► **To cite this version:**

Maxime Poret. Vers un modèle pour la sonification de données scientifiques. Traitement du signal et de l'image [eess.SP]. Université de Bordeaux, 2022. Français. NNT : 2022BORD0435 . tel-03989868

HAL Id: tel-03989868

<https://theses.hal.science/tel-03989868>

Submitted on 15 Feb 2023

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

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



THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE
DE MATHÉMATIQUES ET D'INFORMATIQUE

par **Maxime PORET**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

**Vers un Modèle pour la Sonification de Données
Scientifiques**

Soutenue le 14 décembre 2022

Devant la commission d'examen composée de :

Mme Jenny BENOIS-PINEAU	Professeure Université de Bordeaux	Présidente
M. François PACHET	Docteur Spotify Paris	Rapporteur
M. Sylvain MARCHAND	Professeur La Rochelle Université	Rapporteur
M. Alexander MIHALIC	Président de SAS Alter Instruments	Examineur
Mme Myriam DESAINTE-CATHERINE	Professeure Bordeaux INP	Directrice
Mme Catherine SEMAL	Professeure Bordeaux INP	Directrice

Remerciements

Dans cette seule partie de la thèse où je m'exprimerai à la première personne du singulier, je tiens à exprimer ma gratitude envers celles et ceux qui m'ont aidé à mener à bien ces trois dernières années de travail.

Merci tout d'abord à François Pachet, Sylvain Marchand, Alexander Mihalic et Jenny Benois-Pineau pour avoir accepté de faire partie du jury pour la présente thèse.

Merci à mes directrices Myriam Desainte-Catherine et Catherine Semal pour leur patience, leur écoute, et pour avoir réussi à me convaincre à de nombreuses reprises que je serais réellement capable de finir une thèse sur un sujet aussi ambitieux que celui abordé ici.

Merci à David Auber et Nadine Couture pour avoir fait partie de mon comité de suivi de thèse, et pour leurs remarques sévères mais justes à l'issue de notre première réunion, qui m'ont donné le coup de fouet qui manquait pour réellement consacrer au doctorat le sérieux qu'il méritait.

Merci aux personnes impliquées dans les différentes collaborations qui ont été entamées au cours de la thèse : Nadine Couture, Sébastien Ibarboure, Emmanuel Duc, Sylvain Dupont, Lisa Wingate, Yacine Amarouchène, Grégory Giannone, Olivier Thoumine et Matthieu Lagardère.

Merci aux étudiants qui ont participé à différentes étapes d'implémentation et d'étude de l'existant lors de ces travaux : Johan David de l'Université de Bordeaux, Maelle Andricque de l'ENSEIRB-MATMECA et Martin Devreese de l'ENSC.

Merci à Bernard Serpette, pour sa relecture, ses remarques et ses conseils quant à l'aspect fonctionnel du modèle proposé ici.

Merci à Jean-Michaël Celerier pour avoir pris le temps de m'expliquer l'utilisation d'ossia score et pour sa réactivité à résoudre les bugs signalés.

Merci à Matthias Robine et Pierre Hanna, à qui je dois l'essentiel de mon éducation en informatique sonore.

Merci aux gestionnaires du SCRIME Annick Mersier et Remy Piazza pour leur accueil, leur gentillesse, et pour avoir grandement clarifié et simplifié les tâches administratives depuis nos premiers échanges en 2019.

Merci à Thomas Hermann qui, lors de notre unique rencontre à NordicSMC en 2019, m'a conseillé de suivre ma propre inspiration avant tout. J'espère avoir réussi à suivre ce conseil.

Merci enfin à ma mère, Florence Poret, pour son soutien indéfectible au cours d'une période tourmentée qui m'a par ailleurs coûté bien des relations.

Résumé La sonification est la représentation de données sous une forme auditive non verbale. Il s'agit d'un domaine de recherche relativement jeune, issu de technologies assez récentes, et pour lequel il n'existe pas encore de standard clair. Après quelques décennies de réflexions sur la modélisation générique du concept et des centaines de solutions au cas par cas, il semble intéressant d'effectuer une synthèse des points communs et variables de tout ce que l'on peut appeler une sonification afin d'en extraire un « patron » pour ce type de représentation. Dans cette thèse, nous rapportons nos travaux sur trois cas d'études, ainsi qu'une étude de l'existant plus générale, visant à poser les bases d'une telle modélisation. Notre modèle prend ainsi la forme d'une fonction paramétrée par les données et les interventions de l'utilisateur, dont la sortie est le son construit. On propose un format graphique pour le représenter, avant de réfléchir à son potentiel en tant qu'outil informatique de design.

Mots-clés représentation, représentation auditive, sonification, synthèse sonore, informatique sonore, interface humain-machine, psychoacoustique

Laboratoire d'accueil Laboratoire Bordelais de Recherche en Informatique, 351 cours de la Libération, 33400 Talence

Title Toward a Model for the Sonification of Scientific Data

Abstract Sonification is the use of non verbal sounds for the display of data. It is a relatively new field of research, which stems from somewhat recent technologies and for which there is no clear standard yet. After a few decades of research into generic modeling of that concept, as well as hundreds of ad hoc solutions, it seems worthwhile to compile a synthesis of the commonalities and variables that make up any sonification, so as to describe a “template” for those types of displays. In this thesis, we report our work on three case studies as well as a more general review of existing works, in order to lay down the groundwork for such a model. The model we then propose is defined as a function which takes the data and user input as parameters and outputs an audio signal. We describe a graphical format for its representation before studying its potential use as the basis for a sonification design tool.

Keywords displays, auditory display, sonification, sound synthesis, sound computing, computer-human interface, psychoacoustics

Hosting Laboratory Laboratoire Bordelais de Recherche en Informatique, 351 cours de la Libération, 33400 Talence

Table des matières

Remerciements	i
Résumé étendu en français	1
Introduction	3
I Contexte	9
1 La sonification	11
1.1 Définitions	11
1.2 Motivations	13
1.2.1 Avantages de l’audition	13
1.2.2 Applications	14
1.3 Synthèse sonore et perception	15
1.3.1 Principes généraux	16
1.3.2 Oscillateurs	17
1.3.3 Transformation de Fourier	20
1.3.4 Techniques de synthèse sonore	21
1.3.5 Outils pour la synthèse sonore	21
1.3.6 Paramètres perceptifs	24
1.3.7 Interaction entre paramètres perceptifs	28
1.3.8 Masquage perceptif	28
1.4 Techniques de sonification	29
1.4.1 Audification	29
1.4.2 Parameter mapping	30
1.4.3 Auditory icons	30
1.4.4 Earcons	31
1.4.5 Model-based	32
1.5 Vers un modèle générique pour la sonification de données	32

II	Cas d'études	35
2	Supervision du procédé de fabrication additive wire-arc	37
2.1	Contexte	37
2.1.1	Affiliations	37
2.1.2	Le procédé	38
2.1.3	Enjeux de la représentation en réalité augmentée	38
2.1.4	Contexte sensoriel de la fabrication	39
2.1.5	Intuitions pour le diagnostic par l'écoute	39
2.1.6	Critères de fabrication	40
2.1.7	Simulation de l'étude	40
2.2	Évolution des sonifications proposées	42
2.2.1	Préliminaire : sonifications conçues au cours du stage recherche	42
2.2.2	Aspects récurrents	43
2.2.3	Instances consécutives et adaptations	44
2.2.4	Perspectives pour l'amélioration de la représentation auditive	49
2.3	Protocole d'évaluation « en laboratoire » pour la comparaison de représentations auditives périphériques	49
2.3.1	Préliminaire : tâches d'identification	50
2.3.2	Existant pour l'évaluation de sonification ambiantes	52
2.3.3	Protocole d'évaluation proposé	53
2.3.4	Données récoltées	55
2.3.5	Critères de qualité d'une représentation	56
2.3.6	Perspectives	57
2.4	Impact de l'incohérence écologique sur la qualité d'une sonification	59
2.4.1	Cohérence écologique dans l'existant	59
2.4.2	Notre hypothèse	60
2.4.3	Évaluation comparative d'écologies sonores	60
2.4.4	Résultats	61
2.4.5	Conclusion et perspectives	66
2.5	Perspectives pour la réalisation d'un procédé wire-arc sonifié	67
2.6	Conclusion générale	68
3	Sonification immersive du suivi de protéines individuelles pour l'étude de la motilité cellulaire	71
3.1	Contexte	71
3.1.1	Affiliations	71
3.1.2	Domaine	72
3.1.3	Motivations pour une sonification	74
3.2	Existant	75
3.2.1	Pour le projet TMBASS	75

3.2.2	Projets similaires	77
3.2.3	Réflexions sur la model-based sonification	78
3.3	Sonification proposée	78
3.3.1	Sonification interactive et immersive	78
3.3.2	Mappings perceptifs	79
3.3.3	Mode sonde	80
3.4	Implémentation dans FluoSim	81
3.4.1	Développement initial	81
3.4.2	Retouches ultérieures	82
3.4.3	Usage	83
3.5	Perspectives	85
3.5.1	De meilleurs timbres	85
3.5.2	Fonctionnalités supplémentaires	86
3.5.3	Protocole expérimental envisagé	86
3.6	Conclusion	88
4	Vers une modélisation de la réponse acoustique d'un environne- ment à des phénomènes physiques	91
4.1	Contexte	91
4.1.1	Affiliations	91
4.1.2	Domaine de recherche	92
4.2	Analyse des corrélations entre mesures d'anémomètres et sons de l'environnement	93
4.2.1	Proposition	93
4.2.2	Motivations	93
4.2.3	Difficultés	94
4.2.4	Existant	95
4.3	Expérimentation préliminaire	97
4.3.1	Capture de données synchrones	97
4.3.2	Observation des spectres de nombres d'ondes	100
4.3.3	Étude de la faisabilité	102
4.4	Vers de nouvelles techniques de sonification par apprentissage	104
4.5	Conclusion	106
III	Vers un modèle générique pour la sonification de données	107
5	Description exhaustive d'un outil de sonification	109
5.1	Données d'entrée	110
5.2	Pré-traitement des données	111

5.3	Son en sortie	112
5.4	Paramètres sonores	113
5.5	Fonctions de mapping	114
5.6	Temporalité du son	115
5.7	Interaction	116
5.8	Personnalisation	117
5.9	Conclusion	118
6	Étude de l'existant	119
6.1	Modèles de design	119
6.1.1	Design patterns	120
6.1.2	Espaces de design	121
6.1.3	Discussion	123
6.2	Modèles de mapping	123
6.2.1	Fonction de transfert données-son	124
6.2.2	Mappings entre espaces perceptifs	126
6.3	Outils généralistes de sonification	127
6.3.1	Remarques finales sur les outils généralistes de sonification	131
6.4	Conclusion	132
7	La Cellule de Sonification	133
7.1	Critères pour le modèle	133
7.2	Vue d'ensemble	134
7.2.1	Fonction de transfert de la sonification	134
7.2.2	Représentation graphique de la fonction de transfert	136
7.2.3	Rôle dans le design d'une sonification	138
7.3	Bloc des données d'entrée	138
7.4	Bloc d'interaction utilisateur	139
7.5	Bloc des paramètres réglables	140
7.6	Bloc de la synthèse sonore	140
7.7	Bloc des mappings	142
7.7.1	Trois couches de mappings	142
7.7.2	Description graphique des mappings	143
7.7.3	Fonctions récurrentes	144
7.7.4	Expressions conditionnelles	145
7.8	Temporalité	146
7.8.1	Synchronicité	146
7.8.2	Paramètre temporel	146
7.8.3	Variables mémoires	147
7.9	Conclusion	148

8	Validation d’après cas d’études	149
8.1	Cas d’études de la thèse	149
8.1.1	Supervision du procédé de fabrication additive wire-arc	150
8.1.2	Sonification immersive du suivi de protéines individuelles	160
8.2	Cas d’études externes	166
8.2.1	Audification	166
8.2.2	MBSon : le sonogramme des données	168
8.3	Conclusion	175
9	Perspectives et conclusion	177
9.1	Vers un outil dédié	178
9.1.1	Vers une interface de design	178
9.1.2	Vers une traduction automatique du schéma	179
9.1.3	Prototypage sous ossia score	188
9.2	Conclusion	197
	Conclusion générale	199
9.3	Cas d’étude - Supervision de la fabrication additive	199
9.3.1	Contribution	200
9.3.2	Perspectives	200
9.4	Cas d’étude - Suivi des trajectoires individuelles de protéines	202
9.4.1	Contribution	202
9.4.2	Perspectives	202
9.5	Cas d’étude - Sons environnementaux vs. mesures physiques	203
9.5.1	Contribution	203
9.5.2	Perspectives	204
9.6	Modèle générique de la sonification	204
9.6.1	Contribution	204
9.6.2	Perspectives	206
9.6.3	Conclusion	207
A	Manuel utilisateur pour FluoSim-sonif	209
	Glossaire	219
	Acronyms	223
	Références	225
	Publications	247

Table des figures

1.1	De gauche à droite : un signal sinusoïdal de 1000 Hz généré pour une fréquence d'échantillonnage de 100 000 Hz, puis le même signal échantillonné à 10 000 Hz, puis le signal quantifié sur 3 bits, puis le signal échantillonné à 10 000 Hz et quantifié sur 3 bits.	16
1.2	Un exemple d'onde sinusoïdale pour une fréquence de 1000 Hz échantillonnée à 100 000 Hz	17
1.3	Un exemple d'onde en dents de scie pour une fréquence de 1000 Hz échantillonnée à 100 000 Hz	18
1.4	Un exemple d'onde carrée pour une fréquence de 1000 Hz échantillonnée à 100 000 Hz	18
1.5	Un exemple d'onde triangulaire pour une fréquence de 1000 Hz échantillonnée à 100 000 Hz	19
1.6	Un exemple de bruit blanc échantillonné à 100 000 Hz	19
1.7	De gauche à droite : capture sous <i>Audacity</i> du signal sonore correspondant à une note de piano (A6), puis son spectre calculé à l'aide des fonctionnalités de traitement du signal de <i>scipy.fft</i> , puis son spectrogramme calculé par <i>Audacity</i>	20
2.1	Le dispositif de fabrication wire-arc. À gauche photographié au repos. À droite schématisé avec un code couleur se rapportant aux critères à superviser.	38
2.2	Une capture d'écran de l'interface d'évaluation en ligne utilisée pour [242]	51
2.3	Capture d'écran de l'interface d'évaluation en ligne pour la sonification des critères de fabrication additive. Ici pour la représentation <i>Synth</i>	54
2.4	Histogramme des indices de sensibilité aux différents stimuli employés dans les trois écologies	62
2.5	Histogramme des temps de réaction aux différents stimuli employés dans les trois écologies	63

2.6	Diagrammes box-and-whiskers des vitesses moyennes de recopie des testeurs dans les trois conditions expérimentales	64
3.1	Une trame de vidéo représentant les molécules au sein d'une cellule	73
3.2	Maquette de l'outil de sonification immersive.	78
3.3	La sélection de la hauteur d'après la gamme diatonique majeure de la note la plus basse choisie, et jusqu'à la note la plus haute choisie. Ici entre La_4 et Mi_6 , et pour un taux de diffusivité $D = 0.58$	80
3.4	À gauche, la première version de l'interface de <i>FluoSim-sonif</i> . À droite la version réorganisée. Les contrôles bas niveau de la synthèse sonore et les fonctionnalités de débogage sont dans le bloc « Advanced Options », inactif par défaut.	83
4.1	Un schéma du dispositif de capture	98
4.2	Illustration de la correspondance dans le temps entre les mesures d'anémomètre à 50Hz et les fenêtres du son enregistré à 48 kHz	99
4.3	En haut : spectre des fréquences des vitesses u et w du vent, et des 6 pistes d'audio. En bas : spectre des nombres d'ondes avec zones d'intérêt en gris	101
4.4	Capture sous Audacity des mel-spectrogrammes de la première minute d'audio du 19 février 2021, avant (gauche) et après (droite) le pré-traitement	103
7.1	Une vue d'ensemble du modèle Cellule de Sonification	137
7.2	Détail du bloc des données d'entrée	138
7.3	Détail du bloc d'interaction utilisateur	139
7.4	Détail du bloc des paramètres réglables	140
7.5	Détail du bloc de synthèse sonore	141
7.6	Détail du bloc des mappings	142
8.1	La sonification <i>Mixte1</i> du chapitre 2 décrite selon le format du modèle	151
8.2	Détail de la fonction de synthèse donnant le complexe sonore <i>Mélodie</i>	155
8.3	Détail du synthétiseur <i>Default</i> impliqué dans le complexe <i>Mélodie</i>	156
8.4	Détail du complexe sonore <i>Bourdon</i>	157
8.5	Détail de la construction des complexes sonores <i>Ébullition</i> et <i>Craquèlement</i>	157
8.6	Détail de la construction du complexe sonore <i>Friture</i>	158
8.7	La sonification interactive <i>FluoSim-sonif</i> du chapitre 3 décrite selon le format du modèle	161
8.8	La projection du point \vec{x}_0 dans le repère orthonormé donné par le point d'origine \vec{x}_1 et l'axe $\vec{x}_2 - \vec{x}_1$. La sortie est alors le vecteur (x_{proj}, y_{proj})	164

8.9	Détail de la construction des complexes sonores individuels Son_i pour chaque point	165
8.10	Une représentation du principe de l'audification d'après le format du modèle	167
8.11	Représentation du sonogramme des données d'après le format du modèle	170
8.12	Représentation du complexe sonore $Modèle_i$ utilisé dans le sonogramme des données	171
8.13	Représentation de la fonction $Solution_i$ utilisée dans le complexe sonore $Modèle_i$	172
9.1	Un exemple simple pour le parcours des opérations	181
9.2	Les complexes sonores utilisés dans les exemples des figures 9.1 et 9.3	181
9.3	Un exemple simple d'expression auxiliaire	183
9.4	Un exemple simple d'opération initiale	185
9.5	Le complexe sonore employé dans les exemples des figures 9.4 et 9.7	185
9.6	Un exemple simple d'affectation de variables mémoires	186
9.7	Un exemple simple de factorisation des complexes sonores	188
9.8	Le module de lecture de CSV (à gauche) suivi de la distribution des dimensions des données. Détail de la reconstruction de <i>Mixte1</i> sous <i>ossia score</i>	190
9.9	Capture d'écran du projet <i>ossia score</i> entier pour la reconstruction de <i>Mixte1</i>	192
9.10	Capture d'écran du détail des mappings pour la reconstruction de <i>Mixte1</i> sous <i>ossia score</i>	193
9.11	Capture d'écran du projet <i>ossia score</i> entier pour la reconstruction simplifiée de <i>FluoSim-sonif</i>	195
9.12	Capture d'écran du détail des mappings pour la reconstruction simplifiée de <i>FluoSim-sonif</i> sous <i>ossia score</i>	196

Liste des tableaux

- 2.1 Les critères de fabrication du procédé wire-arc. 40
- 2.2 Les mappings pour la représentation *Mixte0*. 45
- 2.3 Les mappings pour la représentation *Mixte1*. 46
- 2.4 Les mappings pour la représentation *Synth*. 48
- 2.5 Les mappings pour la représentation *Nature*. 48
- 2.6 Résumé des réponses au sondage qualitatif 65

- 6.1 Étude des degrés de liberté offerts par divers outils généralistes de sonification 130

Résumé étendu en français

La sonification est la représentation sonore non verbale de données à des fins de communication et d'interprétation. Ce type de représentation, proposé en supplément ou remplacement de la vue, vise à mettre à profit le sens auditif et les spécificités de son rapport au temps et à l'espace pour des besoins d'exploration de données, de supervision de processus ou d'interaction avec une interface. Les catégories de techniques traditionnellement listées pour la sonification se rapportent à différents niveaux d'abstraction dans l'association de sons aux données : la translation directe signal-signal (audification), la traduction en paramètres perceptifs à différents niveaux de métaphores psychophysiques, naturelles ou musicales (PMSon, auditory icons, earcons), et la traduction en propriétés physiques au sein d'un modèle d'objet sonnant (MBSon).

La sonification ne découlant pas d'une longue histoire, comme celle de la visualisation, ayant pu naturellement mener à une connaissance extensive de l'interprétation de sons, elle ne bénéficie pas encore d'un cadre théorique « universel » guidant la conception de représentations auditives pour n'importe quels usages. L'objectif de la présente thèse est ainsi de contribuer à la définition d'un tel modèle. Nos réflexions ont été alimentées par nos travaux sur trois cas d'études, rapportés dans la thèse, ainsi qu'une étude plus générale de l'existant visant à détecter les points communs et degrés de liberté des diverses approches pour la sonification.

Le premier cas d'étude porte sur la supervision d'un procédé industriel à l'aide de bandes-sons paramétrées, le défi étant de concevoir des représentations non intrusives où tous les éléments sont suffisamment distincts et représentatifs des grandeurs associées pour que leur reconnaissance nécessite le moins d'entraînement et d'attention possible. Dans le chapitre dédié, nous décrivons les itérations successives de la représentation, puis le protocole d'évaluation développé pour les comparer, puis une étude préliminaire effectuée sur l'apport de l'incongruité dans la détectabilité des éléments sonores.

Le deuxième cas d'étude porte sur l'observation de grandes quantités de données évoluant dans le temps et dans l'espace, à des fins d'analyse d'imagerie à super-résolution. Dans le chapitre dédié, nous décrivons la solution d'exploration auditive proposée, puis son implémentation en tant que prototype.

Le troisième cas d'étude a pour but la recherche de nouvelles techniques pour la modélisation des réactions de l'environnement aux phénomènes climatiques, ici en passant par une analyse de la réponse acoustique d'une forêt au vent environnant. Le chapitre dédié rapporte nos études préliminaires sur la possibilité d'impliquer l'IA dans un processus de sonification par reconstitution de cette réponse acoustique en fonction de mesures de la vitesse du vent.

Notre étude des degrés de liberté de « l'objet sonification » (données à représenter, pré-traitements, synthèse sonore, paramètres variables de la synthèse, interactivité, réglages, rapport au temps, mappings) suggère une vision modulaire de ce procédé de représentation basée sur une compartimentation des apports de ces différentes variables. Le modèle que nous proposons consiste ainsi à répartir les sources de données d'entrée selon trois groupes (les valeurs à représenter, les paramètres réglables et les données d'interactions), puis à définir des modalités de rendu sonore et leurs variables, avant d'explicitier les opérations effectuant la liaison entre ces sources et ce rendu. On construit ainsi une expression de la sonification en tant que fonction de synthèse sonore paramétrée par le temps et par les sources de données. Nous posons ensuite les bases d'un langage graphique pour la représentation schématique de ces modalités de représentation, avant d'étudier l'évolution potentielle, à l'avenir, d'un tel modèle vers un outil informatique pour la conception d'outils de sonification.

Introduction

Selon les sources et les spécificités individuelles, le nombre de sens dont dispose l'être humain gravite typiquement autour de cinq [209]. Apparus au fil d'innombrables générations depuis de lointains ancêtres aquatiques, ces sens se sont développés comme une réponse évolutive au besoin de se repérer dans son environnement et d'en détecter les bénéfices ou dangers potentiels. Motivé par sa curiosité et sa pulsion à canaliser les forces de la nature pour ses propres usages, homo sapiens a commencé à se distinguer des autres animaux en usant de ses sens pour observer, corrélérer, comprendre et reproduire les phénomènes du monde physique.

Alors que les progrès constants de cette pratique communément appelée « la science » attirent notre attention vers des situations de plus en plus lointaines, abstraites, dangereuses, infiniment grandes ou petites, l'observation seule finit par ne plus suffire. C'est alors qu'entrent en jeu des appareils permettant la captation de grandeurs physiques représentatives des phénomènes étudiés.

En attendant que l'intelligence artificielle atteigne un niveau qui puisse rendre obsolète l'intervention humaine dans le processus d'analyse des données, il est alors nécessaire de les représenter, c'est-à-dire de les projeter vers un domaine sensoriel à nouveau accessible à un observateur.

La visualisation, qui fait ses preuves depuis l'invention du dessin, est la plus courante de ces représentations, mais des efforts de recherche existent aussi pour l'exploitation des autres canaux sensoriels. Si beaucoup des modalités non visuelles pour la transmission sensorielle d'informations, comme l'odorat ou le goût, ont encore certaines difficultés à sortir du domaine de la littérature d'anticipation, d'autres comme le toucher ou l'ouïe présentent de forts potentiels pour des applications concrètes à la représentation de l'information.

La représentation auditive en particulier offre de tirer parti des facultés propres à l'audition que sont la perception périphérique, constante et inconsciente des sons environnants, la capacité à reconnaître la nature et l'emplacement des sources sonores dans des environnements complexes (analogue à la capacité de la vue

à distinguer les contours des objets), ainsi que la sensibilité aux tonalités, aux rythmes, et même au temps en général, en dehors duquel le phénomène de son n'a pour ainsi dire aucune existence. La modalité auditive peut ainsi être particulièrement judicieuse à exploiter pour des systèmes présentant ce genre de propriétés, ou dans des situations où la vue est indisponible, que ce soit à cause d'un contexte d'observation inaccessible ou dans des cas de handicap.

La sonification

Le terme « sonification » employé dans le titre de cette thèse désigne alors la représentation auditive d'informations sous une forme spécifiquement non verbale. En effet si, techniquement, la parole est une représentation auditive, on s'intéresse ici à toutes les autres façons de communiquer des informations par des sons. Le verbe associé est « sonifier ».

Selon les besoins, les sons non verbaux utilisés peuvent consister en des assemblages de stimuli continus ou ponctuels, se rapportant à des phénomènes naturels ou abstraits, et dont les propriétés perceptives peuvent être paramétrées de diverses façons afin de traduire différents aspects d'intérêt des données. Cette paramétrisation des propriétés du son en fonction des données est centrale au principe de la sonification, et on la désignera dans cette thèse par le terme anglais *mapping*.

Émergence historique

Un survol de la préhistoire de la sonification – pour lequel on se référera à Schoon et Dombois, 2009 [192] et à Worrall, 2018 [231] – montre que l'expression sonore non verbale n'est pas complètement exclusive à l'ère de l'informatique, même si elle n'a pris une forme réellement exploitable que relativement récemment.

En effet, dans le domaine de la composition musicale, il a toujours existé une intention de transmettre au moins des émotions, sinon des scènes entières, par le biais de compositions instrumentales. On pense à l'exemple célèbre de la VI^{ème} symphonie de Beethoven, dite « Pastorale », dans laquelle les timbres des instruments et les différentes mélodies représentent une série de tableaux champêtres. Au siècle dernier, cette expressivité a laissé la place chez certains compositeurs comme Xenakis [42] à l'application objective de règles mathématiques dans le processus de création, une démarche qui s'est avérée annonciatrice du principe moderne de la sonification.

Un exemple souvent cité de « proto-sonification » dans un but non artistique est celui de l'expérience de la planche inclinée de Galilée [59], qui lui permettait de reconstituer par un rythme régulier l'accélération d'un corps due à la gravité.

On y voit un cas historique montrant que l'exploitation des spécificités de l'ouïe pour la recherche d'information dans des sons non verbaux avait déjà été explorée avant d'être facilitée par les technologies sonores modernes.

C'est cependant au cours des XIX^{ème} et XX^{ème} siècles que l'arrivée des technologies d'enregistrement et de traitement du signal, puis de l'informatique du son ont permis au domaine de la représentation auditive de l'information de prendre forme. Les premiers travaux effectuant les retranscriptions audibles de signaux naturels commencent à être rapportés, faisant d'abord usage de hauts-parleurs téléphoniques [17], puis de programmes informatiques [201, 72]. Le compteur Geiger, théorisé dès les années 1910 puis construit pour la première fois en 1928, est un exemple célèbre de capteur utilisant un retour sonore électronique pour signaler une anomalie à son utilisateur. Dans la première moitié du XX^{ème} siècle, des dispositifs électroniques commencent à être employés pour effectuer les premières expériences sur la perception auditive humaine [204, 40, 175].

La phase actuelle de la recherche en sonification résulte d'un fort gain d'intérêt pour le domaine à partir des années 1980, mis en mouvement par les progrès de l'informatique qui ont commencé à permettre l'interaction et le rendu sonore en temps réel. On observe à cette époque l'apparition des premiers outils sonores pour l'interaction utilisateur et l'exploration de données [146, 87, 79], de travaux sur la perception [235, 21] et de thèses portant sur la représentation auditive [22, 81].

Un certain nombre de conférences internationales encore actives sur l'informatique du son ont commencé à être organisées après les années 1980 (en plus d'ICMC, démarrée en 1975) : ICAD en 1992, SMC en 2004 et Audio Mostly en 2006. La publication en 2011 du *Sonification Handbook* [99] marque un nouveau tournant dans la recherche pour la sonification en organisant toutes les connaissances et problématiques contemporaines du domaine pour l'initiation de ses nouveaux participants.

Les outils nécessaires pour la pratique de la visualisation existent depuis des temps antédiluviens, ce qui a largement laissé à celle-ci le temps d'être pratiquée, codifiée et perfectionnée [63]. En comparaison, avant l'apparition relativement récente des technologies d'enregistrement du signal sonore, il n'existait pas de moyen pratique de transmettre une représentation auditive à moins de reconstruire soi-même le dispositif correspondant (une démarche qui a notamment été entreprise à l'époque moderne pour la planche inclinée de Galilée [183]) ou, dans les cas musicaux, de la faire jouer par des interprètes. Ce n'est qu'après l'apparition encore plus récente des technologies informatiques de communication, de rétention et d'automatisation du traitement des données qu'il est devenu réellement pratique de créer et transmettre des programmes de représentation auditive.

Ainsi, si la notion de représenter les informations sous une forme sonore non verbale n'est pas entièrement nouvelle, la synthèse sonore interactive n'est devenue une commodité que depuis une ou deux générations tout au plus, et la sonification n'a pas encore pu bénéficier des siècles de maturation qui permettraient son implication spontanée face à des besoins de représentation.

À ce manque de maturité s'ajoute la difficulté de centraliser la recherche sur ce domaine extrêmement pluridisciplinaire qui implique, comme listé par les auteurs du *Sonification Handbook* dans son chapitre d'introduction [104], des notions d'informatique, d'interface humain-machine, de sciences cognitives, de design sonore, de philosophie et de musicologie entre autres.

Question du cadre théorique

Il existe ainsi un besoin, exprimé par la littérature depuis les années 1990 [128, 70], d'un modèle théorique pour encadrer la conception de sonifications, afin de réduire le temps d'étude préliminaire nécessaire à chaque nouvelle application ad hoc de cette technique de représentation. À l'instar de la sonification elle-même, la recherche d'un tel modèle peut être abordée du point de vue d'un grand nombre de domaines disciplinaires, ce qui mène à différentes approches pour tenter de répondre au besoin d'encadrement. Les trois approches les plus courantes sont résumées ci-dessous.

Problématique du mapping Une des grandes problématiques de la recherche en sonification est le *mapping problem*, formulé par John Flowers en 2005 [70] et nommé ainsi a posteriori par les auteurs qui l'ont cité. Cette problématique part du constat initial qu'il est rarement judicieux de compter sur une émergence naturelle des motifs sous-jacents dans un système de données en représentant tous ses éléments d'un seul coup, d'une seule et même façon. La recherche d'une modélisation se situe alors au niveau des sciences cognitives et consiste à établir un ensemble de règles à appliquer lors de la construction de mappings pour une intelligibilité optimale des représentations auditives.

Outil informatique La sonification ayant pris sa forme actuelle à l'ère de l'informatique, une approche valide peut être de produire une modélisation de la sonification en tant que programme informatique. La finalité est alors de construire un outil de conception spécifique à la sonification mais encore suffisamment générique pour permettre un prototypage souple sur les choix de représentation. Cet axe de recherche rejoint la problématique du mapping dans le sens qu'il doit pouvoir faciliter l'expérimentation pour la construction de modèles cognitifs [217, 70].

Transmission du design Cet axe de recherche étudie les techniques de sonification en tant que solutions à des besoins, dans des contextes plus ou moins spécifiques. La recherche s'intéresse alors au développement de techniques de structuration, organisation et transmission de la connaissance mettant en évidence la corrélation entre le contexte qui motive une représentation auditive et l'ensemble des techniques appliquées pour la créer. La transmission facilitée de la connaissance contribue alors à faire émerger les techniques les plus judicieuses au fur et à mesure de leurs itérations [15, 75].

Le but de cette thèse est de contribuer à la description d'un tel modèle selon une approche qui se situe entre les problématiques de design et d'informatique : le modèle doit à la fois permettre de décrire précisément une solution de sonification afin d'offrir un cadre à la reproductibilité, et être propice à la création d'un outil informatique. Un certain nombre de bonnes pratiques déjà suggérées par la recherche en psychoacoustique ont été prises en compte, mais nous avons tâché de ne pas trop limiter les possibilités offertes par le modèle afin de laisser une bonne part de liberté pour l'expérimentation.

Démarche

Notre démarche pour arriver à ce but a consisté à travailler sur un échantillon de trois cas d'études inédits afin d'orienter nos réflexions sur le principe de la sonification et de produire nos propres designs. En parallèle, une étude poussée de l'existant a été effectuée afin de détecter les aspects récurrents et variables de la conception de sonifications, ainsi que les tentatives existantes pour modéliser ces techniques de représentation. Ces réflexions ont alors été synthétisées pour proposer un modèle structurant la description de toute solution de sonification.

Structure de la thèse

Le chapitre 1 donne pour commencer une brève mise en contexte sur le principe de la sonification, la synthèse sonore, la perception auditive et les catégories de techniques traditionnellement listées pour répondre à différents besoins de représentation.

Les trois cas d'études réalisés sont ensuite rapportés dans la partie II : la supervision auditive du procédé industriel de fabrication additive dans le chapitre 2, l'analyse de trajectoires individuelles de protéines dans le chapitre 3 et la corrélation de mesures physiques avec les sons d'un environnement dans le chapitre 4.

Dans la partie **III**, nous rapportons nos études et réflexions plus générales sur la description d'un modèle générique de la sonification. Notre étude des aspects variables ou constants de la sonification sous ses différentes formes, détaillée dans le chapitre **5**, offre un cadre pour ensuite effectuer dans le chapitre **6** une étude des tentatives existantes de modéliser la sonification d'après différentes disciplines, qui inclut notamment une interrogation sur la souplesse offerte par les outils de sonification générique existants.

Les considérations les plus prometteuses suggérées par cette étude de l'existant sont reprises dans le modèle que nous décrivons dans le chapitre **7**, nommé « Cellule de Sonification ». Une validation partielle du modèle et un rapprochement avec les cas d'études de la thèse sont effectués dans le chapitre **8** en réécrivant sous le format du modèle les outils développés.

La partie **III** est ensuite conclue par une réflexion, dans le chapitre **9**, sur la possibilité d'utiliser le modèle proposé comme un cadre théorique pour la création d'un outil informatique de conception de sonifications.

Première partie

Contexte

Chapitre 1

La sonification

Dans ce premier chapitre, nous offrons une mise en contexte abrégée sur le principe de la sonification. Dans la section 1.1 nous commençons ainsi par donner un certain nombre de définitions clarifiant ce qu'est – et n'est pas – la sonification. Ensuite nous donnons dans la section 1.2 une liste des aspects de la perception humaine et des applications pratiques pouvant motiver la mise en œuvre d'une sonification.

La création de sons numériques paramétrés joue un rôle très important dans le domaine de la sonification. Pour cette raison, nous donnons dans la section 1.3 le contexte nécessaire pour la compréhension des principales techniques de synthèse sonore, ainsi que des outils et langages informatiques typiquement mis en œuvre pour programmer des sons paramétrés.

Dans la section 1.4, nous décrivons les grandes catégories de sonification traditionnellement listées par la littérature, puis dans la section 1.3.6 nous listons les paramètres perceptifs les plus couramment utilisés pour la communication d'informations.

Enfin, nous indiquons dans la dernière section 1.5 comment la présente thèse s'inscrit dans le domaine de recherche pour la sonification en visant à décrire un modèle générique de design du mapping et de l'interaction pour la représentation auditive.

1.1 Définitions

La *sonification* est une famille de techniques de représentation de données s'inscrivant dans le domaine plus large de la *représentation auditive*.

La *représentation auditive* englobe tous les moyens pouvant être mis en œuvre pour communiquer des informations par le biais d'un signal sonore. Cela inclut tous les niveaux de la chaîne de communication employés pour l'interaction avec

l'utilisateur, pour la collecte, le traitement et la transformation de données, puis pour la diffusion du son résultant.

Selon que l'on se réfère au premier ou au deuxième chapitre du *Sonification Handbook*, la **sonification** se distingue alors de la représentation auditive soit en tant que famille de procédés informatiques plus spécifiquement chargés de faire correspondre des propriétés d'un ensemble de données à des propriétés d'un son numérique [104], soit comme le cadre plus spécifique de la représentation auditive dans lequel le son utilisé est non verbal [222].

La définition souvent considérée comme fondatrice pour la sonification est celle donnée par la Communauté Internationale pour la Représentation Auditive (*International Community for Auditory Display*, ou **ICAD**, en anglais) en 1999 dans le *Sonification Report* [128] :

Sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation.

Ou, en français : « La sonification est la transformation de relations entre des données en relations perçues au sein d'un signal acoustique dans le but de faciliter la communication ou l'interprétation. »

Une définition ultérieure de la sonification fréquemment citée est la révision formulée par Thomas Hermann dans son article *Taxonomy and definitions for sonification and auditory display* de 2008 [97] puis formalisée en une seule phrase sur le site web sonification.de [99] :

Sonification is the data-dependent generation of sound, if the transformation is systematic, objective and reproducible, so that it can be used as scientific method.

En français : « La sonification est la génération de son en fonction de données, sous réserve que la transformation soit systématique, objective et reproductible, de sorte à pouvoir être utilisée comme une méthode scientifique. »

Cette définition plus stricte vise à imposer un caractère scientifique à la sonification en excluant certaines pratiques de composition musicale qui pourraient consister à interpréter subjectivement un concept ou un ensemble de données pour les exprimer en musique. L'utilisation de la sonification à des fins artistiques n'est pas exclue, mais celle-ci doit être construite selon des règles explicites et systématiquement appliquées.

Un aspect important de la définition de la sonification est le fait que les sons utilisés doivent être non verbaux. En effet, si le domaine de la représentation auditive admet parfois la communication d'informations par l'utilisation d'une voix synthétique ou d'une collection de messages vocaux pré-enregistrés, la sonification

traite de toutes les autres façons de communiquer des messages par le biais de métaphores auditives et de modulations de paramètres perceptifs.

Le but est de pouvoir employer des sons plus courts pour communiquer des messages plus simples. Pour un exemple minimal, au lieu d'une voix récitant « la température est de trente-cinq degrés », on pourrait émettre un bref « bip » dénotant le fait que la température a dépassé un seuil de 35 degrés.

1.2 Motivations

1.2.1 Avantages de l'audition

Écoute constante et passive L'ouïe est un sens constamment actif, y compris dans une certaine mesure pendant le sommeil. Une image assez frappante souvent citée pour parler de cette particularité, surtout comparée à la vue, est que l'être humain n'a ni le besoin ni la capacité de « fermer les oreilles ». Dans ce contexte d'apport constant d'information, il n'est pourtant pas constamment nécessaire de comprendre et gérer tous les stimuli présents. Cependant il est alors possible d'être interpellé par certains sons incongrus ou changements dans les propriétés du son sans y avoir été activement attentif précédemment, et de choisir d'accorder son attention à la source de ces événements.

Distinction de flux sonores Sous certaines conditions de continuité et de cohérence des paramètres perceptifs, il est en effet possible de distinguer l'une ou l'autre composante d'un signal acoustique comme constituant un flux, une entité correspondant à une action donnée sur un objet donné.

Cet effet perceptif et cognitif, souvent appelé *cocktail party effect* pour son application dans la capacité à se concentrer sur différentes conversations croisées, a été nommé et étudié par E. C. Cherry dans les années 1950 [40]. Le livre *Auditory Scene Analysis* d'Albert Bregman, publié en 1994, répertorie plus généralement les capacités et limites de l'audition humaine pour la distinction de flux sonores non verbaux [25].

Détection spatiale Sans avoir besoin de tourner la tête vers une source sonore, il est possible non seulement de l'entendre, mais encore d'avoir une estimation de sa position. L'audition est ainsi la seule modalité sensorielle qui permette à l'être humain d'être en relation avec la totalité de l'espace qui l'entoure, la disposition des autres organes sensoriels sur le corps ne permettant que la captation des stimuli dans un seul plan.

Temporalité intrinsèque La perception du son est indissociable de la perception du temps : le son que l'on perçoit résulte de la propagation dans l'air de variations de la pression au cours du temps. La temporalité dans le son peut se traduire par des continuités, des transitions, des discontinuités ou, en cas de répétitions, des périodes perçues comme des tonalités ou des rythmes. Or l'oreille humaine est assez efficace pour détecter ces types de changements. La temporalité d'un son synthétisé peut ainsi constituer une modalité très puissante pour communiquer des relations au sein de systèmes de données, que ce soit pour exprimer l'évolution des valeurs au cours du temps ou l'organisation de points de données les uns par rapport aux autres.

1.2.2 Applications

Supervision de processus Comme mentionné plus haut, la temporalité du son en fait un candidat idéal pour représenter l'évolution d'un système au cours du temps. Certains événements dans les données peuvent alors être associés à différents types d'alarmes ponctuelles, tandis que des variations au cours du temps peuvent aussi être représentées par l'évolution de paramètres perceptifs dans des flux sonores constamment actifs. Après avoir longtemps consisté en une stratégie de notification à tout prix usant d'alarmes fortes et aiguës pour attirer l'attention sur des anomalies [166], la sonification pour la supervision fait depuis les années 90 l'objet de recherches visant à la rendre plus esthétiquement recherchée et moins stressante [82, 180, 156]. Le chapitre 2 de cette thèse traite d'un cas d'étude pour la supervision auditive d'un procédé industriel, pour lequel un tel système de notifications sonores ambiantes a été mis en place.

Exploration de données Avec les progrès de la performance des ordinateurs et des moyens d'interaction humain-machine au cours des quelques dernières décennies, il est devenu possible de développer de nouvelles formes de sonifications plus complexes et plus réactives, permettant le contrôle en temps réel de la synthèse sonore par l'interaction virtuelle de l'utilisateur avec l'espace des données. Le but n'est plus alors de notifier l'auditeur sur l'état du système de données ou de l'alerter sur des événements qui s'y produisent, mais de permettre un repérage dans leur organisation spatiale, souvent par une transposition de cette spatialité dans un son synthétisé en stéréo. C'est l'une des principales motivations de la famille de techniques de sonification dite « model-based », présentée par Hermann en 1999 [107] et résumée dans la section 1.4. Le chapitre 3 de cette thèse traite d'un cas d'étude de sonification pour une exploration de ce type, à la fois spatiale et temporelle pour des points en mouvement dans un espace 2D.

Aide aux gestes La sensibilité de l’ouïe aux rythmes en fait une modalité intéressante à explorer pour l’entraînement à des gestes répétés, par exemple dans des contextes sportifs [26]. En faisant usage de principes intuitifs de gestuelle musicale, il est possible de produire des sonifications interactives pour l’assistance à divers gestes ou postures [114].

Vulgarisation scientifique Sous réserve que la nature réelle des données et leur rapport au son soient bien présentés, la sonification peut être un bon moyen de présenter un domaine de recherche au public sous une forme ludique, esthétique, et qui encourage l’interprétation et la discussion.

Arts La représentation auditive de données n’est pas toujours restreinte à un rôle purement utilitaire. En effet, l’observation de relations mathématiques ou de données tirées de la nature a historiquement eu une certaine tendance à entrer en jeu dans la composition musicale [150, 192], sans aucun autre but que l’expérimentation artistique.

Une représentation artistique peut aider à ajouter une dimension émotionnelle à la présentation d’un ensemble de données pour la meilleure communication d’un message, ce qui rapproche cette motivation de celle déjà évoquée pour la vulgarisation [111]. On pense par exemple aux nombreux travaux de sonifications musicales du changement climatique [78, 3, 88, 144, 169, 198, 212].

Accessibilité Dans les situations où la vue n’est pas disponible, le premier sens qui permet de se repérer dans son environnement est l’ouïe. Ainsi l’un des grands enjeux de la sonification est la conception d’interfaces auditives pour le public déficient visuel [64].

1.3 Synthèse sonore et perception

La sonification est un domaine fortement pluridisciplinaire qui nécessite, entre autres notions de design et d’interface homme-machine, de posséder des connaissances d’informatique pour le son et de tenir compte du potentiel et des limites de la perception auditive humaine.

Dans cette section, nous décrivons brièvement en quoi consiste un signal sonore numérique, quelles techniques sont habituellement appliquées pour le synthétiser et quels outils sont actuellement disponibles pour programmer des sons synthétiques, en particulier pour une paramétrisation en temps réel, avant de lister les paramètres perceptifs les plus courants pour la transmission d’informations par le son.

1.3.1 Principes généraux

Le son Le mot « son » désigne à la fois un percept auditif et le phénomène physique qui lui a donné naissance. Le phénomène physique est une perturbation propagée dans un milieu élastique (en général, l'air). Il est produit par un mouvement matériel vibratoire qui modifie la pression locale du milieu. Ainsi, le son ne se propage pas dans le vide. Il engendre un percept auditif si l'onde de pression diffusée atteint le tympan de l'auditeur et fait vibrer cette membrane avec une amplitude suffisante [54].

Une onde de forme quelconque peut être décrite comme une somme de variations sinusoïdales qui diffère par la fréquence et l'amplitude. Cette description a une pertinence pour le système auditif humain qui réagit seulement aux fréquences comprises entre environ 20 Hz et 20 000 Hz, la valeur maximale étant variable selon l'âge et les dégradations potentiellement déjà subies par le système auditif. Du point de vue perceptif, un son stable se caractérise par des attributs distincts : hauteur tonale, sonie, timbre et durée.

Le signal sonore numérique est alors une approximation par discrétisation de ces fluctuations de la pression d'air, d'une part dans le temps (échantillonnage), d'autre part dans la quantité de valeurs qui peuvent être atteintes (quantification).

Échantillonnage L'échantillonnage consiste à mesurer un plus ou moins grand nombre de points du signal sonore au cours du temps. Cette quantité est donnée par une grandeur appelée la fréquence d'échantillonnage, exprimée en Hertz (Hz). Il s'agit du nombre d'échantillons constituant une seconde de signal, donc de l'inverse du laps de temps séparant chaque échantillon sonore. Selon le théorème de Nyquist-Shannon, une fréquence d'échantillonnage donnée F_e permet de représenter des fréquences pouvant aller jusqu'à $F_e/2$. Pour l'audition humaine, qui peut aller jusqu'à une fréquence de 20 kHz, il faut donc une fréquence d'échantillonnage d'au moins 40 kHz. Le format habituellement utilisé par le CD, par exemple, a un taux d'échantillonnage de 44 100 Hz.

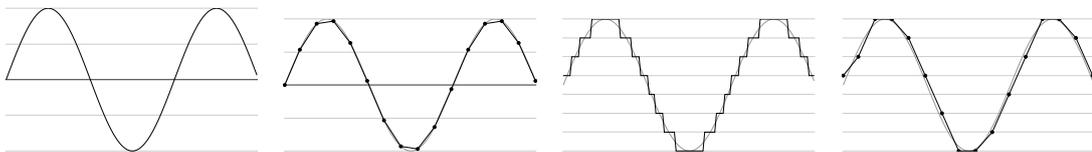


FIGURE 1.1 – De gauche à droite : un signal sinusoïdal de 1000 Hz généré pour une fréquence d'échantillonnage de 100 000 Hz, puis le même signal échantillonné à 10 000 Hz, puis le signal quantifié sur 3 bits, puis le signal échantillonné à 10 000 Hz et quantifié sur 3 bits.

Quantification Pour la conversion analogique-numérique du son, la quantification consiste à décider de la quantité de bits attribués à chacun des points du signal. Cette quantité de bits donne un certain nombre de valeurs que pourra prendre le point : pour n bits, la quantification présente 2^n niveaux. Un nombre de bits plus élevé permet un plus grand niveau de détail et une meilleure qualité sonore, mais il est à choisir prudemment pour ne pas surcharger inutilement le flux de données. La quantification habituelle pour le format **CD**, par exemple, est de 16 bits, soit 65 536 niveaux.

1.3.2 Oscillateurs

Cette sous-section décrit quelques-unes des fonctions périodiques les plus souvent utilisées comme oscillateurs en synthèse sonore. Ceux-ci entrent en jeu dans les différentes techniques de synthèse sonore souvent employées pour la sonification, décrites par la suite.

Onde sinusoïdale Souvent considérée comme la brique de base de la synthèse sonore (voir la sous-section 1.3.3 sur la transformée de Fourier), cette onde est obtenue à partir d'une discrétisation de la fonction *sin*. La figure 1.2 représente un exemple de ce type de signal. Pour une fréquence f , une fréquence d'échantillonnage Fe et un entier t , la valeur du signal sinusoïdal de fréquence f au pas de temps numéro t est calculée selon l'opération 1.1 :

$$\text{Sinus}(f)(t) = \sin\left(2 * \pi * f * \frac{t}{Fe}\right) \quad (1.1)$$

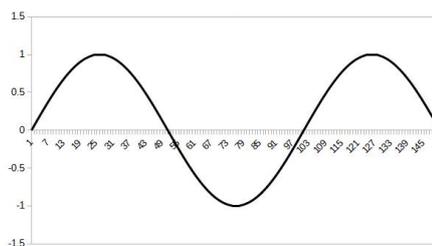


FIGURE 1.2 – Un exemple d'onde sinusoïdale pour une fréquence de 1000 Hz échantillonnée à 100 000 Hz

Onde en dents de scie Ce signal consiste en une répétition de rampes montantes et de chutes abruptes, formant une onde dont la forme rappelle les dents d'une scie. La figure 1.3 représente un exemple de ce type de signal. Pour

une fréquence f , une fréquence d'échantillonnage Fe et un entier t , la valeur du signal en dents de scie de fréquence f au pas de temps numéro t est calculée selon l'opération 1.2 :

$$Scie(f)(t) = 2 \left(f * \frac{t}{Fe} - \left\lfloor \frac{1}{2} + f * \frac{t}{Fe} \right\rfloor \right) \quad (1.2)$$

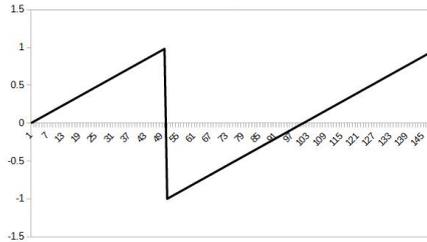


FIGURE 1.3 – Un exemple d'onde en dents de scie pour une fréquence de 1000 Hz échantillonnée à 100 000 Hz

Onde carrée Ce signal alterne instantanément entre deux niveaux, formant une onde en créneaux carrés. La figure 1.4 représente un exemple de ce type de signal. Pour une fréquence f , une fréquence d'échantillonnage Fe et un entier t , la valeur du signal carré de fréquence f au pas de temps numéro t est calculée selon l'opération 1.3 :

$$Carrée(f)(t) = 2 \left(2 \left\lfloor f * \frac{t}{Fe} \right\rfloor - \left\lfloor 2 * f * \frac{t}{Fe} \right\rfloor \right) + 1 \quad (1.3)$$

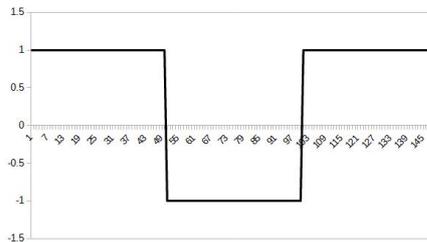


FIGURE 1.4 – Un exemple d'onde carrée pour une fréquence de 1000 Hz échantillonnée à 100 000 Hz

Onde triangulaire Ce signal consiste en une répétition de rampes montantes et descendantes symétriques. La figure 1.5 représente un exemple de ce type de signal. Pour une fréquence f , une fréquence d'échantillonnage Fe et un entier t , la valeur du signal triangulaire de fréquence f au pas de temps numéro t est calculée selon l'opération 1.4 :

$$Triangle(f)(t) = 2 * \left| 2 \left(f * \frac{t}{Fe} - \left\lfloor f * \frac{t}{Fe} + \frac{1}{2} \right\rfloor \right) \right| - 1 \quad (1.4)$$

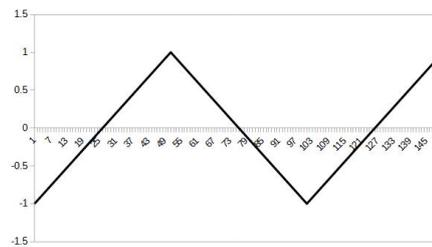


FIGURE 1.5 – Un exemple d'onde triangulaire pour une fréquence de 1000 Hz échantillonnée à 100 000 Hz

Bruit Un signal sonore généré aléatoirement. La figure 1.6 représente un exemple de ce type de signal. Dans le domaine fréquentiel (voir sous-section 1.3.3), il est possible de produire des bruits correspondant à différentes lois de probabilité, ou de filtrer un bruit blanc pour le concentrer dans différentes plages de fréquences, présentant alors différentes propriétés perceptives : bruits blanc, rose, brownien.

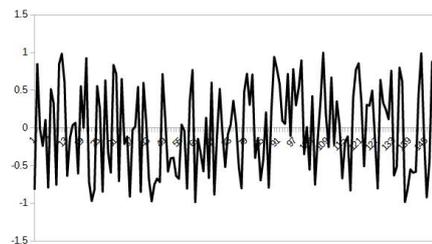


FIGURE 1.6 – Un exemple de bruit blanc échantillonné à 100 000 Hz

1.3.3 Transformation de Fourier

Spectre fréquentiel D'après l'application de la transformation de Fourier à un signal discret s de N échantillons, il est possible de l'exprimer sous la forme d'une somme finie de sinusoides, de sorte qu'à tout instant t :

$$s(t) = \sum_{k=0}^{N-1} A_k \sin(2\pi f_k t + \phi_k) \quad (1.5)$$

L'ensemble des couples de fréquences et amplitudes (f_k, A_k) constitue alors le **spectre** du signal, qui est une représentation indépendante du temps permettant d'observer la teneur en fréquences du signal et d'effectuer des opérations dessus. Cette transformation peut ensuite être inversée afin d'obtenir un signal temporel à nouveau.

Variantes Afin de rapprocher cette représentation de la perception humaine des hauteurs, l'échelle du spectre peut prendre différentes formes.

La sensibilité de l'oreille étant plus élevée pour des fréquences relativement basses (en dessous de 2 500 Hz), l'échelle logarithmique est parfois utilisée pour donner plus d'importance à cette région du spectre.

L'échelle Mel, une forme plus spécifique de l'échelle logarithmique, a été déterminée expérimentalement par Stevens et al. en 1937 [204] pour correspondre à la sensibilité humaine aux écarts de hauteurs dans une échelle musicale. Elle est souvent utilisée pour conditionner des réseaux de neurones pour la synthèse vocale (voir chapitre 4).

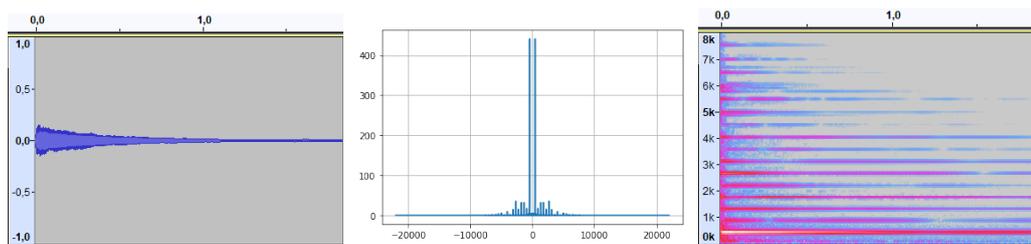


FIGURE 1.7 – De gauche à droite : capture sous *Audacity* du signal sonore correspondant à une note de piano (A6), puis son spectre calculé à l'aide des fonctionnalités de traitement du signal de *scipy.fft*, puis son spectrogramme calculé par *Audacity*

Sonogramme Le sonogramme est une adaptation temporelle du principe du spectre. Au lieu d'effectuer l'analyse du signal entier, on le découpe en fenêtres d'un nombre relativement petit d'échantillons qui sont analysées individuellement. La représentation se fait alors avec l'axe du temps en abscisse, l'axe des fréquences en ordonnées, et un code couleur pour l'amplitude correspondant à chaque fréquence. Cette forme permet d'obtenir une représentation de l'évolution de la teneur spectrale du signal au cours du temps. On parle aussi souvent de « **spectrogramme** », par rapprochement avec le terme employé en anglais.

Filtres Une opération de filtrage permet d'isoler ou de supprimer l'une ou l'autre partie du spectre. Un filtre passe-haut consiste ainsi à préserver les fréquences au dessus d'un seuil donné et un filtre passe-bas consiste à les préserver en dessous de ce seuil. Un filtre passe ou coupe bande consiste à préserver ou supprimer les fréquences dans une bande de fréquences déterminée soit par ses seuils haut et bas, soit par une fréquence centrale et un facteur de qualité donnant la largeur de la bande d'intérêt.

1.3.4 Techniques de synthèse sonore

Additive Une synthèse sonore de type additif consiste à additionner des oscillateurs basiques à différentes intensités, fréquences et phases pour obtenir des signaux complexes.

Soustractive À partir d'un signal déjà formé (souvent de timbre riche, par exemple un bruit ou une onde en dents de scie), on applique des opérations de filtrage ou des convolutions pour sculpter le signal.

Modulation (AM/FM) Cette technique consiste à utiliser des oscillateurs pour paramétrer l'amplitude (AM) ou la fréquence (FM) d'autres oscillateurs, afin de produire des timbres riches à partir de nombres réduits d'opérations [43].

Tables d'ondes Cette technique de synthèse consiste à stocker l'information d'une forme d'onde dans un tableau et la lire selon différents taux d'échantillonnage pour obtenir différentes hauteurs. Son avantage est son faible coût en temps de calcul, bien que la qualité du son obtenu soit souvent amoindrie.

1.3.5 Outils pour la synthèse sonore

Dans cette section, nous décrivons brièvement les outils et langages de programmation actuellement disponibles pour la synthèse sonore : *CSound*,

Max/MSP, *Pure Data*, *SuperCollider*, *ChucK*, *FAUST* et *ossia score*. Pour les besoins de cette thèse sur la sonification, nous nous intéressons en particulier à ceux qui permettent la paramétrisation de sons synthétiques en temps réel, et nous nous interrogeons sur l'apport de chacun, existant ou potentiel, pour des tâches de sonification.

CSound *CSound*¹ est un langage de programmation pour la musique électronique développé en 1985 d'après le modèle des premiers langages de synthèse sonore des années 50 (série *MUSIC*). Il permet de programmer des instruments synthétiques et de les mettre en œuvre dans des partitions musicales. Initialement plutôt prévu pour la programmation de partitions statiques, le langage a depuis été adapté pour l'interaction en temps réel, l'interface avec les protocoles **MIDI** ou **OSC**, ou en tant qu'**API** pour des programmes écrits dans d'autres langages.

Ce langage peut être utilisé pour une application de sonification à condition de créer, dans un autre langage, un programme s'occupant d'accueillir les données acquises et d'effectuer leur traduction en paramètres musicaux, de sorte à les communiquer ensuite à un module en *CSound* pour contrôler la synthèse sonore.

Max/MSP et Pure Data *Max/MSP*² est un logiciel propriétaire de synthèse sonore et musicale développé par Miller Puckette et distribué pour la première fois en 1990. Pour la programmation de **DSP**, cet outil propose un langage visuel consistant à brancher les uns aux autres des blocs d'opération qui gèrent des signaux audio et des paramètres. On parle de *patching*. *Pure Data*³ est une alternative libre de *Max/MSP* distribuée gratuitement en ligne depuis 1996.

Ces deux logiciels font l'objet de contributions régulières par leurs communautés de chercheurs et développeurs, et sont souvent au centre de la conception d'œuvres musicales électroniques et expérimentales. Dans les articles présentés à l'occasion de la conférence internationale *Sound and Music Computing* de 2022, par exemple, on en dénombre ainsi sept utilisant *Max/MSP* [207, 4, 154, 160, 188, 55, 177] et six utilisant *Pure Data* [194, 225, 195, 159, 179, 76].

De par leur prise en main relativement facile et leur flexibilité pour l'interaction avec du matériel externe, ces outils sont souvent utilisés dans des travaux de sonification. En 2004, Pauletto et al. ont tenté de formaliser ce procédé en proposant un outil de sonification interactive pour *Pure Data* [167].

1. <https://csound.com> – visité le 4/1/23
2. <https://cycling74.com/> – visité le 4/1/23
3. <https://puredata.info/> – visité le 4/1/23

SuperCollider *SuperCollider*⁴ est une plateforme de programmation pour la synthèse sonore publiée par James McCartney en 1996 et libre depuis 2002. Il s'agit d'un ensemble formé par le serveur audio *scsynth* qui permet la synthèse sonore en temps réel, le langage *sclang* qui permet la définition de générateurs de signal audio et la construction de structures musicales, et l'environnement de développement intégré *scide*.

Il est relativement facile de prendre en main les bases du langage pour rapidement commencer à produire des prototypes, mais les fonctionnalités plus avancées et la syntaxe fonctionnelle peuvent aussi vite poser problème aux programmeurs novices. Par ailleurs, *SuperCollider* n'offre pas beaucoup de flexibilité en termes de portabilité. En effet, s'il existe bien des options pour produire des applications *SuperCollider* pouvant fonctionner sans installer la plateforme, elles ne semblent fonctionner correctement que sous MacOS X actuellement.

On note cependant à quel point ce langage peut faciliter la sonification : le chapitre 13 du *SuperCollider Book* est entièrement dédié à l'utilisation de *SuperCollider* pour la sonification [52], pour un grand nombre des techniques usuelles (voir section 1.4) incluant l'audification, la **PMSon** et diverses formes d'interactivité.

ChuckK *ChuckK*⁵ est un langage de programmation pour la synthèse sonore développé depuis 2002 par Perry Cook et Ge Wang. Sa particularité est de permettre au programmeur d'indiquer explicitement l'écoulement du temps dans le code, pour un contrôle à la fois précis et flexible de la temporalité.

Ce langage permet l'interface avec les formats **OSC** et **MIDI**, donc la paramétrisation en temps réel de la synthèse sonore. Il est ainsi souvent utilisé comme moteur de rendu audio pour la sonification [123, 229, 164, 89, 200, 36]. Cette utilisation de *ChuckK* pour la sonification a d'ailleurs fait l'objet d'un cours à la conférence **ICAD** de 2012 [45].

FAUST *FAUST (Functional Audio Stream)*⁶ est un langage de programmation fonctionnel pour la création de **DSP**, développé au sein du Générateur de Ressources et d'Activités Musicales Exploratoires (GRAME) de Lyon.

Sa syntaxe, suffisamment simple pour une prise en main rapide, est axée sur la création d'embranchements pour le traitement de signaux audio. L'environnement de développement sur navigateur donne accès à une boîte à outils extensive de générateurs et d'effets. Un grand avantage de ce langage est la compilation

4. <https://supercollider.github.io/> – visité le 4/1/23

5. <https://chuck.stanford.edu/> – visité le 4/1/23

6. <https://faust.grame.fr/> – visité le 4/1/23

sous de nombreux formats permettant l'intégration dans des projets en C, C++, WebAssembly ou Rust, entre autres.

FAUST n'est pas directement utilisable pour la sonification, mais peut être utile pour rapidement produire des outils de synthèse audio paramétrables en temps réel, qui peuvent ensuite être intégrés dans des applications effectuant l'acquisition des données et leur traduction en paramètres.

ossia score *ossia score*⁷ est un logiciel libre de séquençage intermédia axé sur la création de partitions interactives. Il permet de construire des séquences musicales écrites en dur et de conditionner leur enchaînement en fonction de variables et d'évènements. De par sa forte inter-opérabilité (notamment avec *FAUST* et *Pure Data* mentionnés plus haut) et son support pour de nombreux formats de communications comme **OSC** et **MIDI**, l'outil vise à offrir une plateforme unique pour la réalisation d'applications de synthèse musicale interactive. Le langage visuel proposé par l'interface permet de créer des intervalles temporels pouvant s'enchaîner et interagir de diverses façons, et contenant des sections d'audio, des courbes d'automatisation, ou des patches de synthèse et traitement du son à la façon de *Max/MSP* ou *Pure Data*. Ces patches peuvent eux-mêmes impliquer des programmes écrits en C++ ou en Javascript.

Si *ossia score* est actuellement surtout utilisé pour des œuvres purement artistiques (voir la page *Gallery* du site officiel) et impliquant peu ou pas la sonification, toutes les fonctionnalités nécessaires y sont en place pour que cet outil puisse servir à prototyper des représentations auditives de données, du branchement d'un module externe pour l'acquisition des données à la production de son en temps réel en fonction de ces données.

1.3.6 Paramètres perceptifs

La plupart des techniques de sonification consistent à faire en sorte qu'un certain nombre de caractéristiques perceptives du signal soient contrôlées dans la synthèse du son en fonction de l'évolution ou l'organisation des données. Dans cette section, nous listons les plus courants de ces paramètres. Ceux-ci peuvent se situer à différents niveaux de la constitution du son : proche du signal (hauteur, sonie, certains paramètres donnant le timbre), ou plutôt proche de la façon dont le son se présente à un auditeur (la spatialisation, par exemple). D'autres paramètres, à un niveau encore supérieur, sont ceux qui entrent en jeu dans la construction d'une musique (mélodies, registres, timbres d'instruments), ou l'utilisation d'une métaphore sonore tirée du monde réel (propriétés du couple objet-action produisant le son).

7. <https://ossia.io/> – visité le 4/1/23

Sonie

L'intensité perçue, ou **sonie**, est une propriété perceptive donnée par l'amplitude du signal. Le plus souvent cette propriété est contrôlée par un facteur de multiplication du signal, situé entre 0 et 1 pour l'atténuer ou au dessus de 1 pour l'amplifier.

Cette propriété est assez polyvalente pour l'association métaphorique à des grandeurs physiques variées, à des fins de représentation auditive (Walker et Kramer, 2005 [218]). Plus généralement, un paramètre de sonie est souvent utilisé pour nuancer l'émergence, l'importance ou l'urgence d'une notification. Cet usage peut être dû à un rapprochement perceptif de la sonie avec l'idée de proximité, donc d'urgence, d'une notification. Ho et al. [112] ont en effet montré par l'expérience que des notifications semblant se produire plus près de l'utilisateur permettaient des réactions plus efficaces.

Hauteur tonale

La **hauteur tonale** est la propriété d'un son harmonique (c'est-à-dire un son dont les composantes spectrales de plus forte énergie ont un diviseur commun) qui correspond à sa perception comme plus ou moins aigu, ou grave.

Cette propriété est donnée par la fréquence fondamentale, qui est la fréquence la plus basse de la décomposition spectrale d'un son harmonique. Dans le cas particulier de la fondamentale manquante, si une quantité suffisamment grande de fréquences harmoniques est présente dans le spectre, la fréquence fondamentale peut toujours être perçue même si elle est absente.

La hauteur tonale est l'un des paramètres perceptifs les plus souvent employés en représentation auditive [60]. À l'instar de la sonie, elle est assez intuitivement associée à des métaphores se rapportant à un certain nombre de grandeurs physiques [218]. Comme la plupart des personnes possèdent une forte sensibilité aux changements de hauteurs [91] (la plus petite différence détectable est inférieure à 1 Hz autour de 100 Hz et croît pour atteindre environ 3 Hz pour 1000 Hz [33]), il s'agit d'un paramètre de choix pour rendre perceptibles les relations entre des états de données, que ce soit pour en représenter l'évolution temporelle ou pour en comparer des instances.

Dans cette thèse, une hauteur est désignée soit par la valeur de sa fréquence fondamentale en Hz, soit, dans les cas où l'on utilise des notes de musiques, par la note de la gamme diatonique à laquelle elle correspond selon la notation française *Do*, *Re*, *Mi*, *Fa*, *Sol*, *La*, *Si*, suivie du numéro de l'octave. Ainsi, par exemple, le *La* de la 4^{ème} octave est noté *La*₄.

Timbre

Le **timbre** est la troisième propriété perceptive de base du son, traditionnellement définie comme étant ce qui permet de différencier à l'écoute deux sons dont la hauteur et la sonie sont identiques.

De façon informelle, il est souvent considéré que le timbre est la propriété perceptive qui donne « l'instrument » à l'origine d'un son. Il s'agit d'une combinaison complexe et encore assez difficile à formaliser de propriétés perceptives liées, entre autres, à la composition et l'évolution du spectre fréquentiel, ainsi qu'à l'enveloppe de la sonie.

Le contrôle paramétrique du timbre peut se situer à différents niveaux : on peut soit choisir le timbre parmi un nombre fini d'options pré-construites (par exemple sous le format **MIDI**, les identifiants de programmes permettant de choisir différents sons d'instruments), soit, à un plus bas niveau, contrôler des paramètres de synthèse entrant en jeu dans des opérations générant des timbres complexes (par exemple des filtres du spectre, des enveloppes de volume ou des paramètres pour la synthèse FM, ou encore une synthèse de formants pour générer des sons de voyelles).

En représentation auditive, le timbre est le plus souvent paramétré de façon discrète pour représenter des catégories distinctes ou des dimensions des données. Par exemple dans le modèle de l'outil *Sonification Sandbox* de Walker et al. [217], le paramètre de timbre est choisi manuellement par l'utilisateur pour représenter de façon statique une des dimensions d'un tableau de valeurs. Cependant il arrive aussi que le timbre soit paramétré de façon continue pour représenter l'évolution de données au cours du temps, par exemple en faisant varier les paramètres d'une synthèse FM [32], ou en modifiant la synthèse de formant pour une synthèse vocale [71].

Spatialisation

Comme indiqué parmi les motivations pour la sonification, l'audition binaurale permet une bonne détection de la source d'un son dans l'espace, au moins sur le plan horizontal. La **spatialisation** consiste à simuler, dans un rendu sonore binaural, le positionnement de sources sonores autour de l'auditeur.

Cette représentation dans l'espace peut consister en un simple mapping à la répartition d'un signal sonore dans deux canaux stéréo ou d'une simulation complexe de la réponse acoustique de la tête de l'auditeur (*Head-Related Transfer Function*, ou **HRTF**), voire de l'acoustique d'un lieu simulé (réponse impulsionnelle permettant de simuler la réverbération).

Différents usages de la spatialisation peuvent alors correspondre à différentes tâches servies par une représentation auditive. Ainsi, on peut faire évoluer le

positionnement stéréo d'un son pour traduire l'avancement d'une opération de sorte que ce son débute à gauche et se termine à droite, faisant écho au sens de lecture du texte dans les pays occidentaux. Dans le cas de données possédant des coordonnées dans un espace, il peut être judicieux de transposer ces positions aux positions de sources sonores dans l'espace de l'auditeur. La répartition dans la stéréo peut encore servir pour représenter un objectif de position dans une supervision de gestes, par exemple pour le contrôle du tangage d'un avion [29].

Temps

Les deux principales facettes de l'utilisation du temps comme une dimension perceptive en représentation auditive sont, d'une part, le fait d'associer le temps des données à celui du son, et d'autre part le fait de représenter l'une ou l'autre propriété des données par une modulation du temps dans le son.

En effet, par comparaison avec la vue, l'ouïe est particulièrement apte à détecter des rythmes, en plus des tonalités traduisant des périodes entre 20 et 20 000 Hz. Ainsi, pour des données évoluant selon un axe temporel et présentant potentiellement un comportement particulier dans le temps (des répétitions donnant lieu à des périodes ou rythmes), une stratégie courante est de les représenter par des sons se déroulant selon la même temporalité, afin de faciliter la détection de ces comportements.

Il a par ailleurs été montré qu'un rythme plus rapide dénote une plus grande urgence [65, 130], ce qui peut être mis à profit pour des tâches de supervision. On peut par exemple utiliser un motif musical se produisant à une vitesse plus ou moins élevée pour dénoter la gravité d'une anomalie.

Métaphores naturelles

Le plus souvent, une personne qui entend un son dans son environnement quotidien n'en analyse pas explicitement les propriétés psychophysiques, comme son enveloppe de sonie, sa hauteur tonale ou sa quantité de bruit, mais forme plutôt une image mentale de l'objet auquel il se rapporte (être vivant, objet unique, grains multiples, substance, etc), le type d'action qui provoque ce son (grattement, coup, secousse, etc), ou les propriétés de l'objet (rempli de liquide, lourd, léger, épais, etc).

Une stratégie de communication non verbale peut alors consister à faire passer des informations par diverses propriétés d'un son naturel – ou synthétisé pour imiter un phénomène naturel – de sorte à suggérer des propriétés physiques de la source virtuelle du son, qui sont à leur tour métaphoriquement associées à différents messages. Par exemple, un bruit dans une bande de fréquence de plus en plus haute peut évoquer le son du remplissage d'un récipient, évoquant lui-même le taux de

progression d'une tâche. Cette stratégie est au centre de la technique de sonification des auditory icons, décrite dans la sous-section 1.4.3.

1.3.7 Interaction entre paramètres perceptifs

Les paramètres perceptifs listés précédemment sont à combiner avec prudence, car certains ont tendance à interagir dans la perception. Un exemple de cette interaction peut être observé pour le timbre et la hauteur dans le cas particulier de la fondamentale manquante : de par la composition du spectre, la hauteur perçue correspond à une fréquence fondamentale qui en est pourtant absente. Une autre interaction des paramètres perceptifs bien connue est celle entre la hauteur et l'intensité perçue, puisque les sons de tonalités pures ont tendance à être perçus comme plus forts pour de plus hautes fréquences (voir les courbes de Fletcher-Munson [69]).

1.3.8 Masquage perceptif

Le masquage est un phénomène perceptif qui fait que, de par leurs proximités temporelles et spectrales, deux sons ou plus peuvent se fondre en un seul. Dans le cadre d'une représentation auditive, cela peut être problématique si les sons confondus sont représentatifs de messages différents et individuellement importants. Le masquage peut concerner soit la relation entre les différents sons constituant une même représentation, soit la relation entre les sons d'une représentation et ceux de l'environnement dans lequel elle est écoutée.

D'une part, le risque de masquage au sein d'un même ensemble de sons peut être amoindri en prenant la précaution d'allouer des bandes spectrales sans chevauchement pour chacun des éléments, ou d'imposer des particularités timbrales distinctes pour les différents éléments (par exemple qu'ils soient joués par des instruments différents, que leurs enveloppes de sonie soient percussives ou prolongées, etc). Pour une expérience impliquant quatorze sons potentiellement simultanés, Gaver et al. ont aussi pris soin de faire en sorte que chaque élément soit un son ponctuel plutôt que continu, joué une fois par seconde (environ) et silencieux entre deux occurrences [82].

D'autre part, pour contrer le risque de masquage vis-à-vis d'un environnement sonore extérieur à la représentation (souvent dans des contextes industriels), il peut être judicieux d'effectuer une étude préalable de cet environnement sonore afin de déterminer quelles bandes spectrales sont indisponibles, ou de quelles natures et à quels niveaux sont les sons déjà présents.

Une précaution supplémentaire contre le masquage perceptif se situe dans la teneur sémantique, plutôt que perceptive, des sons employés, et consiste à choisir les différents éléments d'une représentation auditive de sorte à ce

qu'ils appartiennent à un même contexte logique cohérent (parfois appelé une « écologie »), par exemple des sons de la forêt, de la plage, de phénomènes météorologiques, etc. Cette cohérence renforce les métaphores utilisées en aidant la représentation à se détacher de l'environnement d'écoute (par exemple : des sons de la plage dans un environnement de travail en bureau [156]).

1.4 Techniques de sonification

Une tâche de sonification des données consiste généralement à synthétiser des sons en fonction de données. Cette opération peut prendre un grand nombre de formes, selon la teneur des données à traduire, l'information à extraire de ces données, et les mécanismes perceptifs et cognitifs que l'on souhaite mettre en œuvre dans la forme donnée aux sons.

Afin d'encadrer le choix de solutions en fonction des besoins de la représentation, cinq grandes familles de techniques de sonification, qui correspondent à différents contextes typiques auxquels on peut vouloir appliquer une représentation auditive, sont ainsi traditionnellement listées dans la littérature : audification, parameter mapping sonification, auditory icons, earcons et model-based sonification.

Ces techniques sont décrites dans les sous-sections qui suivent. Il est important de noter que cette taxonomie ne prétend aucunement être exhaustive et que, les familles de techniques n'étant pas mutuellement exclusives, il n'est pas rare qu'une sonification pour une problématique spécifique s'inspire d'aspects provenant de plusieurs de ces familles à la fois.

1.4.1 Audification

L'**audification** est une technique de sonification qui consiste à transposer un signal acoustique (ou tout signal suffisamment long sur une seule dimension) dans le domaine de l'audible. L'opération consiste ainsi à modifier, si nécessaire, l'échelle d'amplitude des données pour qu'elles correspondent à des niveaux audibles et supportables pour l'écoute, puis à lire le signal comme un son selon une fréquence d'échantillonnage telle que ses fluctuations soient dans le domaine de l'audible.

Dans les cas où le signal original est dans un domaine fréquentiel particulièrement bas, il est nécessaire de disposer d'un grand nombre d'échantillons pour que le résultat accéléré ne soit pas trop court pour l'écoute : pour une fréquence d'échantillonnage de 44 100 Hz, il faut 44 100 échantillons pour obtenir une seconde de son.

Cette technique permet de mettre à profit les capacités humaines innées d'analyse du signal sonore, même bruité, de sorte à en extraire intuitivement

des propriétés de périodicité (pouvant se traduire selon l'échelle de durée par des tonalités ou par des rythmes), d'amplification, d'atténuation ou d'évolution sur le long terme. Ainsi, par exemple, l'audification s'est avérée un choix de représentation particulièrement populaire pour l'étude de mesures sismiques [201, 72, 95, 56, 57, 145], dans lesquelles différentes phases successives de montée en puissance, déclenchement et atténuation finale peuvent être détectées à l'écoute.

Comme elle ne nécessite pas de technologies de traitement du son particulièrement avancées, cette technique de sonification est nettement plus ancienne que les autres. Il est souvent considéré en effet que les premières expériences d'audification des données datent des années 1880, quand Julius Bernstein a utilisé le haut-parleur du téléphone pour émettre le son de potentiels d'actions musculaires [17].

1.4.2 Parameter mapping

La « **parameter mapping sonification** », parfois abrégée « **PMSon** », est une technique de sonification qui consiste à construire un son paramétré par l'évolution de données souvent multidimensionnelles, pour une lecture du rendu selon un axe temporel. Cet axe temporel peut être celui de l'échantillonnage des données ou n'importe quel autre axe de l'espace des données désigné comme celui des abscisses. Il peut même être contrôlable par l'interaction de l'utilisateur dans des cas d'utilisation pour l'exploration de données.

Les **courbe auditive** (*auditory graphs* en anglais), sont une application assez primitive et très répandue de la **PMSon** qui consistent à représenter des séries de valeurs d'une seule dimension chacune selon un axe des abscisses, le plus souvent en modulant des paramètres psychophysiques basiques comme la hauteur tonale ou la sonie [70, 170].

Les techniques **auditory icons** et **earcons**, décrites ci-après, sont parfois considérées comme des sous-catégories de la **PMSon**. En effet, celles-ci consistent aussi à suivre l'évolution de données en mettant en place des mappings entre leurs différentes dimensions et des paramètres de sons, mais elles se distinguent par les catégories de métaphores sonores mises en place (respectivement des sons naturels ou des sons musicaux) plutôt qu'à des stratégies spécifiques de construction du son.

1.4.3 Auditory icons

Les **auditory icons**, ou icônes auditives en français, sont plus un choix de métaphore sonore qu'une technique exclusive de sonification. Ce choix est motivé par le rapport de l'écoute humaine à la réalité physique d'un environnement sonore. On emploie ainsi des sons naturels, enregistrés ou synthétisés, et souvent

paramétrés par les propriétés des données à représenter, tels que l’objet et l’action qui les provoqueraient dans la réalité peuvent être rapprochés mentalement de l’aspect des données à représenter. Par exemple William Gaver, dans son interface utilisateur sonifiée *SonicFinder*, a proposé que la progression du taux d’accomplissement d’une tâche soit représentée par le son d’une bouteille se remplissant de liquide [79].

Cette technique est souvent associée à la représentation d’évènements ponctuels dans un système de données. Ainsi, à l’instar de *SonicFinder*, elle a longtemps été employée pour représenter d’une façon immersive l’interaction de l’utilisateur dans les interfaces graphiques de systèmes d’exploitation. On se souvient ainsi de la corbeille à papier de Windows produisant un son de papier froissé quand l’utilisateur demande à la vider.

Cependant, cette technique peut aussi être utilisée pour représenter en continu l’évolution de plusieurs dimensions de données, comme dans les expériences sur la supervision de procédés par icônes auditives effectuées par Gaver et al. [82], puis Rauterberg et al. [180].

1.4.4 Earcons

Les **earcons** constituent une famille de choix de métaphores sonores proposée par Blattner et al. [19] à la même époque que les auditory icons (voir la sous-section 1.4.3 précédente), et répondant à des besoins similaires. Dans l’article fondateur, l’auteur présente d’ailleurs les auditory icons comme une variante « représentationnelle » des earcons, mais par la suite ces dernières se sont distinguées par l’utilisation plus spécifique de paramètres musicaux pour la représentation de propriétés des données.

Ces représentations servent, comme les auditory icons, pour la notification, l’interaction et – plus rarement – la supervision continue, mais apportent une plus grande importance à la paramétrisation, de sorte que différentes classes et valeurs caractéristiques des données sont associées à différents enchaînements de notes et autres propriétés (timbre, progression de la sonie, registre) d’une brève séquence musicale [27, 143].

Cette technique se base sur des associations cognitives plus abstraites que les auditory icons. En effet, au lieu de reconnaître le son d’un objet donné réagissant à une stimulation donnée, l’auditeur doit reconnaître une séquence de notes, un timbre d’instrument, une vitesse de déroulement de la séquence. Dans des cas où l’intelligibilité du message est particulièrement importante, les earcons ont tendance à nécessiter une phase d’apprentissage plus longue que les auditory icons.

1.4.5 Model-based

La « **model-based sonification** », parfois abrégée « **MBSon** », a été décrite relativement récemment par rapport aux autres techniques, dans le cadre de la thèse de Thomas Hermann au début des années 2000 [96]. Cette technique consiste à construire un modèle régi par des lois physiques (potentiel gravitationnel, propagation de chaleur, loi de Newton, etc.), dans lequel les points de données sont considérés comme des particules dotées d'un ensemble de propriétés physiques (masse, raideur, charge électrique, température, etc.). Selon la quantité de dimensions à représenter, deux ou trois des dimensions des données peuvent servir de coordonnées à ces particules dans l'espace simulé et, par extension, à leur positionnement dans un rendu binaural.

Un outil interactif est alors associé à ce système, donnant à l'utilisateur la possibilité d'y placer des points d'excitation initiale à partir desquels le modèle évolue, suivant ses lois physiques, jusqu'à atteindre un point d'équilibre. La production de son par le modèle se fait, idéalement en temps réel de l'interaction, par le biais d'une grandeur physique calculée dans le modèle (énergie cinétique, quantité de déplacement) dont les oscillations selon un pas de temps donné jouent le rôle de signal sonore.

Le but de cette technique de représentation est le plus souvent de consulter la répartition spatiale d'une grande quantité de données ne présentant pas d'évolution temporelle, par opposition à la plupart des techniques qui consistent plutôt à suivre l'évolution des données au cours du temps. Il n'est pas alors essentiel de rendre intelligible chacune des dimensions de chacun des points de données, mais plutôt de mettre en évidence les groupes pouvant se former dans l'espace des données, ainsi que les éventuels points isolés.

Pour une description plus exhaustive du principe et un certain nombre d'exemples clarifiant le principe et ses applications, il est recommandé de se référer à la thèse susmentionnée [96].

1.5 Vers un modèle générique pour la sonification de données

Dans ce chapitre, nous avons résumé les connaissances nécessaires sur la synthèse sonore paramétrée et les grandes familles de techniques permettant d'effectuer la sonification de données.

La classification des techniques de sonification donne une idée générale de quelles familles de solutions peuvent être apportées à quels contextes et besoins : interactivité du système de représentation, choix de métaphores sonores, continuité

ou ponctualité des stimuli correspondant aux données, rapport du son à la temporalité des données ou à celle de l'interaction, etc.

Une telle classification fournit un cadre pour la production de sonifications qui suivent les règles de l'état de l'art, mais pour le bien de l'expérimentation et de l'apparition de nouvelles techniques inédites, ces critères pourraient gagner à plutôt être considérés indépendamment les uns des autres. Le chapitre 5 de cette thèse approfondit cette réflexion afin de détacher une liste plus complète des éléments variables pour le design de sonifications.

Ainsi, le but de cette thèse est de définir un format de représentation générique et exhaustif pour le design de représentations auditives. Dans les chapitres 2, 3 et 4 qui suivent, nous décrivons le travail effectué au cours de la thèse sur trois cas d'études de sonifications. Ensuite, nous rapportons dans les chapitres 5 à 9 nos réflexions sur un modèle générique de description d'un outil de sonification et discutons de son applicabilité aux cas d'études, ainsi qu'aux techniques de sonification actuellement connues.

Deuxième partie

Cas d'études

Chapitre 2

Supervision du procédé de fabrication additive wire-arc

Ce cas d'étude porte sur la création d'un ensemble de sons paramétrés pour la supervision en temps réel des critères de fabrication du procédé industriel wire-arc (fabrication additive par fusion) [226], décrit dans la section 2.1. Plusieurs itérations successives de représentations ont été produites et évaluées au cours de la thèse. Celles-ci sont présentées dans la sous-section 2.2. Les problématiques corollaires associées à ces solutions sont ensuite explorées : la construction d'un protocole expérimental pour comparer des représentations auditives dans la section 2.3, puis l'étude de l'influence de la cohérence écologique sur la qualité d'une représentation dans la section 2.4.

2.1 Contexte

2.1.1 Affiliations

L'étude de la sonification pour la supervision du procédé wire-arc a été soumise à notre attention par Nadine Couture de l'École Supérieure des Technologies Industrielles Avancées (ESTIA), Emmanuel Duc de Sigma-Clermont (groupe INP), et Sébastien Ibarboure, alors doctorant à ESTIA Recherche et effectuant une thèse sur la supervision du procédé selon une modalité haptique [115]. La recherche sur les technologies de fabrication d'ESTIA Recherche est menée sur la plateforme Addimadour, qui concentre divers procédés de fabrication additive métallique. Cette plateforme a pour enjeux la robotisation et l'industrialisation de ces procédés pouvant faire l'objet d'un transfert technologique vers les entreprises.

2.1.2 Le procédé

Wire-arc appartient à la famille de procédés de fabrication par fusion, la cinquième des catégories de procédés de fabrication listées par Blum et al. en 2017 [20]. Un apport concentré d'énergie (arc électrique) fait fondre un matériau provoquant son dépôt sur une zone en dessous de la tête d'impression appelée le bain de fusion, ou cordon (*A* sur la figure 2.1). La solidification de la matière le long de la trajectoire de l'outil forme couche par couche une pièce métallique (*B* sur la figure 2.1). Ces procédés sont employés pour diverses applications industrielles comme la production de pièces pour l'aérospatiale ou l'automobile, ainsi que la production de prothèses en titane pour des applications médicales.

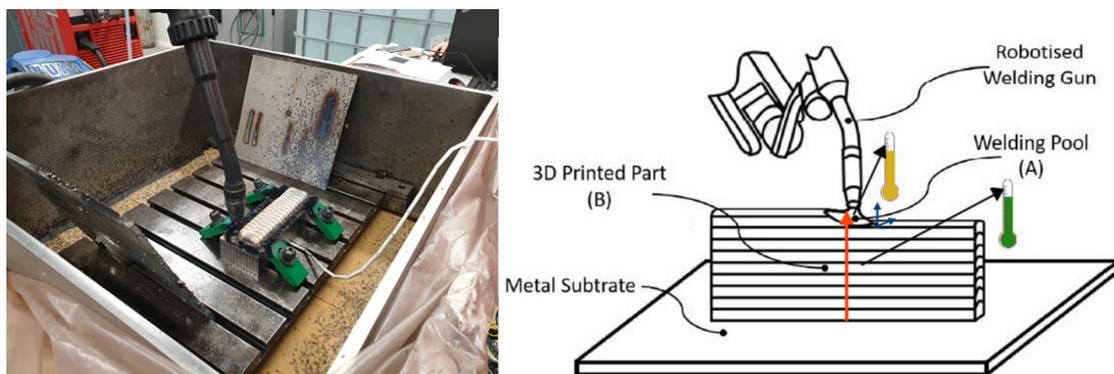


FIGURE 2.1 – Le dispositif de fabrication wire-arc. À gauche photographié au repos. À droite schématisé avec un code couleur se rapportant aux critères à superviser.

2.1.3 Enjeux de la représentation en réalité augmentée

La recherche à laquelle participe ce cas d'étude s'inscrit dans l'optique de l'industrie 4.0, un paradigme de fabrication récent souhaitant proposer un produit personnalisé avec une empreinte écologique faible. L'objectif de l'automatisation, dans cette optique, n'est pas de supplanter l'opérateur humain, mais de l'impliquer dans la boucle pour l'aider à développer ses compétences en résolution de problèmes. Il est donc nécessaire de proposer une alternative à l'automatisation complète des procédés afin de permettre un contrôle et une réaction en temps réel aux défauts de fabrication qui pourraient amener à mettre des pièces fautives au rebut, causant la perte de l'énergie et des matériaux qui ont été nécessaires à leur fabrication. On veut pouvoir diagnostiquer les défauts suffisamment tôt pour passer par des phases de correction quand c'est possible. Dans le cas échéant de la mise au rebut d'une pièce irrécupérable, on veut pouvoir s'y résoudre avant

que la fabrication ait trop progressé dans la mauvaise direction, afin de réduire le gaspillage.

2.1.4 Contexte sensoriel de la fabrication

Le procédé de fabrication additive nécessite différentes mesures de sécurité qui réduisent fortement l'apport de la vue dans la supervision. En effet, la fabrication se produit dans une cuve remplie d'argon, gaz inerte remplaçant l'oxygène, de sorte qu'il faut s'approcher et regarder par-dessus le bord pour avoir une idée de ce qu'il s'y passe. À son tour, cette observation est rendue assez risquée par le fait que des particules de matériau en fusion sont souvent projetées par-dessus le bord de la cuve. Afin de se protéger des jets de particules et de la lumière intense produite par la fusion, il est obligatoire d'observer le procédé à travers un masque de soudure. Ce masque possède une petite fenêtre de verre teinté qui réduit fortement le champ de vision, en plus de diminuer grandement la luminosité visible.

Le son ambiant autour d'une imprimante en pause n'est pas très intrusif. Il est alors toujours possible d'échanger verbalement sans forcer sa voix. Lors d'une passe de fabrication cependant, il est impossible de communiquer à volume normal. Ce son n'est pas douloureux à supporter, mais reste assez désagréable sur le long terme.

2.1.5 Intuitions pour le diagnostic par l'écoute

Lors de la fabrication, le champ visuel est obscurci par les conditions de sécurité de sorte qu'un opérateur est déjà amené, instinctivement, à faire appel à d'autres sens pour obtenir certaines informations de base sur le bon déroulement du procédé. Au cours de conversations informelles avec des opérateurs à Addimadour, il a ainsi été encourageant d'apprendre que l'écoute pouvait déjà servir, dans une certaine mesure, d'outil de diagnostic sur les machines wire-arc. Certains changements dans le grain constituant le bruit de la fabrication peuvent en effet donner des indications sur l'écoulement de matière, comme par exemple le blocage du matériau dans la tête d'impression (amplification du craquèlement puis interruption soudaine du bruit), ou une trop grande distance entre la tête d'impression et la pièce. Ce deuxième exemple a même fait l'objet en 2021 d'une étude pour l'automatisation par analyse spectrale [35], puis d'un chapitre de la thèse d'Alexia Chabot (École centrale de Nantes, Octobre 2020) [34].

Pourtant, le bruit étant assez déplaisant et porteur d'une quantité assez limitée d'informations notre but est de le remplacer, dans un casque antibruit, par un son qui serait synthétisé dans le but précis de communiquer sur l'état du système, et dont l'esthétique aurait été recherchée pour permettre une écoute prolongée pour un taux de stress limité.

2.1.6 Critères de fabrication

Les critères à superviser pour assurer le bon déroulement d'une fabrication sont les suivants :

- La **Hauteur du Cordon** (*HC*).
- La **Largeur du Cordon** (*LC*).
- La **Hauteur de la Pièce** (*HP*) au niveau de la tête d'impression, à comparer au fur et à mesure de l'impression à la **Hauteur théorique de la Pièce** (*HP₀*).
- La **Température du Cordon** (*TC*).
- La **Température de la Pièce** (*TP*).

Le tableau 2.1 détaille les seuils de valeurs à respecter et les taux de tolérance vis-à-vis de ces valeurs.

Critère	Seuil	Tolérance
<i>HC</i>	3.00mm	±10%
<i>LC</i>	4.00mm	±10%
<i>HP</i>	Nombre couches *3.00mm (= <i>HP₀</i>)	±1.50mm
<i>TC</i>	2000°C	±10%
<i>TP</i>	600°C	Aucune

TABLE 2.1 – Les critères de fabrication du procédé wire-arc.

Ainsi, pour s'assurer que les conditions sont les bonnes pour un dépôt et une cristallisation corrects du matériau, il faudrait pouvoir percevoir, entre autres, que *TC* ne dévie pas de plus de 200 °C dans l'une ou l'autre direction autour de 2000 °C, ou que *HC* ne dévie pas de plus que 0.3mm. Il s'agit de divergences extrêmement difficiles, sinon impossibles, à percevoir par la vue seule et, dans les conditions normales de fabrication, la contribution d'autres sens comme l'audition n'apporte pas suffisamment de précision pour détecter des changements de géométrie de l'ordre de 10⁻⁴m.

2.1.7 Simulation de l'étude

D'après des états de l'art par Everton et al. [66] puis Xia et al. [233], diverses expériences ont été réalisées tentant d'effectuer la captation des critères de fabrication en temps réel, mais il ne s'agit pas encore d'une technologie aboutie et largement distribuée. Ce fait, combiné au manque relatif d'accès à des testeurs spécialisés dans le domaine d'application (des opérateurs d'Addimadour ou des chercheurs en interaction humain-machine familiarisés avec ce procédé) donne à la présente étude une nature plutôt hypothétique.

On envisage cependant un scénario où l'imprimante serait dotée de capteurs mesurant les critères de fabrication, et l'opérateur pourrait se trouver à une distance confortable, éventuellement dans une autre pièce, avec un contrôle sur la fabrication (démarrer, arrêter, reprendre, planifier, etc). Les données de capteurs seraient alors communiquées à l'utilisateur à des fins de diagnostic par différents biais de réalité augmentée : haptique, visuel et, dans le cas présent, auditif à travers un casque antibruit.

On part du principe que l'utilisation d'un casque antibruit et la distance à la machine collaborent à rendre le bruit de la fabrication aussi peu masquant que possible vis-à-vis des sons employés pour la représentation. Des études par Molesworth et Burgess ont montré que le port d'un casque antibruit permet non seulement la transmission d'informations critiques dans un contexte de bruit (dans leur cas, en avion) [152], mais qu'il permet aussi de meilleures performances à des tâches concurrentes [153].

Dans ces expériences, l'intensité sonore du bruit est de 78 dB au niveau de la source sonore et 65 dB au niveau de la tête des testeurs [152]. Une étude en cours sur l'intensité sonore des procédés de fabrication a mesuré que celle-ci était entre 74 et 77 dB à proximité de la machine [147]. Sous réserve que l'utilisateur se trouve à une distance raisonnable de la machine durant son fonctionnement, il semble alors vraisemblable que les conditions soient à peu près comparables à celles des expériences de Molesworth et al., donc qu'un casque antibruit puisse suffisamment atténuer le bruit de la fabrication pour permettre l'écoute de notifications auditives.

Afin de palier l'absence de capteurs pour nos expérimentations, les prototypes de sonification produits pour ce cas d'étude ont été réalisés à partir de fichiers traces simulés dans un format **Comma-Separated Values (CSV)** tel que, colonne par colonne, on peut y lire : le temps en secondes, le numéro de la couche en cours de fabrication, la longueur parcourue par la tête d'impression en décimètres, la hauteur du cordon *HC* en millimètres, la largeur du cordon *LC* en millimètres, la hauteur de la pièce *HP* en millimètres, la hauteur théorique de la pièce en millimètres, la température du bain de fusion *TC* en °C, et la température moyenne de la pièce *TP* en °C. Ces valeurs suivent une organisation temporelle telle que chaque ligne représente une mesure prise 0.06 secondes après la précédente.

La méthode pour constituer ces simulations de données en temps réel a consisté à combiner des mesures géométriques, prises a posteriori sur des cordons solidifiés, avec des mesures de températures effectuées par une caméra thermique. On dispose de 30 fichiers traces ainsi construits et représentant divers déroulements possibles de la fabrication, 20 d'entre eux comprenant une situation considérée comme un défaut soit de géométrie, soit de température du bain de fusion.

Comme l'implémentation réelle d'une représentation auditive dans le contexte de la fabrication additive wire-arc est actuellement hypothétique, ce cas d'étude a été l'occasion de soulever certaines questions de psychoacoustique sur la conception et l'évaluation de notifications sonores ambiantes d'une façon plus générale. Ainsi, dans les sous-sections qui suivent, on détaille les choix de représentation testés, puis les problématiques corollaires qui ont été abordées lors du travail sur ce cas d'étude : le développement d'une méthodologie d'évaluation en laboratoire pour les représentations auditives périphériques, et la comparaison de représentations constituées de sons formant des ensembles logiques plus ou moins cohérents.

2.2 Évolution des sonifications proposées

Les choix de représentation pour ce cas d'étude ont fait l'objet d'un développement itératif à l'occasion des différentes expériences successives. Dans cette section nous décrivons la façon dont les retours utilisateurs ont influencé les choix de timbres, de métaphores et d'échelles perceptives pour les sonifications proposées.

Les tableaux de mappings 2.2, 2.3, 2.4 et 2.5 résument, de gauche à droite, la transformation des propriétés extraites des différentes dimensions des données en paramètres perceptifs, puis les flux sonores qui sont contrôlés par ces paramètres.

2.2.1 Préliminaire : sonifications conçues au cours du stage recherche

Ce cas d'étude a été abordé une première fois à l'occasion d'un stage recherche de l'auteur au **Studio de Création et de Recherche en Informatique pour les Musiques Expérimentales (SCRIME)** de l'Université de Bordeaux, portant sur une préparation du présent travail de thèse [176]. Dans ce contexte, trois mappings différents ont été proposés afin d'alimenter les premières réflexions sur les modalités d'évaluation à mettre en œuvre, les techniques de programmation abordables pour la sonification et certaines préférences des testeurs vis-à-vis des choix sonores.

Il avait alors été décidé de produire des prototypes de sonifications construites par un programme simple en C associant aux critères de fabrication divers assemblages d'ondes sinusoïdes et de bruits. La brève vidéo de démonstration de ces sons qui avait été montrée aux testeurs lors de l'évaluation est visible en ligne¹.

Certains aspects récurrents des représentations proposées par la suite ont été mis en place à cette étape du développement, en particulier la recherche d'un ensemble

1. <https://youtu.be/GFLGhfUGHy0> – visité le 4/1/23

de flux sonores sans masquage : les critères étaient représentés par quatre flux de différentes granularités, dans différents domaines de fréquences et de timbres. Les deux critères de géométrie du cordon *HC* et *LC* étaient représentés par deux mappings de paramètres pour un même flux sonore. Deux flux sonores avaient été choisis pour se dérouler constamment, tandis que deux n'émergeaient qu'en cas d'anomalie.

D'après les premiers retours d'auditeurs, il était nécessaire que les timbres soient plus complexes afin de produire un effet moins machinal et pénible à l'écoute. Le souci de précision qui avait initialement amené à utiliser des hauteurs microtonales pour une traduction exacte de la fluctuation des valeurs a été remis en question par le besoin de confort des auditeurs, nécessitant d'utiliser le plus possible de notes sélectionnées dans la gamme tempérée. L'évaluation avait aussi établi que le mapping au volume seul était insuffisant pour communiquer des variations précises. Ces besoins de mappings plus complexes et de timbres plus riches nous ont amenés à utiliser l'outil de synthèse sonore *SuperCollider* pour produire les prototypes ultérieurs.

2.2.2 Aspects récurrents

Suivant les recommandations de Hildebrandt et al. [110], chacune des sonifications proposées consiste au moins en partie en une collection de flux sonores jouant continuellement. Cette partie constante du son, que l'on appelle ici « baseline », est prévue pour jouer un rôle de confirmation ambiante que l'imprimante et la sonification fonctionnent, rôle pour lequel le silence serait plus ambigu puisqu'il pourrait aussi dénoter un problème généralisé du système. Passées les expérimentations de départ produites au cours du stage de Master [176] il a été décidé que cette baseline serait toujours associée aux critères géométriques, tandis que les critères de températures seraient relégués à des alertes plus ponctuelles uniquement audibles en cas de défauts.

Des expériences réalisées par Peres et al. en 2005 ont suggéré qu'une stratégie de mapping redondant particulièrement efficace consistait à combiner, pour une même information, des mappings à la hauteur et au volume [170]. La plupart des flux sonores construits dans les différentes représentations, à l'exception de ceux correspondant au critère *TP* qui n'avaient pas besoin d'être paramétrés, ont ainsi employé cette stratégie sous une forme ou une autre.

Des stimuli sonores se produisant à une plus grande vitesse sont perçus comme indicateurs d'un plus grand taux d'urgence [65], plus particulièrement pour des

sons dont le temps d'attaque est court [130]. Cette perception a été mise à profit pour communiquer la quantité de divergence de la largeur du cordon, produisant un son répété de plus en plus rapidement en cas de formation trop étroite ou trop large.

Afin de réduire le nombre de flux sonores concurrents à superviser, nous avons toujours fait en sorte que les deux critères *HC* et *LC* donnant la géométrie du cordon soient traduits par les propriétés perceptives d'un même flux sonore. Le choix de ces deux grandeurs en particulier est motivé par le fait qu'elles qualifient un même objet physique et que le simple diagnostic d'un défaut de géométrie, sans détail sur sa direction ou sa quantité, est considéré comme suffisant dans un premier temps.

2.2.3 Instances consécutives et adaptations

Mixte0 Cette représentation a été conçue comme un nouveau point de départ pour la série de prototypes conçus sous *SuperCollider*. On décide d'une première collection de plages de valeurs pour les hauteurs, volumes et vitesses, en prévoyant qu'un premier retour utilisateur nous indiquera quels aspects du design devront être adaptés par la suite. La conception et l'évaluation de cette représentation sont rapportées dans l'article [242]. Le tableau 2.2 détaille les échelles perceptives des mappings pour cette représentation. Une démonstration de ces sons est donnée dans la présentation vidéo associée à la publication ².

- Baseline géométrie :
 - *Mélodie* : une séquence de trois notes jouées en boucle par le synthétiseur par défaut de *SuperCollider*, un timbre de piano simplifié. Cette séquence consiste en une suite « tonique - tierce - quinte » sur l'échelle majeure, paramétrée pour que *HC* détermine la tonique de départ et *LC* détermine l'intervalle entre chaque note. Les deux critères de géométrie sont aussi associés au volume sonore du flux, de sorte qu'une divergence de l'un ou l'autre soit aussi marquée dans le son par une amplification.
 - *Bourdon* : un son tonal continu construit à partir d'une onde en dents de scie filtrée. La hauteur de ce son devient plus basse si *HP* est en dessous de sa valeur théorique, plus haute si *HP* est au dessus, et son volume augmente avec la quantité de divergence.
- Alertes température :
 - *Pluie* : un enregistrement du son de la pluie, désactivé en situation normale et dont le volume augmente en cas d'anomalie. Un mapping

2. https://youtu.be/86uT2e_-204 – visité le 4/1/23

expérimental supplémentaire a été ajouté à la vitesse de lecture pour que le son devienne plus lent en fonction de la quantité de divergence, mais le changement s’est avéré quasiment inaudible.

- *Tonnerre* : un enregistrement du son du tonnerre. Sélectionné et filtré pour que le son de la pluie en soit absent.
- Retour utilisateurs : les testeurs ont demandé à ce que les sons de la baseline soient plus lents, moins forts, plus bas. *Bourdon* en particulier était trop sensible et avait tendance à produire, pour de petites divergences dans les données, des fluctuations ponctuelles de hauteurs jugées désagréables. Malgré le filtrage qui avait été effectué sur le son de tonnerre pour qu’il se distingue du son de pluie, il est arrivé que les participants signalent entendre l’un dans l’autre. Ces mêmes sons ont par ailleurs été jugés trop effrayants par certains testeurs. Enfin, pour l’itération suivante, nous avons cherché à remplacer le mapping à la vitesse de *Pluie* par une sélection indiquant plus clairement la direction de l’anomalie de température.

<i>HC</i> <i>LC</i>	Différence hauteur	Hauteur tonique : Do_5 (normal) à Fa_7 (pire)	<i>Mélodie</i>
	Différences H & L	Amplitude : 0.05x à 0.2x	
	Différence largeur	Intervalles : 0.1 à 1.0 s	
<i>HP</i>	Différence relative	Hauteur : $Do_4 \pm 10$ demi-tons	<i>Bourdon</i>
	Différence absolue	Amplitude : 0.1x à 0.4x	
<i>TC</i>	Différence absolue	Vitesse de lecture : 0.25x à 2.0x	<i>Pluie</i>
	Différence absolue	Amplitude : 0.0 à 0.9	
<i>TP</i>	Seuil	Déclenche : échantillon sonore	<i>Tonnerre</i>

TABLE 2.2 – Les mappings pour la représentation *Mixte0*.

Mixte1 Cette représentation a été proposée comme une refonte de *Mixte0* selon les retours des testeurs. Voir l’article [243] pour l’introduction de cette représentation. Le tableau 2.3 détaille les échelles perceptives des mappings pour

cette représentation. Une démonstration vidéo de ces sons est disponible en ligne³.

- Baseline géométrie : afin de créer une baseline moins anxiogène, nous avons réutilisé les flux *Méloдие* et *Bourdon* en leur imposant des échelles de hauteurs plus basses, des intervalles plus grands et des échelles de volumes moins élevées en situation normale.
- Alertes température : remplacement des sons négativement connotés de la pluie et du tonnerre par des sons du quotidien.
 - *État Eau* : émergence en volume de sons d'eau en ébullition ou de craquèlement de glace choisis comme métaphores respectivement pour la surchauffe ou la sous-chauffe du cordon.
 - *Friture* : son d'eau s'évaporant au contact d'une surface brûlante.
- Retours utilisateurs : comme l'évaluation de l'expérience utilisateur pour *Mixte0* n'impliquait aucune quantification du ressenti, il ne nous est pas possible de confirmer objectivement que *Mixte1* a constitué une amélioration. Les commentaires libres semblent cependant confirmer que les nouveaux choix d'échelles perceptives ont rendu la baseline moins stressante. Ces commentaires ont aussi soulevé les mêmes problèmes que précédemment quant à la sensibilité, pourtant diminuée, du *Bourdon*. Par ailleurs une nouvelle confusion a été rapportée par certains testeurs entre le son de *Friture* et le son d'eau en ébullition.

<i>HC</i> <i>LC</i>	Différence hauteur	Hauteur tonique : <i>Do</i> ₅ (normal) à <i>Fa</i> ₆ (pire)	<i>Méloдие</i>
	Différences H & L	Amplitude : 0.02x à 0.2x	
	Différence largeur	Intervalles : 0.5 à 1.5 s	
<i>HP</i>	Différence relative	Hauteur : <i>La</i> ₃ ±5 demi-tons	<i>Bourdon</i>
	Différence absolue	Amplitude : 0.1x à 0.4x	
<i>TC</i>	Signe de différence	Sélection : craquements (-) ou ébullition (+)	<i>État Eau</i>
	Différence absolue	Amplitude : 0.0x, 0.2x ou 0.5x	
<i>TP</i>	Seuil	Déclenche : échantillon sonore	<i>Friture</i>

TABLE 2.3 – Les mappings pour la représentation *Mixte1*.

3. <https://youtu.be/7-xDvA5L3n0> – visité le 4/1/23

Vers deux représentations cohérentes Les deux représentations *Mixte0* et *Mixte1* sont basées sur l’hypothèse que l’on peut faciliter la distinction de flux sonores en les sélectionnant à partir de catégories cognitives volontairement différentes (ici, de la musique synthétique pour la géométrie et des bruits du quotidien pour la température). Les recommandations de la littérature étant le plus souvent contraires à ce principe, nous avons voulu tester concrètement l’intérêt d’une telle séparation. Voir la section 2.4 ou l’article [241] pour le détail de cette étude. Les deux ensembles cohérents *Synth* et *Nature* décrits ci-dessous ont ainsi été construits à partir de *Mixte1* afin d’être utilisés dans une évaluation comparative. Une démonstration de ces sons est disponible en ligne⁴.

Synth Une collection de stimuli tonals. Les sons musicaux déjà utilisés dans *Mixte1* sont réutilisés et deux nouveaux sons tonals sont choisis pour remplacer les métaphores naturelles associées à la température.

- Baseline géométrie : *Mélodie* et *Bourdon*, idem que *Mixte1*.
- Alertes température :
 - *Sonnerie* : émergence en volume d’un grain tonal à une fréquence valant soit 220Hz, soit 660Hz selon la polarité de l’anomalie de *TC*.
 - *Cloche* : un son de cloche synthétique déclenché en cas de dépassement du seuil pour *TP*, constitué de 3 composantes inharmoniques piochées aléatoirement entre 220 et 880 Hz.

Nature Une collection de métaphores naturelles. Les sons naturels déjà utilisés dans *Mixte1* sont réutilisés et deux nouveaux sons naturels sont choisis pour remplacer les flux tonals associés à la géométrie.

- Baseline géométrie :
 - *Gouttes* : un son de goutte d’eau répétitif dont la vitesse de lecture suit un mapping à *HC* et dont le rythme suit un mapping à *LC*.
 - *Oiseaux* : un mix de 3 flux sonores constitués de chants d’oiseaux de différentes espèces en fonction de la divergence de *HP* : canards pour trop haut, corbeaux pour trop bas, et divers petits oiseaux pour la situation normale. Ce choix est inspiré par les différentes connotations émotionnelles associées respectivement aux corbeaux et aux canards dans le folklore occidental. Le corbeau est traditionnellement vu comme un charognard et un mauvais présage (donc valeurs « basses »), tandis que le canard est un animal familier, voire comique (donc valeurs « hautes »).
- Alertes température : *État Eau* et *Friture*, idem que *Mixte1*.

4. <https://youtu.be/s7kGfN0d47A> – visité le 4/1/23

<i>HC</i> <i>LC</i>	Différence hauteur	Hauteur tonique : Do_5 (normal) à Fa_6 (pire)	<i>Mélodie</i>
	Différences H & L	Amplitude : 0.02x à 0.2x	
	Différence largeur	Intervalles : 0.5 à 1.5 s	
<i>HP</i>	Différence relative	Hauteur : $La_3 \pm 5$ demi-tons	<i>Bourdon</i>
	Différence absolue	Amplitude : 0.1x à 0.4x	
<i>TC</i>	Signe de la différence	Hauteur : 220 Hz (-) ou 880 Hz (+)	<i>Sonnerie</i>
	Différence absolue	Amplitude : 0.0x, 0.1x ou 0.2x	
<i>TP</i>	Seuil	Déclenche : Composantes aléatoires	<i>Cloche</i>

TABLE 2.4 – Les mappings pour la représentation *Synth.*

<i>HC</i> <i>LC</i>	Différence hauteur	Vitesse de lecture : 0.5x à 2.0x	<i>Gouttes</i>
	Différences H & L	Amplitude : 0.1x à 0.8x	
	Différence largeur	Intervalles : 0.5 à 1.5 s	
<i>HP</i>	Différence relative	Mixage : canards (+), divers, ou corbeaux (-)	<i>Oiseaux</i>
	Différence absolue	Amplitude : 0.0x à 0.3x	
<i>TC</i>	Signe de la différence	Sélection : craquements (-) ou ébullition (+)	<i>État Eau</i>
	Différence absolue	Amplitude : 0.0x, 0.2x ou 0.5x	
<i>TP</i>	Seuil	Déclenche : échantillon sonore	<i>Friture</i>

TABLE 2.5 – Les mappings pour la représentation *Nature.*

2.2.4 Perspectives pour l'amélioration de la représentation auditive

L'un des principaux problèmes qui reste à résoudre dans la sélection des sons est celui de la confusion entre *Friture* et *État Eau*. Bien qu'il s'agisse de sons occupant des zones du spectre différentes, donc « objectivement » différents, il s'agit aussi de sons associés à des phénomènes similaires d'ébullition plus ou moins violente de l'eau, ce qui a pu être la cause des erreurs. Une voie qui resterait à explorer serait de faire en sorte que ces sons soit représentés par des métaphores associées à des phénomènes réels complètement différents, quitte à sacrifier le rapprochement des données de température à une même catégorie cognitive.

Afin de palier l'hypersensibilité perçue du stimulus *Bourdon*, il pourra être intéressant de plutôt effectuer une variation sur le timbre, un paramètre perceptif prometteur en termes de seuils de détection [117, 116], tout en étant potentiellement moins stressant qu'une fluctuation de hauteur. Après avoir beaucoup compté sur ce paramètre pour créer des catégories sonores nettes pour chacun des critères de supervision, il pourrait être intéressant d'en user aussi pour la représentation de valeurs continues. On pensera par exemple, pour *Bourdon*, à passer par une synthèse FM dont la fréquence porteuse serait constante et dont la fréquence modulante suivrait les variations de *HP*. Plus généralement, le ressenti des testeurs étant encore globalement assez négatif, il y a encore beaucoup d'explorations possibles sur la qualité des timbres à employer pour créer une expérience esthétique optimale.

2.3 Protocole d'évaluation « en laboratoire » pour la comparaison de représentations auditives périphériques

Pour le développement itératif d'un outil, il est nécessaire d'acquérir des données permettant de comparer selon les mêmes critères la qualité d'une version par rapport à la précédente. Une telle évaluation doit permettre de quantifier la qualité de l'outil à la fois vis-à-vis de la tâche qu'il est censé rendre possible et de la façon dont il est subjectivement perçu par les utilisateurs.

Le procédé auquel notre sonification est vouée n'étant pas encore doté des moyens de la tester dans les réelles conditions d'utilisation, on doit se contenter dans la phase actuelle de protocoles d'évaluation dits « en laboratoire » visant à simuler ces conditions réelles. Par ailleurs, une majorité de ce travail de thèse s'est déroulée dans une période de crise sanitaire qui n'a pas permis de convier

les testeurs à participer à l'expérience dans un contexte contrôlé impliquant, selon l'usage, la présence d'un expérimentateur pour clarifier les instructions de l'exercice, ainsi qu'un poste de travail, une souris et un casque audio constants.

Dans cette section, nous décrivons donc la conception d'un protocole expérimental à effectuer chez soi pour la comparaison de représentations auditives périphériques. On considère qu'il s'agit d'une évaluation « en laboratoire » dans la mesure où le protocole est basé sur les recommandations habituellement données pour la simulation en laboratoire des conditions de travail pour une supervision périphérique, tout en gardant à l'esprit la perspective de l'appliquer dans un contexte mieux contrôlé à l'avenir.

2.3.1 Préliminaire : tâches d'identification

Dans un premier temps, nous décrivons les processus expérimentaux appliqués pour les premiers prototypes présentés dans [176] et [242]. La principale différence entre ces deux expériences est que, pour [176], l'intervention du testeur se faisait sur un support papier à partir de ressources numériques à télécharger et déployer soi-même, tandis que l'expérience pour [242] était centralisée sur une page web en **HTML**, **CSS** et **PHP**. L'expérience totale durait environ une demi-heure.

L'expérience Les testeurs commencent par lire une introduction textuelle au procédé supervisé, puis regardent une vidéo de démonstration de 3 minutes présentant chacun des mappings entre les flux sonores et les états de la fabrication. Ils doivent ensuite écouter une série de bandes sons de 30 secondes, une seule fois chacune, et cocher les cases du formulaire adjacent qui correspondent aux anomalies entendues. Une capture d'écran de ce formulaire est visible sur la figure 2.2.

Les simulations sont sélectionnées de sorte qu'il y ait une moitié de passes sans anomalies et l'autre de diverses combinaisons d'anomalies. Dans cette version du protocole expérimental, les stimuli étaient toujours présentés dans un même ordre choisi aléatoirement lors de sa conception.

Motivation pour ce format d'expérience L'intention était alors de mesurer la capacité des testeurs à détecter et reconnaître les notifications proposées, puis à refaire les associations cognitives entre les sons et les états de la fabrication.

Suggestions d'après les résultats Les résultats de ces expériences ont semblé montrer qu'en l'absence d'une phase prolongée d'entraînement au diagnostic, qui serait difficile à implémenter sur le cours terme, les réponses seraient

Son	Dimensions Cordon	Hauteur Pièce	Température Cordon	Température Pièce
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Remarques, commentaires (ne pas sauter de ligne, cela envoie la réponse):

Envoyer

FIGURE 2.2 – Une capture d’écran de l’interface d’évaluation en ligne utilisée pour [242]

inévitablement « faussées » par le manque de familiarité des testeurs avec le procédé sonifié.

Une certaine confusion a été rapportée par les testeurs de [176] quant aux rapprochements cognitifs à faire avec les états de la fabrication. Dans une expérience similaire proposée pour l’itération suivante de la représentation, nommée *Mixte0* [242], deux testeurs ont rapporté avoir eu des difficultés à refaire le rapprochement entre les sons et les critères, mais ont aussi affirmé dans le même commentaire qu’ils étaient confiants en leur capacité à s’y adapter après plus de temps d’entraînement. À l’issue de cette seconde expérience, aucune des associations cognitives n’a été jugée comme aberrante ou contre-intuitive, la seule confusion clairement marquée dans les résultats étant celle, perceptive, entre *Pluie* et *Tonnerre*. La plupart des erreurs dans ces résultats semblaient plutôt liées à la difficulté à interpréter les fluctuations de *Bourdon* (voir la description des sons de *Mixte1* dans la section 2.2) qu’à sa ressemblance avec un autre flux sonore.

Motivation pour l'évolution du procédé Les mesures obtenues à l'aide de ce premier protocole étaient suffisantes tant qu'il s'agissait de mesurer la détectabilité d'un ensemble de stimuli, mais afin d'obtenir une quantification de sa valeur en tant que représentation périphérique, nous avons besoin d'acquérir une mesure plus précise de la charge d'attention requise par le son.

Il s'agissait aussi de pouvoir mesurer plus précisément le temps de réaction à une alerte, en particulier dans une situation où la réaction pouvait nécessiter de se séparer d'une tâche principale active.

Au cas où l'ordre de présentation des stimuli pouvait avoir une influence sur la progression de la réaction des testeurs au cours de l'exercice, nous avons aussi décidé de faire en sorte qu'ils soient présentés dans un ordre aléatoire dans toutes les expériences qui ont suivi.

Faute de pouvoir implémenter une expérience sur le long terme comprenant une phase d'entraînement et une étude des progrès des testeurs, il nous a fallu choisir soit d'évaluer la perception des stimuli eux-mêmes, soit d'évaluer leur rapprochement cognitif avec les défauts. Comme les résultats pour *Mixte0* semblaient prometteurs sur ce second aspect et que les itérations suivantes seraient des évolutions selon les mêmes principes, nous avons choisi la première de ces options pour le temps de l'étude, tout en continuant d'apporter un soin particulier, dans notre sélection des métaphores et paramètres, à préparer le terrain pour la deuxième.

2.3.2 Existant pour l'évaluation de sonification ambiantes

Différentes approches sont possibles pour évaluer une représentation périphérique, qu'elle soit auditive ou pas. Idéalement, le moyen le plus robuste est d'étudier son utilisation directement dans son contexte d'usage prévu (on parle d'implémentation *in situ*) [189, 210]. Cependant, à des stades préliminaires, cela n'est pas toujours possible ou judicieux, souvent à cause d'un manque d'équipement ou parce que l'outil est toujours à un stade trop expérimental pour être représentatif de ce à quoi le public visé pourrait s'attendre.

Dans beaucoup de cas, le fait de simplement demander aux testeurs de juger leur expérience par le biais d'entrevues et de sondages peut suffire pour collecter des informations sur les qualités esthétiques et ambiantes d'une représentation vouée à un contexte d'utilisation peu critique : la maison, le bureau, ou les espaces publics [113, 168, 156, 124, 186].

Cependant, dans les cas où la représentation a besoin de transmettre des informations critiques, l'évaluation doit aussi tenir compte de l'intelligibilité de ces informations. Il est alors nécessaire de récolter des données quantitatives sur la performance des testeurs utilisant la représentation, le plus souvent à l'aide d'une tâche d'identification [23, 141, 242].

Quand une représentation d'information critique doit pouvoir servir dans un contexte de travail plus large impliquant potentiellement d'autres tâches, il est aussi nécessaire d'obtenir une mesure de la distraction. McCrickard et al. recommandent ainsi d'employer une méthode d'évaluation à deux tâches concurrentes [142], l'une étant l'identification des notifications et l'autre un exercice simple permettant de quantifier en temps réel la capacité à gérer d'autres activités. Cette méthodologie a aussi été recherchée par Hausen et al. en 2014 [93], ainsi que dans le cadre de la thèse de Maxime Daniel en 2018 [46], et elle a été implémentée à l'occasion de nombreuses expériences sur la représentation auditive périphérique [208, 9, 170, 121, 112, 109, 102, 110].

Dans le cadre de ce travail, l'expérimentation *in situ* n'est pas encore faisable à défaut de capteurs pouvant fournir les mesures à superviser. Notre but est cependant de proposer une représentation permettant de superviser le procédé sans nécessiter d'exposition directe à la fabrication. Pour cela, nous devons envisager d'autres activités rendues disponibles par cette liberté sensorielle comme, par exemple, le fait de vérifier ses e-mails ou de préparer une autre fabrication. Ainsi, notre expérience doit tenir compte à la fois de l'intelligibilité des données et de la capture de l'attention, ce qui est fait par le biais d'un protocole expérimental à deux tâches. Afin de toujours récolter des informations sur la qualité subjective de la représentation entendue, ce protocole comporte aussi un sondage final.

L'une des tâches de l'expérience est l'identification des notifications et, bien qu'il s'agisse techniquement de la tâche « principale » pour l'évaluation de la représentation auditive, nous la désignons plutôt par l'expression « tâche de supervision ». En effet la nomenclature usuelle dans ce genre d'expérience préfère désigner l'exercice de distraction comme la « vraie » tâche principale au vu de sa place prioritaire dans le partage de l'attention au cours d'un exercice.

2.3.3 Protocole d'évaluation proposé

Une capture d'écran de l'interface d'évaluation en ligne est visible sur la figure 2.3. L'application commence par afficher une page décrivant l'exercice et rappelant les quelques instructions auparavant annoncées dans l'appel à participations : utiliser une souris, porter un casque audio ou au moins des écouteurs, et faire l'exercice sans bruit environnant. L'expérience est conçue pour durer une demi-heure environ, le temps écoulé entre chaque niveau étant à la discrétion du testeur.

La tâche principale proposée dans cette expérience est basée en grande partie sur celle conçue par Maxime Daniel [46]. Elle consiste à cliquer sur des boutons afin

2.3. Protocole d'évaluation « en laboratoire » pour la comparaison de représentations auditives périphériques

de recopier des séquences de symboles 'X' et 'O' apparaissant dans une interface au milieu de la page. À chaque fois qu'une séquence est recopiée, elle disparaît et une autre est générée et affichée.

<p>Niveau 1/10</p> <p>Lancez la lecture ci-dessous, puis complétez autant de séquences que possible et cochez les anomalies.</p> <table border="1"><tr><td>o</td><td>x</td><td>o</td><td>x</td><td>x</td></tr><tr><td>o</td><td>x</td><td>o</td><td></td><td></td></tr></table> <p style="text-align: center;">x o</p> <p>Séquences réécrites : 3</p> <p>Dernier temps de réécriture : 1.74</p>	o	x	o	x	x	o	x	o			<p><input type="checkbox"/> Melodie</p> <p><input type="checkbox"/> Bourdon</p> <p><input type="checkbox"/> Sonnerie</p> <p><input type="checkbox"/> Cloche</p>
o	x	o	x	x							
o	x	o									

FIGURE 2.3 – Capture d'écran de l'interface d'évaluation en ligne pour la sonification des critères de fabrication additive. Ici pour la représentation *Synth*

Étapes de l'exercice

- Une introduction à l'exercice de tâche principale consistant à recopier 10 séquences avant de pouvoir passer à la suite.
- Introduction à l'exercice d'annotation anomalie par anomalie : écoute de chacun des sons avec des indications textuelles sur les divergences à repérer et le nom de la case à cocher dans ces situations.
- Introduction du son complet de la baseline produite en situation normale.
- Introduction à l'exercice complet passant par 3 niveaux d'entraînement (2 avec anomalies puis 1 sans). Le testeur effectue les passes complètes telles que prévues dans l'exercice d'évaluation à venir : copie de séquences et annotation occasionnelle de notifications. À la fin de chacun de ces niveaux, un retour est présenté au testeur indiquant la proportion d'annotations correctes, ainsi qu'une liste des annotations qui étaient attendues pour le niveau. Il n'est permis de passer à la suite que si l'annotation des anomalies au cours de cet entraînement est réussie à 90%. Un testeur ne remplissant

pas ces conditions peut refaire les niveaux d'entraînement afin d'améliorer son score.

- Exercice d'évaluation : 10 niveaux correspondant chacun à une passe de 30 secondes. 5 passes normales et 5 passes avec anomalies sont réparties aléatoirement dans la présentation de ces 10 niveaux. À la fin de chaque niveau, un écran donne le bilan de la performance du testeur : le pourcentage d'annotation correcte et le nombre de recopies correctement effectuées.
- À l'issue de cette expérience, les testeurs sont redirigés vers un sondage leur permettant de quantifier leur ressenti vis-à-vis de la représentation entendue et son rapport aux deux tâches, principale et de supervision. Un champ de commentaires libres est proposé pour leur permettre de détailler leurs réponses.

2.3.4 Données récoltées

Exercice La page web de l'expérience est codée en **HTML**, **PHP** et JavaScript. Chaque utilisateur se voit attribué au moment de sa connexion un numéro utilisé pour enregistrer les données de sa performance de façon anonyme dans un fichier numéroté. Au début d'un niveau, l'application enregistre le numéro de la passe simulée correspondante afin de pouvoir recouper les résultats pour chaque niveau indépendamment de l'ordre de présentation aléatoire.

À chaque fois qu'une séquence a été recopiée pour la tâche principale, la date en millisecondes depuis la dernière séquence recopiée (ou depuis le début du niveau pour la première) est enregistrée dans le fichier de résultats.

À la fin de chaque niveau, l'application enregistre la date depuis le début du niveau, en millisecondes, à laquelle ont été cochées les cases qui sont restées cochées jusqu'au bout, de sorte que des cases qui auraient été cochées puis relâchées plus tard dans le niveau ne seraient pas considérées comme des annotations.

Sondage final Les sondages d'expérience utilisateur traditionnellement prévus par la littérature [92, 126, 28, 131, 193] nous ont paru trop généralistes pour l'évaluation d'une représentation auditive périphérique. Par ailleurs, dans le cadre d'une expérience en ligne, leur longueur faisait courir le risque de perdre l'engagement de beaucoup de testeurs. Pour cette raison, nous avons plutôt construit notre propre sondage spécifique et concis. Les sept affirmations ainsi proposées (voir le tableau 2.6) visent à orienter les retours des testeurs sur l'influence de la variable « représentation » vis-à-vis du contexte de tâches concurrentes. Pour chaque affirmation, un testeur peut quantifier son approbation

en cochant l'une des cases d'une échelle de Likert à trois degrés : « Un peu », « Moyennement » et « Beaucoup ».

2.3.5 Critères de qualité d'une représentation

À partir des données collectées par l'interface d'expérimentation, nous proposons une collection de quatre critères pouvant être calculés pour l'évaluation de la qualité de la représentation, en nous basant sur une synthèse des recommandations de l'existant pour ce type d'évaluation.

Ceux-ci caractérisent les différents aspects d'une représentation permettant la supervision de données multidimensionnelles (sensibilité et réactivité) dans un contexte périphérique (performance à la tâche principale), et visant à être suffisamment agréable pour ses utilisateurs (retours qualitatifs).

Sensibilité La sensibilité donne une mesure de la qualité des annotations produites par les utilisateurs au cours des tests d'identification, donc de l'intelligibilité des notifications choisies. On obtient cette valeur en calculant l'indice de sensibilité, noté d' , de chaque participant pour chaque notification, en fonction de ses taux de réussites et de fausses alertes [138]. On en calcule ensuite la moyenne sur l'ensemble des testeurs dans chaque condition évaluée, afin d'obtenir une seule valeur indicatrice de la qualité des stimuli.

Cet indice est tel qu'une valeur plus élevée dénote une meilleure intelligibilité pour le stimulus correspondant. Comme il n'est associé à aucune unité ou échelle, l'indice de sensibilité ne peut pas être utilisé pour déterminer la qualité d'un signal dans l'absolu, en dehors du fait que si $d' > 1$ le signal a des chances meilleures que le hasard d'être correctement interprété. Cependant il peut être utilisé pour comparer des signaux différents représentant une même information et évalués de la même façon.

Réactivité La réactivité est donnée par le temps de détection, puis de réflexion, nécessaire à l'annotation correcte d'un signal. Il s'agit donc, comme la sensibilité, d'un indicateur de l'intelligibilité des notifications. On mesure cette grandeur en millisecondes.

Performance à la tâche principale Ce critère consiste en une mesure de la capacité des testeurs à effectuer la tâche principale dans le contexte expérimental. Dans le cas de l'exercice proposé ici, il s'agit de la vitesse moyenne de copie des séquences pour chaque représentation testée.

Retours qualitatifs des testeurs Les données pour ce critère sont récoltées à l'aide d'une série d'échelles de Likert à trois degrés quantifiant l'appréciation des testeurs pour la représentation testée. Les appréciations à quantifier sont conçues pour obtenir une mesure de la nature potentiellement stressante de la représentation, ainsi que son influence sur la façon dont le contexte de travail simulé est perçu.

On obtient une agrégation de ces résultats en faisant leur somme pondérée par des valeurs attribuées aux degrés, puis en divisant par le nombre de réponses. Avec « Un peu » à 0, « Moyennement » à 50 et « Beaucoup » à 100, le résultat est un pourcentage indiquant la quantité d'approbation récoltée par chaque affirmation. On peut ainsi compter combien d'appréciations majoritairement positives ou négatives ont été données à chaque représentation.

Pour les deux critères de performance à l'annotation que sont la sensibilité et la réactivité, on ne compte que les cas où un utilisateur a coché la bonne case après le début de l'anomalie dans la bande-son. Ainsi on évite de tenir compte de « prédictions » d'anomalies, qui seront considérées comme des faux positifs.

On peut classer plusieurs représentations évaluées selon ce protocole en fonction des valeurs moyennes pour chaque critère et en déduire quelle représentation est la meilleure pour le plus grand nombre de critères. Éventuellement, on peut aussi déduire des différents classements obtenus quelle représentation devrait être favorisée en fonction de quels critères sont considérés comme prioritaires.

Ce protocole expérimental a été présenté une première fois dans [243]. Dans ce premier travail, il s'agissait d'abord d'effectuer l'évaluation dans l'absolu d'une seule représentation sans mettre à profit l'aspect comparatif. Pour un exemple de mise en œuvre comparative de cette méthode, voir l'évaluation présentée dans l'article [241] ou dans la sous-section 2.4, appliquée à une problématique corollaire pour laquelle on étudie l'impact de l'incohérence écologique sur la qualité d'une sonification.

2.3.6 Perspectives

Déploiement

À l'origine, ce protocole expérimental devait permettre l'évaluation d'une représentation auditive en laboratoire. Cependant ses deux déploiements ont finalement eu lieu en ligne, chaque testeur effectuant l'exercice dans son environnement propre, ce qui a posé problème vis-à-vis de l'homogénéité des résultats. Pour résoudre ce problème à l'avenir, on espère voir la méthode appliquée

dans son contexte présentiel original, idéalement avec un public sélectionné pour sa connaissance du contexte d'application prévu. Un doctorat devrait d'ailleurs commencer à la rentrée 2022 à ESTIA Recherche, avec pour objectif de produire une évaluation croisée des solutions de représentation auditive proposées avec une autre représentation vibrotactile développée par Sébastien Ibarboure au cours de sa thèse [115].

Méthodologie

La méthode d'évaluation proposée permet déjà dans une certaine mesure d'obtenir des données indicatrices de la qualité d'une représentation et comparables d'une condition à une autre, mais il reste des points sur lesquels elle pourrait être améliorée pour fournir des valeurs plus précises.

Tout d'abord, en fait de protocole d'évaluation d'une représentation à proprement dit, il s'agit plutôt actuellement d'un protocole d'évaluation d'un système de notifications : on ne mesure pas encore l'apprentissage du sens des sons vis-à-vis du processus supervisé. Il s'agit d'une prochaine étape de l'expérimentation qui sera implémentée à l'occasion d'un travail sur un plus long terme, laissant la place à des phases d'entraînement prolongées auprès d'un public familier avec le processus.

Afin d'acquérir plus de données sur la performance d'un utilisateur à une tâche principale, le protocole d'évaluation devrait donner plus d'importance à la mesure de l'erreur dans cette tâche. D'une part, l'application pourrait récolter, en plus des instants de copies réussies, une plus grande quantité d'informations sur les clics ratés, ce qui permettrait de reconstituer une réelle trace de la capacité d'un testeur à se concentrer sur l'exercice. D'autre part il serait utile de rendre la tâche principale globalement plus difficile en encourageant l'utilisateur à commettre le moins d'erreurs possible. On pense par exemple à faire en sorte que la quantité de clics ratés soit pénalisante pour un score obtenu par le testeur au cours de l'exercice. Cette complexification pourrait aussi consister à y ajouter d'autres types d'interactions occasionnelles, de sorte que tout en restant relativement facile, la tâche nécessite réellement un certain niveau de concentration pour être effectuée.

Le fait que la tâche de supervision et la tâche principale sont effectuées par la même main manipulant la même souris induit un biais, lié au temps de déplacement de la souris, dans le temps de réaction mesuré. Ce biais peut dépendre de l'habitude de chaque personne à passer rapidement d'une tâche à une autre dans son interaction avec un ordinateur, voire du dispositif de clic dont chaque personne

dispose (souris de différentes qualités, pavés tactiles). On pourrait envisager soit d'ajouter un calcul préalable de ce biais qui consisterait à faire effectuer des allers-retours chronométrés à l'utilisateur pour un débiaisage ultérieur des mesures, soit de faire en sorte que les deux tâches soient effectuées chacune par une main, avec par exemple une interaction à la souris pour l'une et une interaction au clavier pour l'autre.

Généralement, on remarque qu'il serait utile dans le protocole de produire plus d'expériences témoins évaluant les différentes conditions isolées les unes des autres : mesurer la performance à la supervision sans tâche de recopie, ou la performance à la tâche de recopie sans stimuli sonores, ou la performance à la tâche de recopie avec les stimuli sonores, mais sans tâche de supervision.

Enfin, on pourrait renforcer la simulation du contexte de travail prévu dans l'interface d'évaluation en faisant en sorte que l'audio contienne, en plus du son de la sonification, une version partiellement filtrée et atténuée du son de la fabrication. La tâche principale parallèle à la supervision pourrait elle-même inclure des parasites auditifs induits par l'interaction, simulant par exemple la physicalité d'un objet sur lequel on appuierait sur des boutons, de la même façon que les activités auxquelles pourraient s'adonner les opérateurs ne seraient pas forcément entièrement silencieuses.

Il est prévu de réutiliser ce protocole expérimental adapté selon certaines des remarques ci-dessus dans un travail à venir de Calandra et al. sur la charge cognitive associée à la segmentation de phrases et sections musicales [31].

2.4 Impact de l'incohérence écologique sur la qualité d'une sonification

2.4.1 Cohérence écologique dans l'existant

Traditionnellement, les recommandations pour la conception de sonifications indiquent que les ensembles de sons choisis doivent respecter une certaine cohérence écologique, c'est-à-dire qu'ils doivent être exempts de masquage vis-à-vis des sons de l'environnement d'application prévu, ainsi que distincts les uns des autres dans la perception par une allocation délibérée de la bande passante spectrale pour chacun des éléments [180, 156, 214].

Notre but dans cette étude n'est pas de remettre en question ces précautions anti-masquage, car elles semblent trivialement justifiées. Cependant, une règle supplémentaire souvent implicite que semblent suivre les auteurs de sonifications est que l'ensemble de sons doit correspondre à la définition d'une *soundscape* comme décrite dans le cadre du *World Soundscape Project* de R. Murray

Schafer [178], n'impliquant que des événements et flux sonores qui forment un ensemble logique vraisemblable. Les représentations ainsi conçues correspondent alors à des thèmes, ou « écologies », tels que par exemple la forêt, la plage ou divers genres musicaux.

D'autre part, la recherche récente sur le cheminement mental de l'information auditive suggère que la différenciation de sons se produit à un niveau neuro-physiologique en passant par une hiérarchie de classifications parallèles [155, 136, 132]. D'après une étude de 2016 par Isnard et al. [117], les capacités humaines de séparation des sources sonores seraient si efficaces que des sons naturels, même fortement raccourcis et simplifiés, pourraient toujours être reconnus au milieu d'enchaînements de sons musicaux. On se demande alors s'il pourrait être avantageux de sélectionner des sons appartenant à des catégories écologiques clairement différentes afin de permettre aux processus de classification du cortex de les séparer plus efficacement.

2.4.2 Notre hypothèse

L'hypothèse que nous formulons pour cette expérience est donc la suivante : la sélection de sons appartenant à des catégories écologiques incohérentes permet de construire des systèmes de notifications qui sont plus efficaces, tout en restant périphériques.

Il s'agit en fait d'une hypothèse qui était sous-entendue dès notre première expérience publiée sur ce cas d'étude [242], mais pour laquelle nous n'avions pas encore développé la procédure expérimentale permettant de comparer des représentations (voir la section 2.3).

Il est assez courant que les travaux de sonification pour la supervision proposent plusieurs ensembles de stimuli basés sur des stratégies différentes de représentation. Il s'agit alors souvent d'une dichotomie entre des icônes auditives d'une part et des représentations musicales d'autre part [44, 86, 102, 133, 2]. Il est cependant plus rare que ces travaux proposent une représentation « hybride » consistant en un mélange de sons naturels et musicaux [156, 140] et aucun, à notre connaissance, n'a produit d'évaluation comparative complète de ces représentations au-delà de sondages qualifiant l'expérience utilisateur.

2.4.3 Évaluation comparative d'écologies sonores

En plus de la représentation hybride nommée *Mixte1* introduite et évaluée dans [243], nous avons produit deux écologies sonores cohérentes *Synth* et

Nature (voir la section 2.2). Ces deux écologies ont été conçues à partir des stimuli employés dans *Mixte1*, de sorte à compléter respectivement sa moitié musicale/tonale et sa moitié métaphorique/naturelle.

La procédure expérimentale comparative décrite dans la section 2.3 avait précédemment permis de récolter des données sur *Mixte1*. Dans le cadre d'une nouvelle campagne d'appel à participations, nous l'avons appliquée dans des conditions aussi comparables que possible aux deux nouvelles écologies *Synth* et *Nature*.

Afin d'obtenir le même nombre de participants pour les deux groupes de conditions représentés par ces deux écologies, nous avons ajouté à la page d'évaluation un comptage automatique, au moment de la connexion d'un usager, des participations déjà effectuées de part et d'autre de la condition expérimentale. Selon le groupe le plus représenté, un numéro est attribué à la session, indiquant si les sons et explications doivent traiter de l'écologie *Synth* (1) ou *Nature* (2). Si les deux écologies ont été testées le même nombre de fois à la connexion, le numéro 1 ou 2 est pioché aléatoirement.

2.4.4 Résultats

Participants

- *Mixte1* : 43 personnes (20M, 23F) d'âges entre 18 et 67 ($\mu = 32.3, \sigma = 12.4$).
- *Synth* : 15 personnes (8M, 5F, 2 non spécifiés) d'âges entre 20 et 69 ($\mu = 35.9, \sigma = 15.9$). Quatre ont signalé avoir participé à une itération précédente de l'expérience.
- *Nature* : 16 personnes (7M, 9F) d'âges entre 19 et 62 ans ($\mu = 24.75, \sigma = 10.5$). Deux ont signalé avoir participé à une itération précédente de l'expérience.

Sensibilité La figure 2.4 donne un histogramme des taux moyens de sensibilité obtenus pour chaque type de stimulus (*HC+LC*, *HP*, *TC*, *TP*) dans chaque écologie (*Mixte1*, *Synth*, *Nature*). Ces résultats semblent suggérer les tendances suivantes :

- *HC+LC* : *Gouttes* est nettement moins efficace que *Mélodie* (*Nature* vs. *Mixte1* : $F(1, 56) = 24.37, p < 10^{-4}$; *Nature* vs. *Synth* : $F(1, 29) = 20.45, p < 10^{-4}$). *Mélodie* fonctionne un peu mieux dans *Mixte1* que dans *Synth* ($F(1, 55) = 0.85, p \simeq 0.36$).

- **HP** : *Oiseaux* est nettement plus efficace que *Bourdon* (*Nature* vs. *Mixte1* : $F(1, 55) = 5.76$, $p \simeq 0.02$; *Nature* vs. *Synth* : $F(1, 29) = 4.90$, $p \simeq 0.03$).
- **TC** : les trois écologies ne présentent pas de différence particulièrement marquée (*Synth* vs. *Nature* : $F(1, 29) \simeq 0.00$, $p \simeq 0.95$; *Synth* vs. *Mixte1* : $F(1, 56) = 0.20$, $p \simeq 0.66$). *État Eau* semble fonctionner à peu près aussi bien dans *Mixte1* que dans *Nature* ($F(1, 57) = 0.15$, $p \simeq 0.70$).
- **TP** : *Friture* est détectée de façon plus efficace au sein de *Mixte1* que dans *Nature* ($F(1, 57) = 9.76$, $p < 0.01$).
- Globalement : *Mixte1* est à peu près équivalente à *Synth* ($F(1, 56) = 0.0050$, $p \simeq 0.94$), et *Nature* est un peu moins efficace que les deux autres écologies (*Nature* vs. *Mixte1* : $F(1, 57) = 2.41$, $p \simeq 0.13$; *Nature* vs. *Synth* : $F(1, 29) = 2.04$, $p \simeq 0.16$).

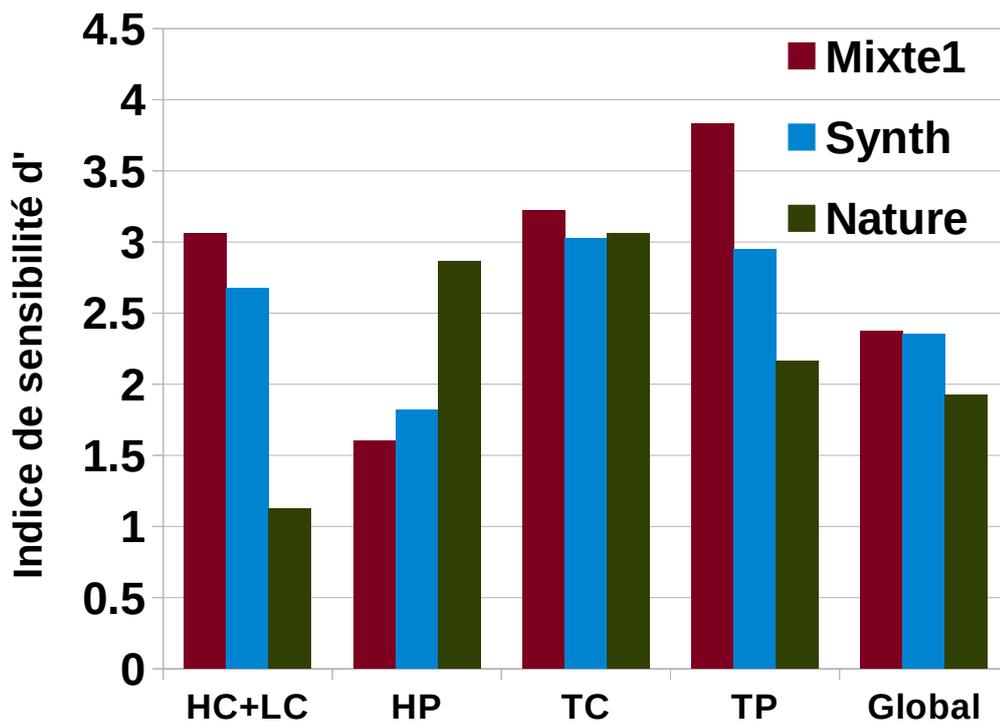


FIGURE 2.4 – Histogramme des indices de sensibilité aux différents stimuli employés dans les trois écologies

Réactivité La figure 2.5 donne un histogramme des temps moyens d’annotation correcte pour chaque type de stimulus (*HC+LC*, *HP*, *TC*, *TP*) et dans chaque écologie (*Mixte1*, *Synth*, *Nature*). En comparant ces valeurs, on remarque que *Synth* semble avoir permis d’obtenir de meilleurs temps de réaction en général (*Synth* vs. *Nature* pour *HP* : $F(1, 82) = 2.74$, $p \simeq 0.10$; *Synth* vs. *Mixte1* pour *HP* : $F(1, 145) = 3.37$, $p \simeq 0.07$; *Synth* vs. *Nature* pour *TC* : $F(1, 128) = 6.10$, $p \simeq 0.01$; *Synth* vs. *Mixte1* pour *TC* : $F(1, 247) = 8.10$, $p < 0.01$; *Synth* vs. *Nature* pour *TP* : $F(1, 19) = 0.24$, $p \simeq 0.63$; *Synth* vs. *Mixte1* pour *HP* : $F(1, 50) = 1.24$, $p \simeq 0.27$)

Seul le stimulus Mélodie (*HC+LC*) a été annoté plus rapidement en moyenne dans le cadre de *Mixte1* que dans *Synth* ($F(1, 297) = 3.90$, $p \simeq 0.05$).

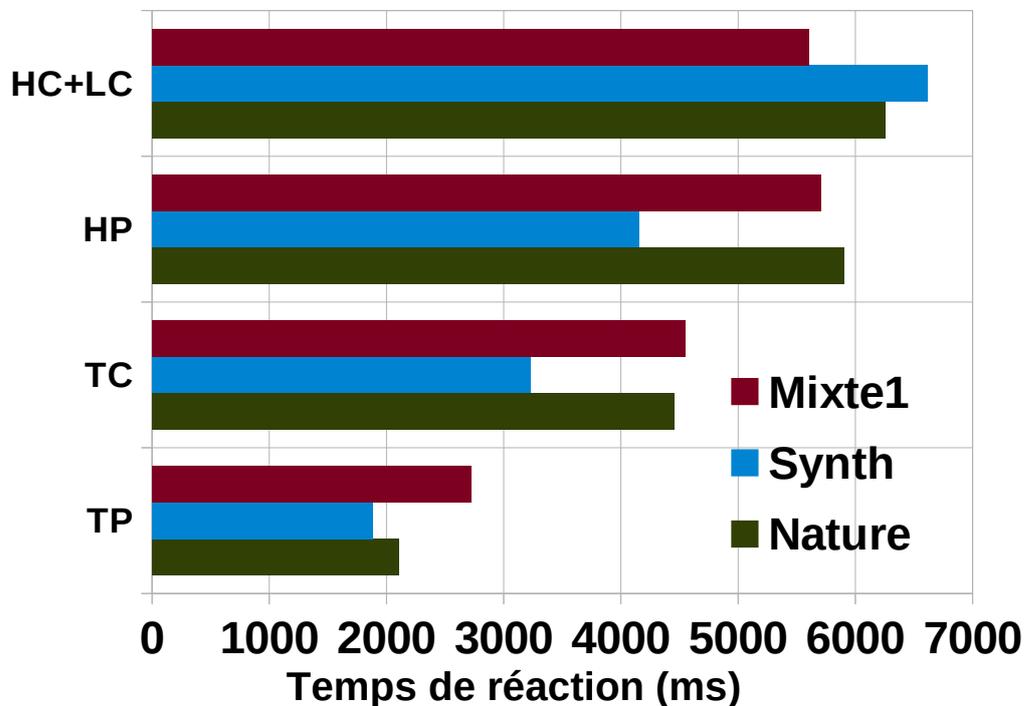


FIGURE 2.5 – Histogramme des temps de réaction aux différents stimuli employés dans les trois écologies

Performance à la tâche principale La figure 2.6 représente sous un format box-and-whiskers les distributions de performances à la tâche principale en fonction de l’écologie testée.

Nature semble avoir produit les meilleurs résultats pour ce critère, suivie par

Synth ($F(1, 29) = 3.56, p \simeq 0.07$). La distribution des scores est beaucoup plus large pour *Mixte1*, l'expérience ayant reçu un plus grand nombre de participations.

Cette distribution semble indiquer une moins bonne performance à la tâche principale pour *Mixte1* que pour les deux autres écologies ($F(1, 57) = 11.86, p < 0.01$ par rapport à *Nature* et $F(1, 56) = 0.67, p \simeq 0.42$ par rapport à *Synth*).

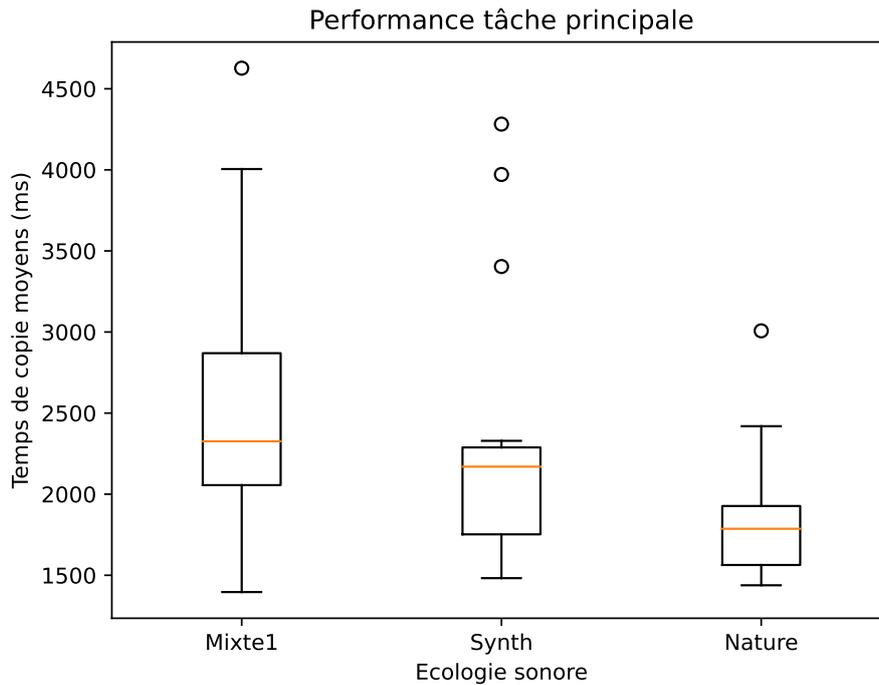


FIGURE 2.6 – Diagrammes box-and-whiskers des vitesses moyennes de recopie des testeurs dans les trois conditions expérimentales

Retours utilisateurs Le tableau 2.6 résume les scores obtenus par chaque écologie pour chaque affirmation. En gras, les scores les plus élevés obtenus pour chaque affirmation. Les scores ayant les connotations les plus positives pour chaque affirmation sont surlignés en vert et les scores ayant les connotations les plus négatives sont surlignés en rouge.

Commentaires libres À la fin de l'évaluation, les testeurs sont encouragés à soumettre librement toute remarque sur l'exercice. Les affirmations proposées dans le sondage sont prévues comme un moyen d'orienter leurs réflexions sur la relation entre l'exercice réalisé et la bande son entendue. Nous mettons donc en parallèle

2. Supervision du procédé de fabrication additive wire-arc

les scores d'approbation attribués aux affirmations dans le tableau 2.6 avec ces retours libres.

	<i>Mixte1</i>	<i>Synth</i>	<i>Nature</i>
J'ai trouvé les éléments sonores faciles à distinguer	71.43	82.14	83.33
Le son m'a gêné.e pour faire l'exercice	7.14	25.00	14.29
J'ai trouvé le son stressant	19.05	42.86	30.00
J'ai trouvé l'exercice stressant	16.67	25.00	16.67
J'ai trouvé l'exercice amusant	66.67	46.43	73.33
J'ai trouvé l'exercice compliqué	14.29	21.43	13.33
L'exercice m'a empêché.e de suivre le son	11.90	21.43	3.33

TABLE 2.6 – Résumé des réponses au sondage qualitatif

Généralement, *Synth* semble avoir été considérée comme la plus stressante des écologies (lignes 2-7) y compris, en particulier, dans le ressenti des utilisateurs vis-à-vis de la tâche principale (ligne 4).

Dans un commentaire libre sur l'écologie *Synth*, un testeur a signalé avoir trouvé le stimulus *Bourdon* trop sensible. Ce problème avait aussi été signalé dans l'évaluation initiale pour *Mixte1*. Il est vraisemblable que la plupart des erreurs pour ce stimulus aient ainsi été causées par des fausses alertes plutôt que des faux négatifs.

Les testeurs pour l'écologie *Nature* ont eu tendance à considérer ses stimuli comme faciles à distinguer (ligne 1), malgré une performance moindre à la tâche d'annotation. Cela pourrait être causé par le fait que, même si les sons eux-mêmes sont faciles à différencier, l'évolution de leurs paramètres perceptifs en cas d'anomalies ont pu être plus difficiles à détecter que dans les deux autres écologies. En effet, deux participants ont signalé dans leurs commentaires libres avoir eu des difficultés à différencier les situations normales et anormales pour le stimulus *Gouttes*, un problème clairement visible dans sa mesure de sensibilité (figure 2.4) et qui a vraisemblablement été rencontré par plus de participants que ceux qui l'ont signalé.

Pour *Mixte1* comme pour *Nature*, un certain nombre de testeurs ont signalé leur difficulté à différencier les stimuli *Friture* et le son d'ébullition d'*État Eau*. Malgré les différences perceptives objectives observables entre ces deux sons (*Friture* se situe dans une bande de fréquences plus élevée et présente une enveloppe percussive caractéristique tandis que l'ébullition est un bruit continu de bulles), il persiste que les deux sont perçus essentiellement comme des sons d'eau en ébullition.

Les sons employés dans *Mixte1* ont plus souvent été notés comme difficiles à différencier (ligne 1). Cela peut être dû au fait que cette écologie combine les défauts des deux autres en termes de confusion entre *Friture* et *État Eau*, et d'hypersensibilité du stimulus *Bourdon*.

On remarque que les écologies impliquant des sons naturels ont eu tendance à être jugées de façon plus positive :

- Les stimuli pour *Nature* sont considérés comme plus faciles à distinguer (ligne 1).
- *Mixte1* est l'écologie la moins distrayante et la moins stressante (lignes 2–3).
- *Mixte1* et *Nature* sont à égalité pour l'influence la moins stressante sur la tâche principale (ligne 4).
- La tâche principale est considérée comme plus amusante et moins difficile quand la bande son utilise *Nature* (lignes 5–6).
- L'écologie *Nature* est considérée comme plus facile à suivre en s'occupant de la tâche principale (ligne 7).

Cette préférence pour l'écologie *Nature* semble confirmer les arguments souvent avancés quant au caractère moins agaçant des icônes auditives comparées aux notifications musicales [82, 80].

2.4.5 Conclusion et perspectives

D'après les classements des différentes écologies selon les différents critères, on constate que :

- *Nature* permet une meilleure performance à la tâche principale et un meilleur ressenti des utilisateurs.
- *Mixed* permet une meilleure sensibilité aux stimuli.
- *Synth* permet une meilleure réactivité aux stimuli.

Ces résultats sont partiellement encourageants par rapport à l'hypothèse formulée en 2.4.2 puisque l'écologie incohérente *Mixte1* semble bien permettre une meilleure sensibilité de l'utilisateur aux stimuli sonores. Cependant les critères qualifiant l'aspect périphérique (performance à la tâche principale et retours utilisateurs) semblent plutôt favoriser l'écologie cohérente *Nature*. Nous ne pouvons donc pas affirmer que notre hypothèse de départ est vérifiée.

On remarque que si l'écologie *Nature* présente des résultats globalement moins bons que les deux autres en termes de discrimination des stimuli, son stimulus *Oiseaux* a été beaucoup plus efficace que *Bourdon* pour produire des notifications sur l'état du critère *HP*. Si l'on considère que les écologies incohérentes présentent effectivement un intérêt pour les systèmes de notifications ambiants, on peut alors se demander si une meilleure écologie de type *Mixte* pourrait être construite en utilisant ce stimulus naturel à la place du stimulus musical.

Plus généralement, afin de pouvoir déterminer d'une façon aussi exhaustive que possible quelle combinaison de stimuli permet la meilleure détection des notifications il faudrait encore expérimenter sur toutes les combinaisons possibles avec les différents stimuli proposés. En faisant cela, on pourrait aussi étendre l'expérience à l'hypothèse que des dimensions physiques de même nature gagnent à être représentées par la même catégorie de sons. En particulier, on vérifierait ainsi si les critères *TP* et *TC* seraient mieux différenciés si leurs sons appartenaient à des catégories entièrement différentes (avec nos exemples : une combinaison *Cloche* + *État Eau* ou *Friture* + *Sonnerie*).

Dans le cadre de ce travail, nous avons conçu des ensembles de stimuli que nous considérons comme représentatifs de catégories sonores respectivement musicales et naturelles. Si l'expérience effectuée semble favoriser l'écologie incohérente *Mixte1* pour la sensibilité aux notifications, on garde cependant à l'esprit que ces résultats peuvent ne refléter que nos propres choix de design, encore imparfaits, plutôt que les choix écologiques à proprement dit. C'est pourquoi nous espérons qu'à l'avenir un corpus plus large d'expériences comparant l'efficacité de représentations écologiquement cohérentes et incohérentes verra le jour et qu'une conclusion plus certaine pourra en être tirée sur l'impact de l'incongruité dans les systèmes de notifications.

2.5 Perspectives pour la réalisation d'un procédé wire-arc sonifié

Bien que nous ayons effectué toutes nos expérimentations sur des fichiers traces simulés sous un format *CSV* et transformés en sons par différentes variantes d'un script *SuperCollider*, un des buts de ce cas d'étude est de se projeter dans un avenir où la captation en temps réel des critères de fabrication sera réellement possible. Le programme de sonification permettant leur supervision pourrait alors être adapté à partir des prototypes *SuperCollider* sur lesquels nous avons expérimenté au cours de cette thèse.

Cette perspective donne un exemple concret de motivation pour le format de modèle générique que cette thèse vise à définir (voir d'ailleurs sa réécriture d'après ce format dans le chapitre 8) : offrir une représentation à la fois synthétique et exhaustive des opérations pour la sonification afin de pouvoir transmettre le design à quelqu'un d'autre dont la tâche serait de la recréer, potentiellement plusieurs années plus tard ou avec d'autres outils.

2.6 Conclusion générale

Pour ce cas d'étude, nous avons travaillé sur la conception d'une sonification pour la supervision auditive d'un procédé industriel de fabrication additive wire-arc.

Nous avons proposé une série de solutions consistant en des itérations successives d'une sonification par parameter mapping. Le principe de base de ces représentations était de prendre la forme d'un mélange de plusieurs flux sonores combinant diverses propriétés perceptives pour aider l'auditeur à les distinguer sans nécessiter de trop grands efforts d'attention : la moitié des sons est active constamment dans un état de repos et l'autre moitié est activée seulement en cas d'alerte, la moitié des sons consiste en des notes de musique jouées par des timbres synthétiques et l'autre moitié en des enregistrements de sons naturels, la moitié des flux sonores est continue et l'autre moitié est ponctuelle.

Le besoin de comparer différentes solutions successives nous a amenés à décrire un protocole expérimental spécifique pour l'évaluation en laboratoire d'un système de notifications auditives périphériques. Ce protocole permet de récolter des informations sur la détection (et potentiellement l'intelligibilité) des données, sur la distraction induite dans un contexte de travail, et sur le ressenti subjectif des utilisateurs. Pour des raisons de crise sanitaire à l'époque des expériences, celles-ci n'ont pas réellement pu être effectuées en laboratoire. À la place, les testeurs ont été invités à se connecter à une page web depuis chez eux, dans des conditions de tests qui ne pouvaient pas réellement être contrôlées.

Ce protocole a ensuite servi pour expérimenter sur l'importance de la cohérence écologique pour la qualité d'une sonification en comparant les scores obtenus par différentes sonification construites, ou non, à partir de sons logiquement cohérents. Si les résultats suggèrent un léger avantage dans la détection de notifications au sein d'une sonification incohérente, celle-ci n'est cependant pas la plus favorable aux autres critères d'évaluation. Il serait par ailleurs nécessaire d'expérimenter

à l'avenir dans des conditions plus homogènes afin d'obtenir des réponses plus robustes à la question soulevée par cette expérience.

Nos réflexions sur la sonification de la fabrication additive wire-arc ne constituent encore qu'un des premiers pas vers la conception d'un système complet de supervision augmentée pour ce procédé. Il sera encore nécessaire à l'avenir d'implémenter de réels capteurs pour les données permettant la supervision, de créer un logiciel autonome effectuant la sonification (potentiellement inspirée des prototypes sous *SuperCollider* décrits dans ce chapitre) à partir du flux de données généré par ces capteurs, d'ajuster les choix de mappings aux nouvelles conditions induites par ces progrès, et d'effectuer une évaluation en situation réelle du résultat, en particulier combiné aux représentations développées en parallèle pour les autres modalités, entre autres haptique.

Chapitre 3

Sonification immersive du suivi de protéines individuelles pour l'étude de la motilité cellulaire

Ce cas d'étude porte sur la construction d'un outil d'observation des protéines en mouvement lors de l'activité cellulaire. Les données à sonifier sont acquises à l'aide des techniques modernes d'imagerie nanoscopique à super-résolution, et des analyses préalables permettent de les exprimer dans un format propre à l'étude de leurs trajectoires individuelles (*Single-Particle Tracking* en anglais).

Dans la section 3.1, nous situons la problématique de représentation dans son contexte de recherche et donnons les motivations pour le passage à une modalité auditive. Ensuite, dans la section 3.2 nous étudions des projets similaires tirés de la littérature qui nous permettent d'extraire des pistes à explorer pour notre propre outil d'exploration, qui est alors décrit dans la section 3.3. Nous décrivons dans la section 3.4 la façon dont nous avons conçu un prototype pour cet outil dans l'interface existante FluoSim. Ce prototype n'a pas encore fait l'objet d'une évaluation, mais nous décrivons dans la section 3.5 les suggestions qui ont déjà été faites pour l'améliorer, ainsi qu'un plan pour un protocole expérimental permettant d'étudier l'efficacité de la représentation proposée dans des problèmes de localisation.

3.1 Contexte

3.1.1 Affiliations

Ce cas d'étude a initialement été soumis à notre attention par Gregory Giannone et Olivier Thoumine de l'IINS. Par la suite, il a aussi impliqué Matthieu

Lagardère en tant que développeur du programme de simulation et analyse des trajectoires *FluoSim* [129] dans lequel nous avons intégré un prototype de la sonification proposée.

L’Institut Interdisciplinaire de NeuroSciences (IINS) est une unité mixte de recherche rattachée au CNRS et à l’université de Bordeaux, qui conjugue des compétences interdisciplinaires pour étudier les mécanismes moléculaires et cellulaires impliqués dans l’activité cérébrale. Un des axes de recherche majeurs de l’IINS est le développement de nouvelles méthodes d’imagerie, de biocapteurs et d’outils biomimétiques pour la mesure et la perturbation des fonctions biologiques au niveau subcellulaire.

L’étude de la représentation auditive du comportement cellulaire a déjà fait l’objet, pendant quelques années, du projet *Turning Molecular Behavior into Architecture Shape Sound* (TMBASS) mené par l’IINS, le Centre Interactif d’Art Cognitif (CIAC) et le SCRIME. Deux outils de sonification précédemment conçus dans le cadre de ce projet sont étudiés dans la section 3.2.

3.1.2 Domaine

Ce cas d’étude s’inscrit dans la recherche sur la motilité cellulaire à l’aide de techniques d’imagerie nanoscopique à super-résolution.

À cause du phénomène de diffraction de la lumière, il existe une limite physique à la résolution pouvant être atteinte par des moyens de microscopie optique traditionnelle. Dans le domaine de la lumière visible, la plus petite résolution possible est ainsi d’environ 250nm. Un certain nombre de techniques modernes permet cependant de dépasser cette limite en faisant appel à d’autres phénomènes physiques. Par exemple la technique *PhotoActivated Localization Microscopy* (PALM) décrite par Betzig et al. en 2006 [18], consiste à marquer les protéines à suivre avec une molécule fluorescente activée par le balayage d’un laser. Les points lumineux ainsi activés sont toujours flous, mais peuvent plus facilement être séparés en faisant en sorte que chacune des protéines soit excitée très brièvement, devenant alors visible individuellement.

Le logiciel *MetaMorph*¹ est utilisé pour piloter le matériel (microscope, lasers et caméra) nécessaire à de telles captures, ce qui permet d’acquérir des piles d’images de molécules individuelles. Ensuite, la macro spécifique *PALM-Tracer*², écrite pour

1. <https://fr.moleculardevices.com/products/cellular-imaging-systems/acquisition-and-analysis-software/metamorph-microscopy> – visité le 4/1/23

2. <https://neuro-intramuros.u-bordeaux.fr/displayresearchprojects/70/11> – visité le 4/1/23

3. Sonification immersive du suivi de protéines individuelles pour l'étude de la motilité cellulaire

MetaMorph par Jean-Baptiste Sibarita et son groupe, permet de reconstituer les trajectoires de molécules individuelles. On obtient ainsi des structures de données indiquant les positions successives de chaque point identifié, au cours du temps, pour une résolution temporelle pouvant atteindre 20 à 50 Hz et une résolution graphique allant jusqu'à 20nm. La figure 3.1 montre un exemple d'image obtenue.

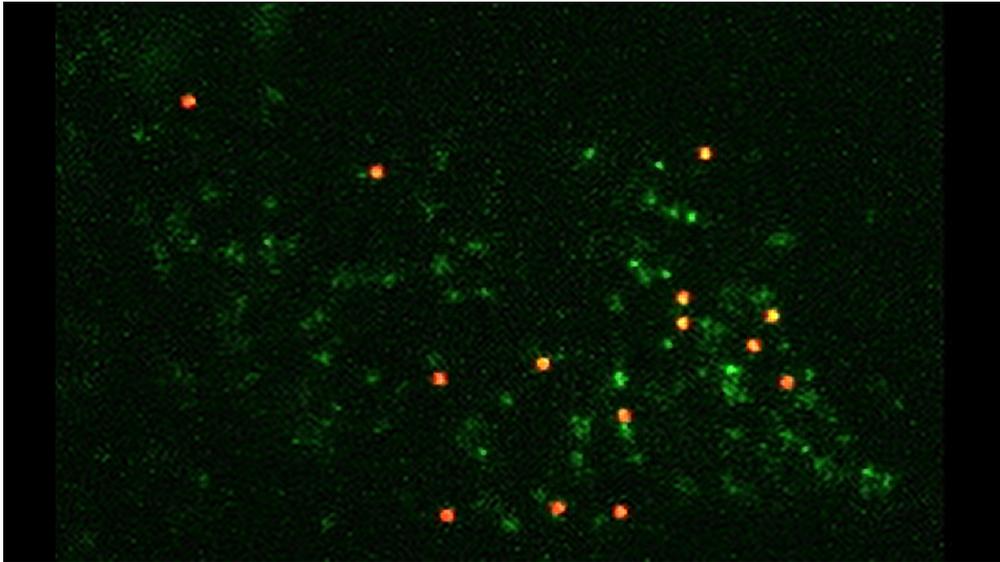


FIGURE 3.1 – Une trame de vidéo représentant les molécules au sein d'une cellule

L'étude de ces trajectoires est appelée *Single-Particle Tracking* (**SPT**) en anglais, et consiste à effectuer des opérations d'analyse statistique sur les données de positions des protéines au cours du temps. Ces techniques de visualisation et l'étude par **SPT** permettent de comprendre comment les interactions dynamiques entre molécules d'adhésion (intégrines), cytosquelette d'actine et leurs régulateurs contrôlent la motilité et la morphologie cellulaire [187, 165, 38, 37, 77]. Elles permettent entre autres d'obtenir des mesures de forces à l'échelle protéique dans les cellules [85, 118].

Le taux de diffusivité, noté D , est une donnée caractéristique de la trajectoire d'une protéine qui entre souvent en jeu pour déterminer son type de mouvement, et à laquelle on s'intéresse particulièrement dans le cadre de la représentation. Il s'agit d'une grandeur en $\mu m^2/s$ qui peut être calculée approximativement à partir de la surface moyenne couverte par la protéine au cours de sa trajectoire (**Mean Squared Displacement** (**MSD**)). À un instant t : $MSD(t) = 4Dt + MSD(0)$ [191].

Un passage obligé de l'analyse de données est leur observation, et c'est à ce niveau que se situe notre travail dans ce cas d'étude. Dans les articles sur l'étude de la motilité cellulaire, il n'est pas inhabituel de voir des illustrations montrant l'organisation des trajectoires selon des codes couleurs sur leurs taux de diffusivité ou leurs types de mouvements [187, 165, 38]. Cela semble indiquer qu'une intuition première consiste à rechercher des motifs dans les représentations visuelles des données afin de formuler des hypothèses qui pourront ensuite être prouvées par des calculs plus objectifs. Ainsi, la motivation de départ de ce travail est d'étudier la possibilité de fournir des intuitions similaires, voire supplémentaires, à l'aide d'une nouvelle modalité d'observation auditive de ces données.

3.1.3 Motivations pour une sonification

La représentation auditive des données de **SPT** est motivée par plusieurs axes décrits dans les paragraphes ci-dessous : tirer parti d'une représentation multimodale des données, mettre à profit la sensibilité de l'ouïe à l'évolution temporelle, offrir un support à la vulgarisation scientifique pour le domaine d'application et contribuer à l'accessibilité du travail de recherche pour un public non voyant.

Utiliser le son comme une modalité supplémentaire à la vue

En mettant en œuvre plusieurs canaux perceptifs, ici la vue et l'ouïe, de façon à produire des stimuli cohérents les uns avec les autres, on espère augmenter la détection d'information pour les deux, tout en réduisant la charge attentionnelle nécessaire pour l'un ou l'autre. Des expériences par Spence et al. ont prouvé que, sous réserve d'un design soigneux, la multimodalité audiovisuelle présente un avantage sur les modalités exclusivement auditives ou visuelles [202, 112].

Permettre la détection de mouvements concertés de protéines

La représentation auditive est particulièrement appropriée pour des domaines d'application dans lesquels les données présentent des comportements périodiques, rythmiques ou temporels, tant que ceux-ci restent dans les domaines de perceptibilité du tempo [119, 120]. Or, des travaux de Giannone et al. ont montré le caractère périodique de certains phénomènes de motilité cellulaire [83, 84], ce qui semble encourageant quant à la possibilité de détecter des motifs de ce type au cours d'observations auditives des trajectoires.

Offrir une représentation ludique de l'analyse des données pour SPT

Le fait de présenter des résultats d'expériences scientifiques d'une façon inhabituelle qui invite à l'interaction, l'interprétation et la discussion peut être un bon moyen de disséminer la science auprès de publics non scientifiques. Schertenleib et Barrass en particulier ont défendu le potentiel de la sonification comme support pour la vulgarisation scientifique [190].

Contribuer à l'accessibilité du domaine de recherche

L'un des grands enjeux de la sonification est la conception d'interfaces auditives pour le public non voyant [64]. Les chercheurs en biologie ne disposant actuellement d'aucune alternative largement adoptée pour l'observation visuelle des captures à super-résolution, il est intéressant de s'interroger sur la pertinence d'une autre modalité, ici auditive, pour l'inclusion de personnes non voyantes dans ce domaine.

3.2 Existant

Cette étude de l'existant porte d'abord sur des outils de sonification déjà produits sur le même sujet dans le cadre de projets antérieurs, puis sur l'état de l'art plus général pour des situations nécessitant la sonification de grandes quantités de données positionnées dans un espace et évoluant au cours du temps. Les aspects clés de l'existant retenu ici sont « grandes quantités de données », « temporalité » et « spatialisation ».

3.2.1 Pour le projet TMBASS

DoReMolecule Le projet *DoReMolecule* a été réalisé en 2015 par un groupe d'étudiants de l'ENSEIRB-MATMECA [5] selon une optique arts et sciences, visant à concilier les visions de deux interlocuteurs venant des domaines de recherche en neurosciences et en musique. Le choix entre deux modes d'utilisation, « Chercheur » ou « Compositeur » est ainsi présenté dès le lancement du programme. La principale différence entre ces deux modes est dans la façon dont l'utilisateur peut sélectionner des régions de l'espace dans lesquelles les protéines seront représentées par différents timbres.

La sélection de zones d'intérêt et leur représentation par différents timbres jouent en effet un rôle très important dans la présentation de cet outil. L'utilisateur peut effectuer une sélection manuelle des régions à représenter selon différents instruments de musique. Les trajectoires disponibles pour la représentation sont ensuite listées selon un classement en fonction de leurs types de mouvements (immobile, diffusif, aléatoire, directionnel), de sorte que l'utilisateur puisse

sélectionner celles qui doivent être utilisées ou pas dans la sonification. Les types de mouvements sont caractérisés par un ensemble de seuils sur le taux de diffusivité et sont associés dans le son à différents types d'évènements musicaux (glissandi, notes soutenues, arpèges comprenant plus ou moins de notes).

Dans cet outil, la sonification se fait séparément de la sélection des paramètres. L'utilisateur définit depuis une première interface graphique les zones de données à sonifier et les paramètres de timbres. Dans une deuxième il peut choisir les trajectoires à représenter selon différents critères. Ensuite seulement il est possible de demander à lancer un rendu en partition **MIDI**, ce qui ouvre une autre interface d'exploration du résultat audio-visuel avec un contrôle du positionnement temporel dans la lecture. L'absence de spatialisation rend l'interprétation des flux sonores entendus assez difficile, même avec un rendu visuel simultané. Le fait d'utiliser le format **MIDI** facilite la manipulation de timbres riches, mais rend la quantité de contenu sonore plutôt limitée (16 pistes simultanées max) et oblige à régler le déroulement temporel selon un tempo bridé à 300 bpm, soit un intervalle entre chaque trame d'environ 0.02s, qui ne peut pas être modifié dynamiquement.

GPseyeSONIC Le programme *GPseyeSONIC* a été développé au **CIAC** dans le cadre du projet **TMBASS**. Il subsiste assez peu de ressources en ligne sur ce projet qui avait pourtant gagné le prix d'excellence 2016 de l'université de Bordeaux. L'essentiel des remarques dans ce paragraphe sont faites d'après l'observation des vidéos de démonstration toujours disponibles sur YouTube³.

L'outil semble mettre à profit la répartition spatiale des protéines, à la fois dans la spatialisation des sources sonores et dans un mapping aux hauteurs en fonction de la quantité de mouvement d'une trame à la suivante. L'intensité de la fluorescence semble par ailleurs être utilisée pour déterminer l'intensité du son joué.

L'objectif de ce projet est apparemment plus artistique que scientifique. En effet le procédé de sonification semble passer par la production d'une partition à partir des données sélectionnées, puis son interprétation par un musicien humain, ce qui n'est pas optimal pour en faire un usage quotidien dans un contexte de recherche. Les vidéos de démonstration apportent très peu d'informations sur les mappings effectués, au profit de remarques sur la qualité esthétique des compositions obtenues par le mapping d'évènements musicaux à la réalité physique de la motilité cellulaire. Si l'intelligibilité ne semblait pas être le but premier de la démarche, on note tout de même le souci d'immersion qui a amené à effectuer une spatialisation des sources sonores en fonction des points, ainsi que la construction de rendus audiovisuels indiquant de façon sélective quelle protéine est à l'origine de quelle partie du son.

3. Par exemple : <https://youtu.be/FtguevX57fk> – visité le 4/1/23

3.2.2 Projets similaires

Interactive Software for the Sonification of Neuronal Activity Verrier et al. [213] ont proposé en 2020 une sonification pour un projet relativement similaire au présent cas d'étude. Il s'agit dans leur cas de sonifier une technique de visualisation par fluorescence de l'activation des neurones : l'imagerie au calcium. On se place ainsi à un plus haut niveau en étudiant les neurones eux-mêmes plutôt que leur contenu macro-moléculaire. De façon similaire à notre cas d'étude, la démarche a pour but de faciliter la détection de motifs spatiaux et temporels, ainsi que des phénomènes synchrones, dans des données temporelles et spatiales. Cependant, plutôt que de s'intéresser aux mouvements des objets étudiés, on s'intéresse à leurs fluctuations de fluorescence.

La technique de sonification proposée consiste en une compression de la quantité de données à représenter par une sélection manuelle des régions d'intérêt dans une carte visuelle des neurones. Les taux de fluorescence moyens de ces régions d'intérêt sont alors associés à des paramètres de synthèse FM. Les positions des données ne sont pas traduites par une spatialisation du son. Cependant l'outil permet de paramétrer la sélection des données en temps réel du rendu sonore, ce qui contribue à ce que les phénomènes dans le retour audio-visuel puissent plus facilement être associés aux phénomènes groupés d'activation des neurones.

Rotator Autre exemple d'outil de sonification permettant une sélection de la zone spatiale d'intérêt des données, *Rotator* [41] présente la particularité d'être plus générique que les autres outils listés jusqu'ici en permettant l'utilisation de n'importe quelles données qui présentent à la fois une évolution dans le temps et une position dans l'espace.

L'interface permet de définir une zone d'intérêt englobant plus ou moins de sources de données réparties dans un espace – par exemple des mesures physiologiques à différents emplacements du corps humain – et de décider de modes de représentation pour ces données en combinant diverses techniques génériques de sonification et de visualisation.

Cependant, si l'outil permet de gérer la représentation de sources de données multiples situées dans un espace, il ne semble pas particulièrement prévu pour une quantité de points particulièrement grande, tels que les « essaims » de trajectoires étudiés en **SPT**. Par ailleurs, la position géographique des points fait bien partie des paramètres gérés par l'outil, mais il ne semble pas prévu que celle-ci puisse évoluer au cours du temps.

3.2.3 Réflexions sur la model-based sonification

Le besoin de représentation de grandes quantités de points dans un espace rappelle aussi diverses techniques d'exploration auditive de type **MBSon**, comme *Tangible Data Scanning* [24] dans laquelle l'interaction permet de se situer dans un espace de points, ou *Principal Curve Sonification* [106] dans laquelle un auditeur virtuel se déplace sur la courbe principale d'un ensemble de points en déclenchant le son caractéristique et spatialisé de chacun au fur et à mesure qu'il les croise.

Cette technique de sonification serait la plus indiquée traditionnellement pour représenter de grandes quantités de points dans un espace, mais elle présente ici le défaut de ne pas prévoir que les données aient une temporalité propre [98]. La temporalité du son créé est plutôt celle induite par l'interaction de l'utilisateur (qui est aussi la temporalité de la simulation physique qui produit le son) et son évolution est basée sur le sondage de l'espace des données, pas sur l'évolution des données elles-mêmes. Si on reprend donc dans le présent travail certains aspects importants de la **MBSon** que sont l'interactivité, l'immersion et le sondage de l'espace, c'est pourtant un autre type de sonification hybride plus proche de la parameter mapping sonification qui est proposé.

3.3 Sonification proposée

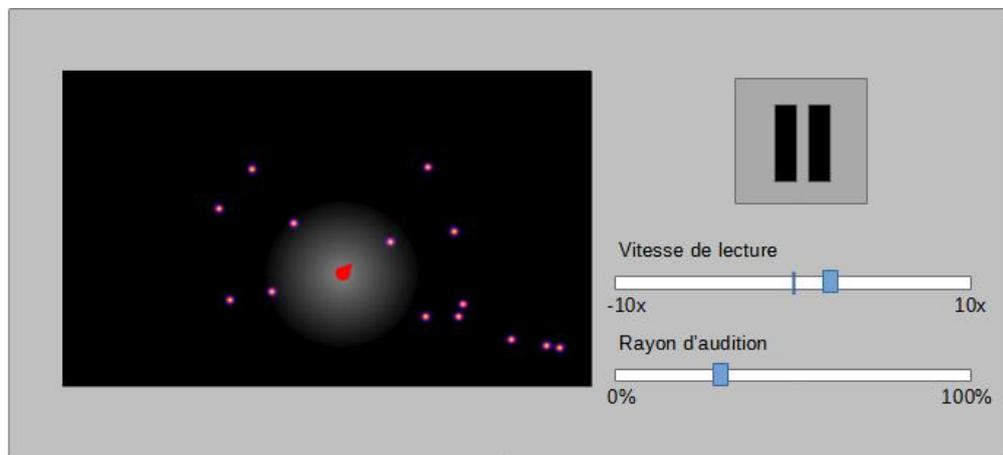


FIGURE 3.2 – Maquette de l'outil de sonification immersive.

3.3.1 Sonification interactive et immersive

L'outil de sonification proposé se base sur l'idée d'immerger l'utilisateur dans l'espace des protéines. Suivant un principe souvent appliqué dans des sonifications

de type model-based, on fait en sorte de fournir l'information de la répartition spatiale des points en lui associant la spatialisation des sources sonores autour de l'utilisateur. Une maquette de cet outil est visible sur la figure 3.2 : l'avatar de l'utilisateur au sein de la cellule consiste en une position (rond rouge), une orientation (triangle rouge sur le rond), et un rayon d'audibilité (cercle flou blanc) contrôlables par l'utilisateur, à l'aide respectivement des touches du clavier et du curseur « Rayon d'audition ». Un point à un instant donné constitue une source sonore, selon un ensemble de mappings à des paramètres perceptifs décrit ci-dessous. L'utilisateur parcourt l'enchaînement temporel des images à l'aide d'un lecteur vidéo dont la vitesse est réglable à l'aide d'un curseur (« Vitesse de lecture » sur la maquette).

3.3.2 Mappings perceptifs

Notre stratégie de représentation des points consiste à mettre à profit autant que possible des paramètres perceptifs dits « bas niveau » traditionnellement employés en sonification (hauteur, volume, timbre, position dans l'espace) de sorte que toutes les informations caractéristiques d'un point puissent être présentées dans sa transposition en son.

Ainsi pour chaque point, à chaque instant :

- Le taux de diffusivité instantané normalisé entre 0 et 1 est associé à la hauteur du son selon un mapping à une fréquence brute. Le premier degré de fréquence fondamentale supérieure à cette valeur est ensuite sélectionné dans la gamme diatonique majeure de la note la plus basse autorisée, jusqu'à la note la plus haute autorisée. La figure 3.3 illustre cette sélection de la hauteur pour une note minimale La_4 , une note maximale Mi_6 et un taux de diffusivité de 0.58.
- Le taux de fluorescence et la distance à l'auditeur virtuel sont associés au volume, chacun informant un facteur d'amplitude entre 0 et 1, de sorte qu'une fluorescence plus forte donne un volume plus élevé et une distance plus grande donne un volume plus bas.
- L'angle du point par rapport à l'orientation de l'auditeur virtuel est associé à une spatialisation binaurale de la source sonore correspondante.
- Le choix du timbre est laissé à l'utilisateur dans l'interface. Dans un premier temps, on se contente d'un ensemble de timbres correspondant à quelques formes d'ondes basiques : sinusoïdale, carrée, dents de scie, triangulaire. Pour plus tard, il est envisagé de mettre en place une réelle synthèse sonore

de timbres d'instruments de musique. La plupart des options permettent de sélectionner un seul timbre pour caractériser tous les points, mais une option expérimentale supplémentaire permet d'effectuer une sélection du timbre pour chaque trajectoire en fonction de son taux de diffusivité global. Cette sélection est alors faite selon une série de seuils définissant différents types de mouvements, comme ceux utilisés dans *DoReMolecule* [5].

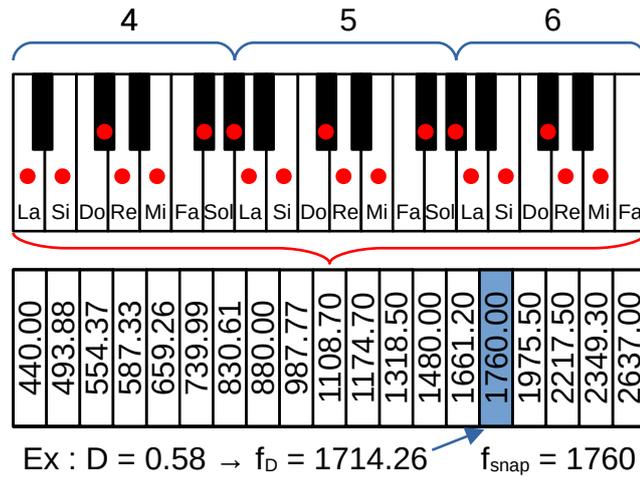


FIGURE 3.3 – La sélection de la hauteur d'après la gamme diatonique majeure de la note la plus basse choisie, et jusqu'à la note la plus haute choisie. Ici entre La_4 et Mi_6 , et pour un taux de diffusivité $D = 0.58$.

Cette représentation se veut un compromis entre la précision vis-à-vis de la teneur des données et la généralisation vis-à-vis de leurs comportements dans l'espace. En effet d'une part on veut que chaque protéine soit représentée par un son ponctuel caractéristique de toutes ses propriétés, mais d'autre part il n'est pas prévu que l'observation selon cette modalité permette d'extraire les propriétés exactes de chaque point individuellement. Une telle détection serait à peu près impossible d'après les capacités humaines de détection de flux sonores [25], d'autant moins si l'on permet que plusieurs points distincts possèdent la même hauteur et le même timbre. Le but est plutôt de permettre à l'utilisateur de se déplacer au sein d'un nuage de sons paramétrés pour tenter de détecter des tendances dans l'évolution de leur répartition spatiale au cours du temps.

3.3.3 Mode sonde

On propose en plus un mode d'interaction dit « sonde » qui permet, quand la vidéo est en pause, de lancer une représentation selon une version simplifiée du

principe de sonogramme tel que décrit par Hermann et Ritter [107]. Un front d'onde se propage en cercle autour du point de départ donné par l'utilisateur et applique une « excitation » aux points rencontrés sur son passage. Au lieu de construire le son, comme dans le modèle de Hermann, en suivant l'évolution d'un ensemble de ressorts dont les raideurs seraient définies à partir de propriétés des données, on fait simplement en sorte que chaque point, quand le front d'onde le rencontre, produise un bref son percussif indicateur des paramètres de hauteur, timbre, volume et emplacement spatial qui lui ont été attribués. Une telle représentation doit permettre de consulter la répartition des points de différentes natures dans l'espace autour d'une position donnée, indépendamment de leur évolution dans le temps.

3.4 Implémentation dans FluoSim

*FluoSim*⁴ est une application de simulation, exploration et analyse d'images à super-résolution développée par Matthieu Lagardère en 2020 [129]. Ce programme présente l'avantage d'être open source et de déjà disposer d'une interface graphique complète permettant l'affichage des protéines au cours du temps, d'un moteur de gestion des données de trajectoires incluant les calculs de taux de diffusivité, et de nombreuses fonctionnalités pour la génération de trajectoires simulées. Pour le prototypage de la modalité d'exploration expérimentale des trajectoires que nous proposons, nous avons donc opté pour une intégration en tant que plug-in dans l'environnement de *FluoSim*, que nous avons nommé *FluoSim-sonif*.

3.4.1 Développement initial

La première phase d'implémentation de l'outil proposé a été effectuée par Johan David, stagiaire au **SCRIME** d'avril à juillet 2021.

Le travail a consisté à programmer le plug-in *FluoSim-sonif* en C++, de sorte à ajouter un nouvel onglet « Sonification » au panneau de contrôle de *FluoSim*. L'interface graphique a été conçue avec l'API *Qt*⁵ et le moteur audio est géré par *RtAudio*⁶. La spatialisation de l'audio en particulier est simulée à l'aide de l'API *SteamAudio*⁷.

4. <https://github.com/mlagardere/FluoSim> – visité le 4/1/23

5. <https://www.qt.io/> – visité le 4/1/23

6. <https://www.music.mcgill.ca/~gary/rtaudio/> – visité le 4/1/23

7. <https://valvesoftware.github.io/steam-audio/> – visité le 4/1/23

Cette mise en place initiale a été décrite à mi-projet dans un poster pour le séminaire des doctorant.e.s de la Société Informatique de France [244], puis le rapport de stage de Johan David [47] a offert une description plus détaillée des fonctionnalités implémentées et de l'exploration des différentes options logicielles pour la programmation de sons spatialisés en C++.

La plupart des propositions décrites dans la section 3.3 ont été implémentées, mais quelques modifications ont été apportées par la suite, comme décrit dans la section suivante.

3.4.2 Retouches ultérieures

Depuis le prototype programmé et rapporté par Johan David, un certain nombre de corrections de bugs et de modifications de l'interface ont été effectuées par nos soins.

Réduction de la latence Il a été possible de réduire la latence de la synthèse sonore en retirant une réinitialisation redondante du contexte de spatialisation à chaque trame du rendu. La valeur minimale pouvant être attribuée à l'intervalle entre les trames est ainsi passée de 0.1s à 0.05s.

Mapping à la hauteur Par un souci initial de permettre la distinction individuelle de chacun des points, le mapping de la diffusivité à la hauteur avait d'abord été programmé pour que chacune des hauteurs disponibles soit attribuée une et une seule fois à un point pendant toute sa durée de vie à l'écran, de sorte que si une note était déjà associée une diffusivité donnée, la première note supérieure disponible était attribuée à sa place.

L'information de la diffusivité était cependant presque entièrement perdue avec ce choix de représentation puisque la hauteur attribuée à un son dépendait plus souvent de la première note disponible que de la proximité à un quelconque résultat de mapping. Cela ne laissait, par ailleurs, pas beaucoup de hauteurs disponibles pour des nombres de points potentiellement très grands. Ainsi on a malgré tout permis que chaque point puisse se voir attribuer la hauteur la plus proche de celle requise par le mapping, qu'elle ait déjà servi ou pas, comme initialement prévu et décrit dans la section 3.3.

Simplification de l'interface Afin de simplifier l'utilisation de l'interface pour des utilisateurs qui ne seraient pas nécessairement spécialisés dans le design sonore, nous avons désactivé, ou rangé dans des cadres dédiés au débogage, beaucoup des fonctionnalités plus bas niveau initialement présentes dans l'onglet de sonification. Un contrôle a été ajouté pour permettre à l'utilisateur d'indiquer les numéros de

3. Sonification immersive du suivi de protéines individuelles pour l'étude de la motilité cellulaire

notes minimum et maximum pour la sélection des hauteurs. La figure 3.4 montre les deux versions successives de l'interface côte à côte.

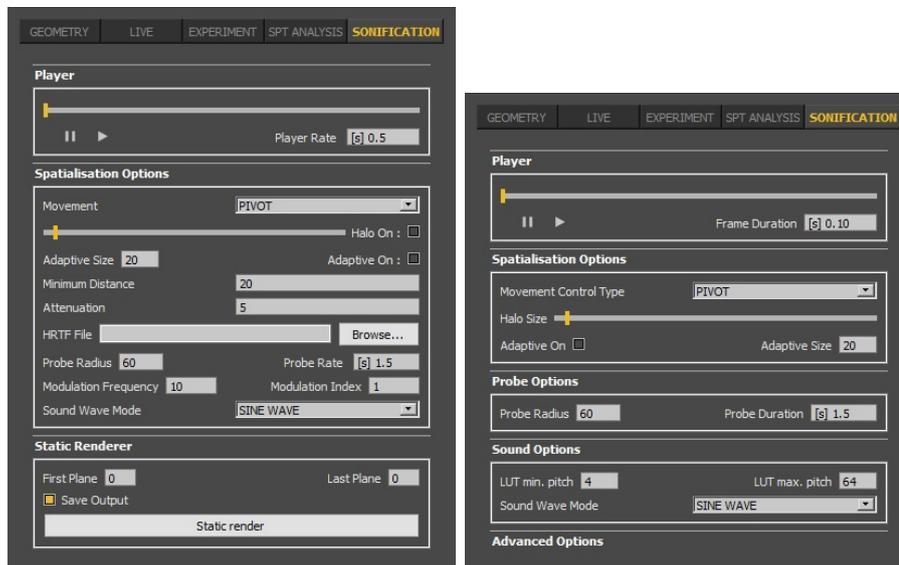


FIGURE 3.4 – À gauche, la première version de l'interface de *FluoSim-sonif*. À droite la version réorganisée. Les contrôles bas niveau de la synthèse sonore et les fonctionnalités de débogage sont dans le bloc « Advanced Options », inactif par défaut.

La gestion de l'atténuation du volume sonore en fonction de la distance à l'auditeur virtuel a aussi été simplifiée. Celle-ci dépendait en effet initialement de paramètres réglables par l'utilisateur (*minimum distance* et *attenuation*), mais pouvait simplement être calculée en fonction du rayon d'audibilité comme $1 - \text{distance} / \text{rayon_audibilite}$, avec une borne inférieure imposée à 0.

3.4.3 Usage

Pour une vue complète des paramètres offerts par *FluoSim-sonif*, voir dans l'annexe A le manuel utilisateur fourni avec la distribution du prototype. Pour une brève démonstration de l'usage de l'interface, voir la vidéo à l'adresse donnée en pied-de-page⁸.

Onglet SPT Analysis Les fonctionnalités d'importation et de filtrage des données sont les mêmes que pour la représentation visuelle déjà présente dans *FluoSim* et peuvent être utilisées depuis l'onglet **SPT Analysis** : sélection des

8. <https://youtu.be/-1mXjgaI3C0> – visité le 4/1/23

trajectoires à maintenir dans l’affichage à partir de seuils sur leurs taux de diffusivité ou leur fluorescence, contrôle sur le code couleur de l’affichage, sur la taille et l’intensité des points lumineux représentant les protéines, ainsi que sur la taille et l’intensité du trait pour les trajectoires.

Onglet Sonification Dans le bloc « Player », l’utilisateur a la possibilité de se situer à différents instants de la trace à l’aide du curseur d’avancement. La lecture de la trace peut être lancée à l’aide du bouton marqué par un triangle pointant vers la droite. La vitesse de lecture peut être réglée depuis le champ « Frame Duration » qui régit la longueur de chaque trame en secondes.

Le réglage du rayon d’audibilité est effectué dans le bloc « Spatialisation Options ». Il peut être défini soit dans l’absolu comme la taille en nanomètres du cercle, soit de façon adaptative pour un nombre donné de points à englober, la taille du cercle évoluant alors automatiquement pour toujours contenir le nombre de points demandés, si leur nombre à l’écran est suffisant.

L’interaction de l’utilisateur pour positionner l’auditeur virtuel peut se faire selon deux modes de déplacement : « Tank » ou « Pivot ». Dans le mode Tank les touches haut et bas permettent de faire avancer ou reculer l’auditeur de 3nm sur son axe d’orientation et les touches gauche et droite permettent de le faire pivoter de 0.15 radians vers sa gauche ou sa droite. Dans le mode Pivot, les touches directionnelles permettent de le déplacer librement sur le tableau pour une orientation fixe et l’orientation peut être réglée en cliquant, avec la touche Contrôle enfoncée, à la position que l’avatar doit viser.

Dans le bloc « Sound Options », il est possible de régler les bornes minimale et maximale entre lesquelles est effectué le mapping de la diffusivité instantanée à la hauteur.

Le mode d’exploration sonde peut être utilisé en mettant la vidéo en pause au moment souhaité puis en cliquant sur le point de départ voulu avec la touche Majuscule enfoncée. L’auditeur virtuel est alors instantanément positionné à la position sélectionnée et le déroulement du sonogramme est effectué autour de ce point de départ. Dans le bloc « Probe Options », deux réglages sont permis pour définir le rayon maximal que doit atteindre le front d’onde de la sonde, ainsi que le temps total que son déroulement doit durer.

Pour une spatialisation plus convaincante, l’outil se doit de permettre une adaptabilité au moins partielle aux divers profils de fonctions de transfert de la tête

(**HRTF**) chez différents utilisateurs. Pour cela, *FluoSim-sonif* offre à l'utilisateur le choix d'utiliser des **HRTF** personnalisées, enregistrés sous le format *.sofa*⁹, depuis le bloc « Advanced Options ». Deux configurations par défaut, tirées de la base de données KEMAR¹⁰ et représentant deux formes d'oreilles plus ou moins grandes, sont présentes dans le répertoire de l'outil.

3.5 Perspectives

Afin d'être en mesure de produire des évaluations représentatives de la qualité de la sonification plutôt que de celle du programme, la priorité dans un avenir proche sera de résoudre les plantages intempestifs et les problèmes de latence qui persistent dans l'état actuel du prototype. Ensuite, nous pourrons nous intéresser aux perspectives notées dans les sous-sections suivantes pour l'amélioration des timbres, l'ajout de fonctionnalités pour l'exploration, et les protocoles expérimentaux envisagés pour étudier l'efficacité de la représentation offerte.

3.5.1 De meilleurs timbres

Dans la première version produite, l'objectif était d'abord de tester l'intérêt du mapping de la diffusivité à des notes de musique, ainsi que l'apport d'une spatialisation interactive pour la tâche d'exploration. Ainsi, un défaut volontairement laissé de côté pour commencer a été l'utilisation de timbres très basiques (sinusoïdes, carrées, dents de scie, triangles), peu satisfaisants d'un point de vue esthétique et dont la faible bande passante spectrale – dans le cas de la sinusoïde en particulier – peut être la cause d'erreurs de différenciation entre l'avant et l'arrière dans le rendu spatialisé [148]. On prévoit ainsi d'ajouter la synthèse de timbres plus complexes et plus plaisants dans de futures versions de l'outil.

La sélection du timbre d'une trajectoire peut être fixée par l'utilisateur ou associée à son taux de diffusivité. Il serait intéressant à l'avenir d'expérimenter sur cette sélection pour déterminer quelle collection de timbres permet la meilleure détection des taux de diffusivité, que ce soit parmi les formes d'ondes basiques déjà implémentées, ou des sons d'instruments synthétisés. Pour créer des timbres plus complexes, on pourra utiliser des techniques de synthèse FM, employer différents types d'enveloppes pour l'amplitude (continue, percussive, montée), ou faire appel à des **DSP** écrits dans des langages dédiés (on pense par exemple à *FAUST*) importés dans le programme en C++.

9. https://www.sofaconventions.org/mediawiki/index.php/Main_Page
– visité le 4/1/23

10. <http://sofacooustics.org/data/database/mit/> – visité le 4/1/23

3.5.2 Fonctionnalités supplémentaires

Il a été proposé lors d'entretiens avec les chercheurs de l'IINS de donner une plus grande importance à la gestion de la géométrie, comme dans des itérations précédentes du projet (*DoReMolecule*, *GPSeyeSONIC*), pour permettre de sélectionner des régions d'intérêt de la capture et pouvoir par exemple se focaliser sur la sonification des trajectoires à proximité de la membrane.

Plutôt que de n'offrir qu'un contrôle manuel direct, il a aussi été suggéré de permettre à l'utilisateur de programmer des chemins à faire suivre automatiquement par l'auditeur virtuel. Par exemple, on pourrait ainsi se placer du point de vue d'une particule choisie et suivre sa trajectoire, ou parcourir la frontière entre deux cellules.

3.5.3 Protocole expérimental envisagé

Dans un premier temps, la distribution de notre outil de sonification est prévue pour au moins permettre à ses utilisateurs d'offrir des retours quant à leur expérience esthétique et ergonomique. Des protocoles expérimentaux plus formels sont aussi envisagés pour la suite afin de déterminer d'une façon plus quantitative si la nouvelle modalité d'exploration auditive ajoutée apporte effectivement un avantage sur celle, déjà présente, de l'observation visuelle.

Expérience utilisateur

Sans permettre une quantification de la facilité qu'aurait un utilisateur moyen à se situer dans un espace sonifié, une expérience utilisateur offrira déjà des données sur le ressenti vis-à-vis de l'ergonomie de l'outil, ainsi qu'un retour sur la qualité esthétique du son produit.

Le manuel utilisateur de l'annexe A a été distribué de façon informelle à l'issue de la première phase de programmation. Dans ce manuel, un lien est fourni vers une page web pour l'évaluation de l'expérience utilisateur (un questionnaire standard [UEQ-S \[193\]](#)), de sorte qu'après avoir découvert l'outil de sonification dans leur contexte de travail habituel, les chercheurs puissent donner dans le questionnaire des appréciations sur leur expérience. Une réelle campagne d'appel à évaluations au sein du personnel de l'IINS, basée sur ce support, aura lieu prochainement.

Test de localisation

À l'aide du protocole décrit dans cette sous-section, on cherche à déterminer la capacité des utilisateurs à localiser des zones de l'espace ayant des propriétés spécifiques en s'aidant de la représentation auditive interactive.

3. Sonification immersive du suivi de protéines individuelles pour l'étude de la motilité cellulaire

Pour un timbre fixé de spectre relativement riche (par exemple une forme d'onde triangulaire) et pour une large fourchette de hauteurs, les conditions variables pour l'expérience seront les différentes combinaisons de modalités possibles : visuelle uniquement, auditive uniquement, et les deux à la fois.

Plusieurs simulations d'activité cellulaire seront préalablement conçues pour présenter différentes configurations possibles de diffusivité, de fluorescence ou de densité des trajectoires. Chacune contiendra une zone minoritaire dans laquelle une ou plusieurs propriétés des trajectoires seront différentes du reste de la cellule : diffusivité du milieu inférieure ou supérieure au reste, fluorescence inférieure ou supérieure au reste, densité de trajectoires inférieure ou supérieure au reste.

Lors de l'expérience, il sera demandé à l'utilisateur de déplacer l'auditeur virtuel au cours de la lecture jusqu'à ces zones minoritaires. Une fois que l'utilisateur pensera avoir atteint la zone recherchée, il effectuera une confirmation de sa réponse, soit en cliquant sur un bouton donné, soit en appuyant sur une touche du clavier à cet effet.

On prédit que la condition visuelle permettra naturellement les meilleures performances, les utilisateurs spécialisés étant mieux habitués à cette modalité, mais elle pourra servir de condition témoin à laquelle sera comparée l'efficacité du retour auditif pour la même tâche : peut-on toujours accomplir la tâche aussi efficacement en s'aidant de l'ouïe qu'en s'aidant de la vue ? Quelle quantité de temps supplémentaire est nécessaire pour se repérer ?

La session expérimentale commencera par une phase d'entraînement à l'occasion de laquelle le testeur apprendra à se servir de l'interface, en particulier des nouvelles fonctionnalités de contrôle de l'auditeur virtuel auxquelles même les testeurs spécialisés n'auront pas été exposés auparavant.

Cet entraînement comprendra aussi une démonstration complète des retours visuels et auditifs qui peuvent être produits par l'outil d'exploration, et finira sur quelques tâches identiques à celles prévues pour l'évaluation mais qui ne seront pas comptabilisées dans les résultats, de sorte à confirmer que chaque testeur comprend l'exercice et parvient à l'effectuer.

Au vu du nombre assez réduit de testeurs spécialisés qui pourront être réunis, on envisage plutôt une évaluation de type « within group », où tous les testeurs seront exposés à toutes les conditions testées.

Test du mapping au timbre

Dans cette expérience, en partie similaire à celle décrite ci-dessus, on testera si une sélection automatique des timbres en fonction des types de mouvements peut aider un utilisateur dans des tâches de localisation.

Les conditions variables de l'expérience seront alors basées sur différentes configurations des timbres : un seul timbre fixé pour tous les points, une sélection expérimentale du timbre en fonction du taux de diffusivité global de la trajectoire, et une sélection témoin pour laquelle chaque point est représenté à chaque instant par un timbre sélectionné aléatoirement.

Comme précédemment, l'exercice consistera à localiser une zone minoritaire de l'espace dans laquelle les protéines ont une diffusivité, une fluorescence ou une densité différente.

Une version de l'interface dédiée à l'expérimentation

Des fonctionnalités de récolte des données expérimentales devront être ajoutées à *FluoSim* afin d'obtenir, pour chaque exercice, une trace de la position de l'avatar au cours du temps. On pourra par exemple enregistrer le numéro de trame, en cours de lecture, pour chaque moment où un déplacement ou une réorientation sera effectuée, ainsi qu'une identification du type de mouvement effectué.

Dans l'analyse ultérieure de ces résultats on s'intéressera, en fonction de la condition de représentation, à la précision des localisations par l'utilisateur. Cela consistera à vérifier si, au moment de l'exercice où l'utilisateur a confirmé sa réponse, il se trouvait effectivement dans la zone d'intérêt. Puis on étudiera, pour les exercices réussis, le temps qui aura été nécessaire aux testeurs pour arriver à la position correcte. Il sera même envisageable d'effectuer une analyse des trajectoires des utilisateurs au cours de l'exercice pour en déduire les habitudes d'exploration qui pourront se former lors des tâches de localisation par l'audition.

3.6 Conclusion

Ce cas d'étude porte sur l'observation auditive de nuages de points en mouvement pour l'étude des trajectoires des protéines lors de la motilité cellulaire. Les propriétés de ces points à représenter sont leurs positions dans l'espace, leurs taux de diffusivité instantanés et leurs taux de fluorescence au cours du temps, ainsi que leurs taux de diffusivité globaux.

La principale difficulté de ce cas d'étude est la très grande quantité d'informations à présenter à un utilisateur. Chaque point de données comporte

3. Sonification immersive du suivi de protéines individuelles pour l'étude de la motilité cellulaire

un assez grand nombre de dimensions et il peut y avoir un très grand nombre de points présents à un instant donné.

Comme solution, nous avons proposé une sonification de chacun des points selon un ensemble de mappings aux paramètres perceptifs habituels : spatialisation virtuelle pour les positions des points, hauteur (restreinte à des fréquences fondamentales de la gamme tempérée) pour le taux de diffusivité instantané, volume pour le taux de fluorescence et, en option, le mapping de la diffusivité globale au timbre. Afin de réduire la surcharge d'information, l'interaction utilisateur est prévue pour permettre une exploration sélective de l'espace des points en mouvement, ainsi qu'un certain degré de personnalisation pour les mappings à la hauteur et pour le timbre.

Un prototype de cet outil a été ajouté à l'interface de simulation et analyse de trajectoires *FluoSim*. Ce prototype présente un certain nombre de défauts encore à corriger, puis plusieurs niveaux d'évaluation de l'expérience utilisateur et de l'efficacité de la représentation sont prévus pour la suite.

Chapitre 4

Vers une modélisation de la réponse acoustique d'un environnement à des phénomènes physiques

Dans ce cas d'étude, en très grande partie théorique, nous proposons d'étudier les corrélations entre des mesures physiques de vitesse du vent au cours du temps et les sons ambiants induits par ce vent sur l'environnement (dans notre cas, une forêt de pins), afin de pouvoir par la suite générer le son que produirait le même environnement pour d'autres mesures de vitesse de vent. Dans la section 4.1 nous décrivons le contexte de recherche dans lequel s'inscrit cette étude, puis dans la section 4.2 nous formulons plus en détail le processus envisagé et l'état de l'art en termes de synthèse sonore conditionnée, en particulier à l'aide de réseaux de neurones artificiels. Ensuite, dans la section 4.3 nous décrivons une première acquisition de données à des fins d'expérimentation ainsi que l'étude de faisabilité effectuée par des étudiants en intelligence artificielle. Enfin nous entamons dans la section 4.4 une réflexion sur l'apparition de nouvelles techniques de sonifications liées à l'apprentissage machine de paramètres sonores pour la représentation des données.

4.1 Contexte

4.1.1 Affiliations

Ce cas d'étude découle de réunions de recherche avec Lisa Wingate (EcoFun) et Sylvain Dupont (ME) de l'UMR Interaction Sol Plante Atmosphère (ISPA), à l'Institut National de Recherche pour l'Agriculture, l'alimentation et l'Environnement (INRAE) de Bordeaux.

4.1.2 Domaine de recherche

Missions de l'ISPA

Les recherches de l'ISPA portent sur l'étude des cycles de matières, des transferts d'énergie et de la physique de l'environnement au sein d'écosystèmes complexes agricoles et forestiers. Le projet de cette unité de recherche est centré sur l'observation, l'expérimentation et la modélisation multi-échelles du fonctionnement des écosystèmes. Il vise à évaluer et proposer des actions pour garantir la durabilité de la production des agro-écosystèmes et des écosystèmes forestiers en quantité et qualité, en préservant l'environnement dans un contexte de changement du climat et des pratiques.

Équipe EcoFun (Relations hydriques et fonctionnement des écosystèmes)

L'équipe EcoFun étudie les réponses physiologiques et adaptations des écosystèmes terrestres aux changements environnementaux, à différentes pratiques de gestion, et aux événements climatiques extrêmes. Cela inclut des observations des mécanismes de photosynthèse et respirations, de l'usage par les arbres de leurs réserves de carbone pour leurs croissances primaire et secondaire, de la réaction des plantes à des niveaux de CO₂ ou des températures élevés, et de la sensibilité des forêts à la sécheresse.

Équipe ME (Mécanique Environnementale)

L'équipe ME étudie l'écoulement atmosphérique, la mécanique du sol et de la plante, et les échanges sol-plante-atmosphère sur une large gamme d'échelles spatiales, de quelques mètres à plusieurs kilomètres.

Utilisation de mesures d'anémomètres

Les mesures d'anémomètres sont utilisées au sein de l'équipe ME pour l'étude de l'interaction entre le vent et les arbres afin d'aider à modéliser et prédire les dommages pouvant être encourus au cours de tempêtes [62], ainsi que pour l'étude des transferts de matière et d'énergie induits par le vent sur le couvert forestier [61].

4.2 Analyse des corrélations entre mesures d'anémomètres et sons de l'environnement

4.2.1 Proposition

La technique de sonification que nous proposons dans le cadre de ce cas d'étude consiste à analyser les relations sous-jacentes entre les mesures d'anémomètre et le son ambiant du vent dans l'environnement de l'acquisition. Cette corrélation permettrait ensuite de synthétiser un son environnemental inédit pour de nouvelles mesures. Afin de traduire efficacement la structure spatiale de l'environnement (emplacements et dimensions des arbres, épaisseur du feuillage et du sous-bois), on cherche à effectuer ces opérations pour un son spatial, au moins binaural.

L'axe d'étude privilégié au cours de la thèse a plutôt été la conception d'un réseau de neurones qui apprendrait à reproduire l'audio à partir des mesures. Ces réflexions constituent la plus grande partie de ce chapitre, mais d'autres discussions ont aussi eu lieu pour rechercher des techniques d'analyse plus transparentes de la corrélation entre le son et les mesures à partir d'analyses spectrales. Celles-ci sont résumées dans la sous-section 4.3.2.

4.2.2 Motivations

Étude a posteriori des mesures d'anémomètre L'une des motivations les plus fréquentes pour une sonification est de permettre l'étude de données selon une modalité auditive. En particulier, un angle d'approche récurrent pour la sonification est l'utilisation de métaphores sonores rappelant le comportement réel du monde physique [81]. Dans le cas présent, l'un des meilleurs indicateurs auditifs associés à des phénomènes de vent est le son induit par ce vent dans son environnement : des bruissements de feuilles ou d'herbe plus ou moins intenses, provenant de différentes directions autour de l'auditeur, ou des sifflements lors de passages du vent dans des zones étroites à des vitesses suffisamment élevées. C'est pourquoi nous proposons comme technique de sonification la génération des sons induits par le vent sur un environnement forestier simulé, afin de permettre la détection de divers comportements du vent.

Bruitage Une telle sonification pourrait être proposée comme une nouvelle façon expérimentale de produire des sons de vent paramétrés pour servir, par exemple, dans la bande-son d'un jeu vidéo ou d'un film. Dans leur article *Visually Informed Sounds* [163], Owens et al. proposent une démarche de reconstitution de sons d'impacts à partir de vidéos, qu'ils justifient à la fois pour des applications dans

le domaine du bruitage et pour l'analyse du rapport entre les conditions physiques et la réponse acoustique, une autre motivation détaillée ci-dessous.

Modélisation de l'environnement Cette motivation n'est pas directement liée à la sonification elle-même, mais est corollaire de l'idée d'un apprentissage des corrélations mesures-sons : un moyen d'effectuer la modélisation de la réaction acoustique de l'environnement pourrait consister à passer par une étude poussée des dimensions, de la disposition, de la flexibilité et de la constitution des arbres présents. Elle pourrait ainsi contribuer aux axes de recherche pour la modélisation des environnements de l'unité **ISPA**, dont des membres collaborent avec nous sur ce cas d'étude, en offrant potentiellement un moyen de n'utiliser que des enregistrements audio (avec information spatiale de préférence) pour reconstituer les propriétés de l'environnement.

4.2.3 Difficultés

Naïvement, on pourrait espérer que le rapport que l'on cherche entre une mesure et une fenêtre de son serait instantané. Alors, pour un point de mesure, la courte fenêtre de son directement après ce point pourrait systématiquement être calculée en fonction de sa valeur.

Cependant, il paraît plus réaliste de supposer que les relations entre les fenêtres de mesures et de sons se font sur un beaucoup plus long terme, et qu'un phénomène du vent pourrait être lié à un phénomène dans le son autant a posteriori qu'a priori.

En effet, le vent et le son se déplacent à des vitesses très différentes. La vitesse du son dans l'air est d'environ 344m/s et celle du vent, plus variable, n'a en tout cas jamais dépassé le mur du son sur Terre. En dehors des tornades, le record mondial de vitesse du vent est à 408 km/h [53], soit 113.33 m/s, ce qui est encore loin du mur du son.

Cette différence de vitesse pourra être à l'origine de situations où une bourrasque causerait un bruit à quelques mètres des micros, qui le captureraient quasi-instantanément, puis seulement quand la même bourrasque atteindrait l'anémomètre – sous réserve qu'elle ne le manque pas, ce qui est un autre problème – elle serait enregistrée à son tour.

Pour une modélisation de la relation entre les mesures et le son, il faudrait donc faire en sorte de tenir compte d'un certain nombre de fenêtres temporelles dans le passé et dans le futur. Ainsi, au cours d'une prédiction du son en fonction de mesures lues dans leur ordre chronologique, et sous certaines conditions de direction et d'intensité du vent, l'algorithme devrait être en mesure de revenir en arrière a

posteriori dans le son déjà généré et d'y ajouter de nouvelles informations sonores liées à l'évènement détecté dans les mesures.

4.2.4 Existant

Notre étude de l'existant pour ce cas d'étude porte sur les systèmes capables d'apprendre les relations sous-jacentes entre des séquences de données et des sons afin de pouvoir ensuite effectuer la traduction de nouvelles séquences de données en sons. Trois applications d'un tel apprentissage qui nous ont paru particulièrement intéressantes à explorer sont celles pour la synthèse vocale à partir de texte (*Text-to-Speech*, ou TTS en anglais), pour la synthèse de sons musicaux à partir de partitions, et pour le bruitage artificiel de vidéos.

Synthèse de parole à partir de texte

La recherche sur les applications de l'intelligence artificielle au TTS a été particulièrement fournie au cours de la dernière décennie avec l'essor de l'apprentissage neuronal [8, 174, 224, 197, 173, 182, 134, 6, 39, 181, 206]. Le procédé consiste le plus souvent en une première traduction des données symboliques sous un format mel-spectrogramme, puis son décodage en tant que signal audio à l'aide d'un vocoder (souvent *WaveNet* [162], une forme dérivée [161], ou plus récemment un GAN pour la synthèse sonore comme *HiFiGAN* [127] ou *WaveGAN* [234]). On note que plus récemment les auteurs ont commencé à utiliser des mécanismes d'attention basés sur le modèle *Transformer* initialement proposé pour la traduction par Vaswani et al. [211]. Ces méthodes avec attention semblent, depuis, constituer l'état de l'art en termes de qualité pour le TTS [174, 173, 182, 134, 6, 39, 181, 206].

Comme indiqué dans la section 4.2.3, il n'est pas garanti dans notre cas d'étude qu'un évènement sonore correspondant à un phénomène du vent se produise exactement au même moment que la mesure d'anémomètre, ni seulement qu'il se produise toujours à la suite de cette mesure. Or, la traduction constitue un domaine d'application existant, en IA, pour lequel l'ordre de la séquence de données en sortie ne correspond souvent pas à l'ordre de la séquence de données en entrée. En effet, toutes les langues n'ayant pas les mêmes règles syntaxiques, il est parfois nécessaire de réordonner les mots résultant de la traduction lexicale initiale pour que la phrase produite en sortie ait du sens [236]. Cet apport particulier des modèles à attention pour la traduction, ainsi que sa compatibilité avec la synthèse vocale, semblent ainsi constituer une piste prometteuse qui pourra être explorée pour notre synthèse de sons naturels.

Synthèse de musique à partir de MIDI

Il est encourageant de constater que les méthodes de synthèse audio qui ont déjà fait leurs preuves pour le TTS ont aussi commencé à être adaptées pour d'autres tâches proches comme, ci-dessous, la synthèse musicale.

Hawthorne et al. [94] ont proposé en 2018 un modèle capable de traduire un enregistrement audio en MIDI, de générer une continuation du morceau sous un format MIDI et – ce qui nous intéresse ici – de traduire la partition MIDI sous forme d'audio. Leur méthode, proche de la synthèse vocale, ne consiste pourtant pas à produire un mel-spectrogramme pour conditionner *WaveNet*, mais à lui passer directement une liste des notes actives et inactives (*piano roll* en anglais). Cette particularité est probablement due au fait qu'il n'est pas nécessaire de paramétrer le timbre de l'instrument dans leur outil, puisqu'il s'agit toujours d'un piano.

Kim et al. [125] proposent en 2019 d'adapter les procédés de synthèse vocale pour du texte à la synthèse musicale pour du MIDI. La séquence de symboles MIDI est d'abord traduite en un *piano roll*, puis *ce piano roll* est lui-même traduit en mel-spectrogramme. *WaveNet* est ensuite conditionné par ce spectrogramme afin d'obtenir l'audio correspondant. Dans ce modèle, le fait d'utiliser un mel-spectrogramme en entrée de *WaveNet* permet de paramétrer le timbre d'instrument qui joue la musique, afin de proposer un ensemble d'instruments pour lesquels le modèle est entraîné, en plus d'interpolations de ces instruments dans un espace du timbre à deux dimensions.

Synthèse de sons à partir de vidéos

Une dernière application de la synthèse sonore par intelligence artificielle qui semble particulièrement courante dans la littérature est celle de la création de sons en fonction de séquences d'images. Certains travaux, en particulier, qui effectuent la reconstitution de sons naturels [163, 239] montrent qu'il est bien possible de reproduire des sons d'interactions physiques déduites de l'évolution de données séquentielles.

Visually Indicated Sounds d'Owens et al. [163] effectue un apprentissage et une restitution des sons produits par divers matériaux selon des interactions de type impact ou grattement, à partir de séquences vidéo de ces interactions se produisant.

En 2018, Zhou et al. [239] proposent un outil pour la génération, à partir de vidéos, des sons naturels pouvant provenir de différentes sources sonores reconnues dans l'image par un réseau de neurones. La même année, Zhao et al.

4. Vers une modélisation de la réponse acoustique d'un environnement à des phénomènes physiques

décrivent un principe similaire pour une application aux vidéos de performance musicale [238, 237].

Milazzo et Buehler ont exploré dans un travail de 2021 l'apport potentiel des flammes dans la création artistique par IA [151]. L'une des applications présentées consistait en particulier à corréliser par apprentissage profond les oscillations et orientations d'une flamme de bougie avec les différentes tonalités auxquelles elle était exposée.

4.3 Expérimentation préliminaire

4.3.1 Capture de données synchrones

Afin de récolter des données sur lesquelles expérimenter pour notre étude de la corrélation entre les mesures et les sons, nous avons organisé début 2021 des sessions d'acquisition au cours desquelles des microphones, prêtés par le **SCRIME**, étaient placés à proximité d'un anémomètre sur un site de capture et enregistraient le son ambiant en synchronisme avec celui-ci.

Références du matériel

Anémomètre sonique Campbell Scientific CSAT3-3B¹.

Microphones Deux micros monophoniques NT5 de RØDE² et un micro ambisonique (4 pistes) SPS200 de SoundField³.

Comme le but de cette acquisition était de capturer en priorité l'interaction du vent sur l'environnement, les micros ont été équipés de filtres anti-pop afin d'éviter autant que possible de capturer le bruit du vent soufflant sur les appareils.

Carte son Une carte son Fast Track Ultra 8R de M-Audio⁴.

1. <https://www.campbellsci.fr/csats3> – visité le 4/1/23
2. <http://www.rote.com/microphones/nt5> – visité le 4/1/23
3. <https://www.soundfield.com/#/products/sps200> – visité le 4/1/23
4. <https://www.sonology.shop/article/m-audio-fast-track-ultra-8r.html>
– visité le 4/1/23

Site de mesures

L'acquisition a été effectuée sur la station de mesures de l'infrastructure européenne *Integrated Carbon Observation System* (ICOS) situé dans la forêt de pins bordant le Chemin de Sillac, à proximité de Salles, 33770⁵.

Dispositif mis en place

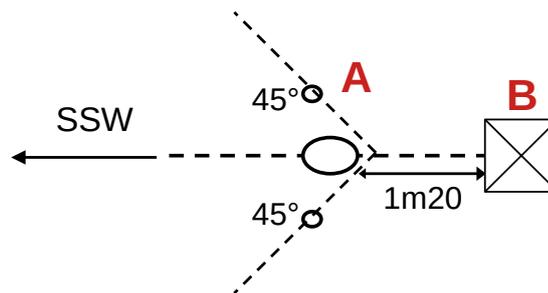


FIGURE 4.1 – Un schéma du dispositif de capture

Le schéma de la figure 4.1 décrit le positionnement, vu d'en haut, du dispositif pour l'acquisition synchronisée. L'anémomètre est sur un mât (*B* sur la figure) à environ 10m de hauteur et orienté vers le sud-sud-ouest. Les microphones (*A* sur la figure) sont placés à environ 1.20m du mât, à une hauteur d'environ 1.20m. Le micro ambisonique central a été orienté vers le sud-sud-ouest et les deux micros monophoniques ont été placés à des angles de 45° de part et d'autre. Tous les micros sont relevés de 45° par rapport au plan horizontal.

Données acquises

Audio Le son acquis est échantillonné à un taux de 48 kHz et couvre un total de 6 pistes. Les deux premières pistes sont enregistrées par les micros monophoniques de part et d'autre de l'anémomètre, tel que le micro à sa gauche donne la première piste et celui à sa droite donne la deuxième. Les quatre autres sont enregistrées par le micro ambisonique.

Vent Les mesures de la vitesse du vent sont échantillonnées à un taux de 50 Hz sur trois canaux (deux composantes horizontales u, v et une composante verticale w). Les valeurs sont données en $m.s^{-1}$ selon le repère de l'anémomètre.

5. <https://xylofront.pierroton.inra.fr/Salles2.html> – visité le 4/1/23

4. Vers une modélisation de la réponse acoustique d'un environnement à des phénomènes physiques

L'interaction entre les taux d'échantillonnages est telle que pour un point de mesure de l'anémomètre, il correspond une fenêtre de $48000/50 = 960$ échantillons sonores (multiplié par 6, le nombre de pistes audio enregistrées) formant les 0.02 secondes qui suivent l'instant de la mesure. La figure 4.2 illustre cette correspondance.

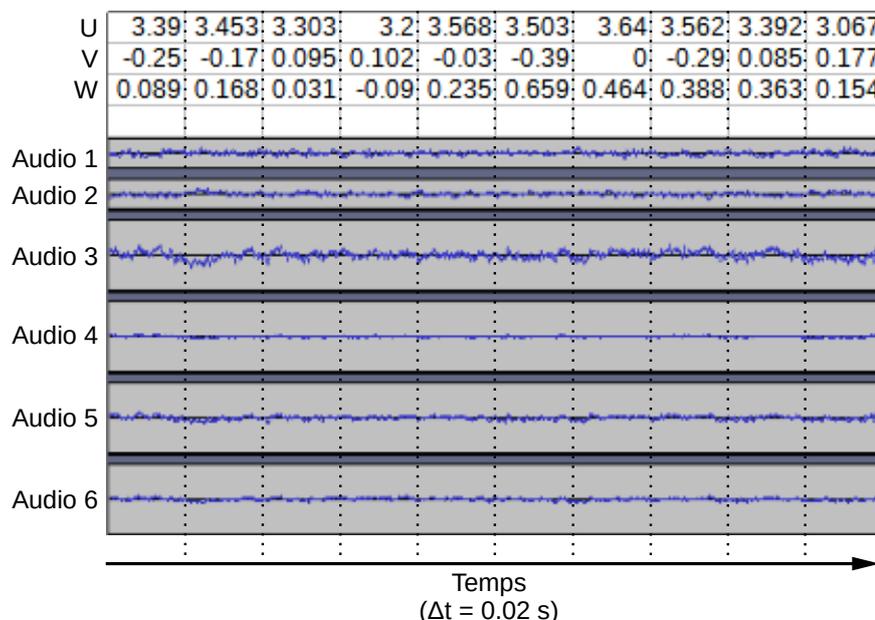


FIGURE 4.2 – Illustration de la correspondance dans le temps entre les mesures d'anémomètre à 50Hz et les fenêtres du son enregistré à 48 kHz

Sessions d'acquisition Deux sessions d'acquisition ont eu lieu :

- Le 1er février 2021 : 30 minutes entre 12h30 et 13h.
- Le 19 février 2021 : 2 heures et 45 minutes entre 10h18 et 12h53.

Les jours pour ces sessions ont été choisis, autant que possible, pour leur météo venteuse avec peu de précipitations.

Défauts des données Sur le lieu de l'acquisition, un appareil de mesure présent à proximité du mât de capteurs produisait un sifflement continu de fréquence légèrement modulée qui a été capturé par le dispositif d'acquisition audio. La fréquence de ce sifflement se situait à 4250 Hz le 1er février et à 3000 Hz le 19 février. Par ailleurs, l'amplitude des données audio est globalement très faible, à peine perceptible à l'écoute.

Notre but sera donc de résoudre ces problèmes, d'après l'observation de ces résultats, à l'occasion de prochaines séances d'acquisition. Le problème de volume sonore est en effet dû au fait que les réglages du matériel étaient essentiellement restés les mêmes du début à la fin de l'emprunt au **SCRIME**, sans certitude qu'ils étaient appropriés pour les conditions de l'acquisition.

Afin de s'assurer que l'ensemble des mesures est représentatif de l'étendue complète des vitesses de vent pouvant se produire dans l'environnement d'étude, il est aussi prévu d'effectuer une ou plusieurs sessions sur des journées particulièrement venteuses, de préférence à l'occasion d'une tempête.

On tâchera enfin de choisir et noter plus soigneusement les positions et orientations des micros par rapport à l'anémomètre et par rapport au plan horizontal. En particulier, il n'a en effet pas été possible de reconstituer a posteriori l'orientation des différentes pistes enregistrées par le micro ambisonique lors des sessions de février 2021.

4.3.2 Observation des spectres de nombres d'ondes

Un protocole d'analyse des données acquises a été proposé par Sylvain Dupont afin d'observer de plus près les rapports potentiels entre les mesures de vitesse du vent et les enregistrements audio.

On trace les spectres de Fourier normalisés des signaux audio pour chaque piste, ainsi que pour la composante verticale w et la composante longitudinale u des mesures d'anémomètres. Afin de situer les spectres dans des échelles comparables, on divise l'axe horizontal de chaque grandeur par sa vitesse moyenne de déplacement : $v_{sound} = 344m.s^{-1}$ pour les pistes de son et $u_{wind} = 3.2m.s^{-1}$ pour le vent.

On passe ainsi d'un spectre des fréquences en Hz (ou s^{-1}) à un spectre des inverses de longueur d'onde (ou nombres d'ondes) en m^{-1} . On peut ainsi observer les pics d'intensité communs dans ces spectres dont on peut déduire des relations mécaniques potentielles entre les fluctuations du vent et les fluctuations du son.

Les deux spectres de la figure 4.3 illustrent ces distributions pour la vitesse longitudinale u et verticale w du vent, et pour les 6 canaux de son.

4. Vers une modélisation de la réponse acoustique d'un environnement à des phénomènes physiques

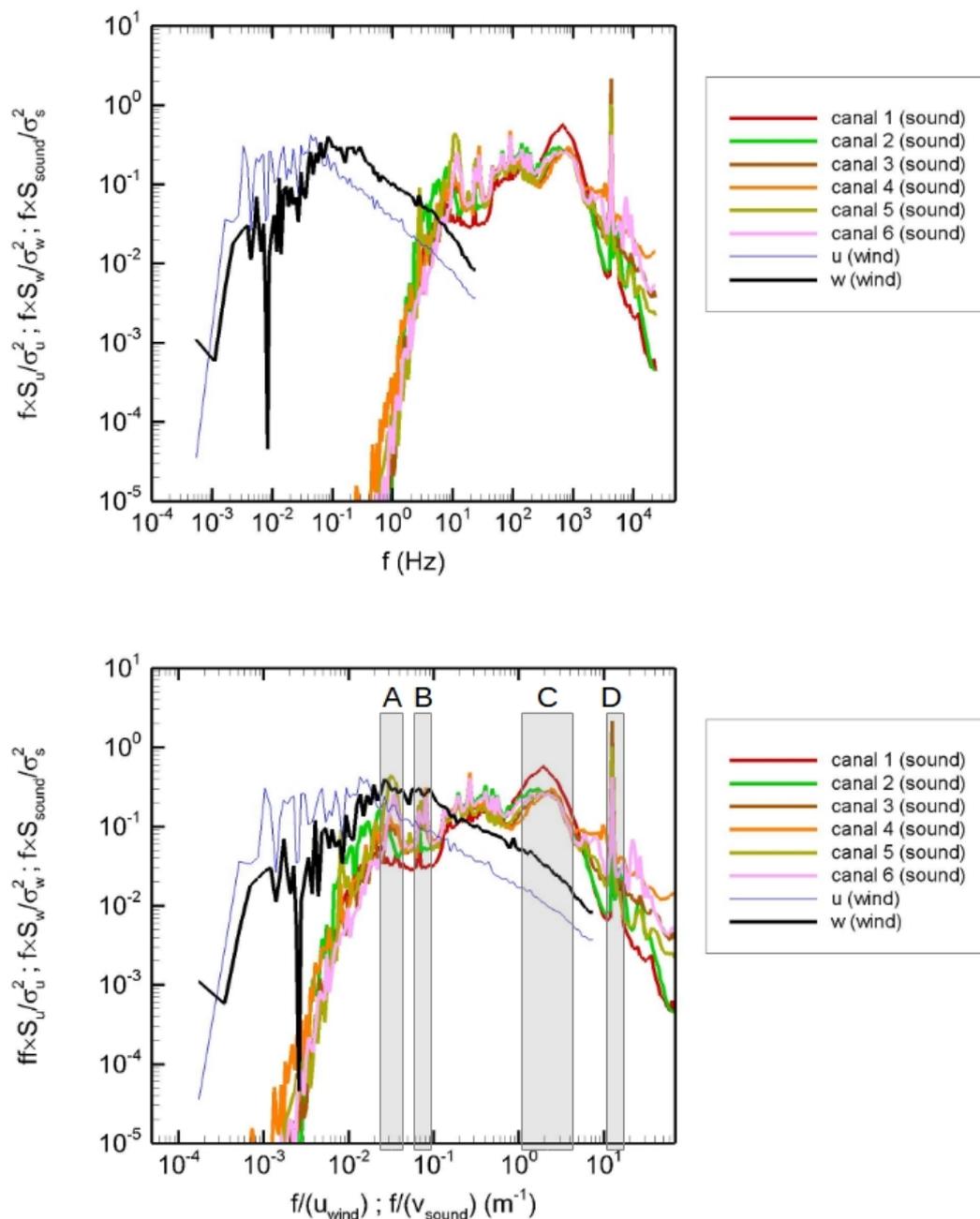


FIGURE 4.3 – En haut : spectre des fréquences des vitesses u et w du vent, et des 6 pistes d'audio. En bas : spectre des nombres d'ondes avec zones d'intérêt en gris

Le spectre du bas de la figure 4.3 comporte une série de zones d'intérêt que l'on tente d'interpréter ci-après :

- A : autour de 0.03 m^{-1} , il semble pouvoir y avoir un lien entre le pic de w et le premier pic basse fréquence des signaux de son. Ce pic est autour de 3 à 4 fois la hauteur des arbres, ce qui paraît cohérent avec la taille des rafales de vent au sommet des arbres.
- B : un deuxième pic autour de 0.08 m^{-1} , soit 12.5 m, apparaît sur les signaux de son. Étant proche de la hauteur des arbres (9-10 m), ce pic pourrait correspondre à la première fréquence propre d'oscillation des arbres.
- C : la « bosse » autour de 2 m^{-1} (soit une échelle de longueur de 0.5 m) pourrait alors correspondre aux oscillations des branches.
- D : le pic nettement marqué vers 10 m^{-1} est celui du sifflement parasite capturé dans le son.

Cette observation préliminaire suggère qu'il pourrait bien exister un lien entre les mesures de vent et les sons enregistrés aux basses fréquences, en particulier induit par la structure des arbres en présence (troncs, branches, aiguilles). Cette hypothèse sera à confirmer une fois que des données de meilleure qualité auront été acquises.

4.3.3 Étude de la faisabilité

Dans le cadre d'un projet semestriel d'intelligence artificielle, un travail d'étude de la faisabilité a été soumis au binôme d'étudiants Maëlle Andricque et Martin Devreese, en 3^{ème} année respectivement à l'ENSEIRB-MATMECA et à l'ENSC. Les données acquises selon la mise en place décrite dans la section 4.3.1 leur ont été fournies et expliquées, et leur but était de déterminer si une synthèse sonore paramétrée par des mesures d'anémomètre pouvait être réalisée selon un apprentissage machine.

Leur démarche, résumée ci-dessous et visible dans le dépôt Google Drive du projet⁶, consiste à nettoyer les données audio, réduire les données d'anémomètre et tester la possibilité de différencier à l'écoute des moments notables de la capture audio.

6. <https://drive.google.com/drive/folders/1IGPH2T2ZJkN7QqMwVJeIPyeFqDieerso?usp=sharing> – visité le 4/1/23

Nettoyage de l'audio

Le son enregistré comportant un sifflement parasite constant, le premier pré-traitement des données audio a consisté à leur appliquer un filtre coupe-bande entre les fréquences 2950 et 3100 Hz. Ensuite, afin de rendre les fluctuations de volume plus détectables, l'intégralité de l'audio disponible a été amplifiée de 30 dB.

La figure 4.4 montre un comparatif des mel-spectrogrammes de la première minute d'audio avant et après ce traitement.

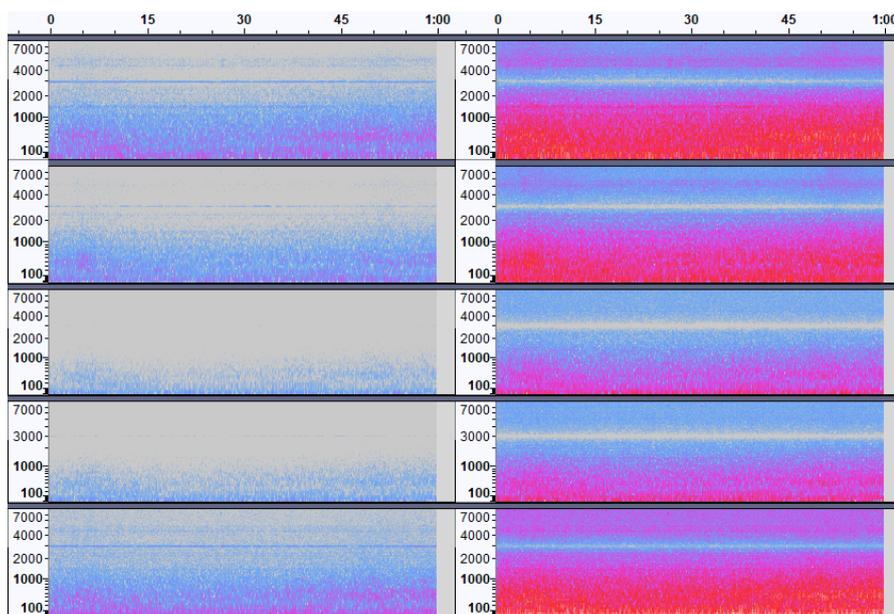


FIGURE 4.4 – Capture sous Audacity des mel-spectrogrammes de la première minute d'audio du 19 février 2021, avant (gauche) et après (droite) le pré-traitement

Réduction des données d'anémomètre

Afin d'attribuer aux diverses combinaisons de coordonnées un nombre fini de classes à partir desquelles un apprentissage pourrait s'effectuer plus facilement, les mesures d'anémomètres ont été associées à des coordonnées spatiales sphériques, donnant ainsi un ensemble de classes qui caractérisent approximativement la direction du vecteur de vitesse du vent. L'intensité de ce vecteur est utilisée comme un attribut supplémentaire à l'identifiant de classe pour former des couples de valeurs caractéristiques de chaque mesure d'anémomètre : (classe, intensité).

Différenciation de l'audio

Les données audio ont été recoupées en fenêtres de 2 millions d'échantillons (environ 41.67 secondes), auxquelles un algorithme d'apprentissage *K-means* a été appliqué. Le nombre de groupes K a été fixé à une valeur relativement faible ($K = 3$) afin de détecter des centroïdes qu'on espère particulièrement différents d'un point de vue perceptif. Cependant, l'écoute de ces centroïdes ne permettant pas de différenciation perceptive, il a été conclu que l'audio produit par un apprentissage des corrélations avec les mesures d'anémomètre ne permettrait pas de détection de leurs variations.

Discussion

Si le nettoyage audio décrit semble indispensable et la méthode de réduction des données peut être intéressante à tester pour un apprentissage, la méthode de différenciation des phénomènes dans le son paraît discutable. En effet, le regroupement des fenêtres d'audio par *K-means* a été effectué sur l'ensemble brut des échantillons sonores, comme si chacun constituait une dimension d'un point dans un espace à 2 millions de dimensions. Une comparaison par extraction de caractéristiques spectrales ou perceptives, comme par exemple les **MFCC**, aurait permis un regroupement plus pertinent, puisque purement sensible à la composition perceptive des sons plutôt qu'à leur organisation temporelle.

4.4 Vers de nouvelles techniques de sonification par apprentissage

Une technique de sonification rare

Au cours de notre étude des possibilités pour ce cas d'étude, la recherche d'existant dans le domaine de l'apprentissage machine (section 4.2.4) a fait ressortir un certain nombre de travaux dont l'objectif de base était assez similaire au nôtre : corrélérer des sons non verbaux avec des données séquentielles supposées synchronisées [163, 238, 239, 237, 151]. Si ces exemples s'occupent plus spécifiquement de créer des bandes sons pour des vidéos, rien ne semble empêcher de généraliser ce concept à toutes sortes d'autres contextes impliquant des données séquentielles et des sons.

Le seul de ces travaux à employer le terme « sonification » est celui de Milazzo et Buehler [151]. Pourtant, ce type d'application correspond bien à la définition d'une sonification. Il s'agit bien, en effet, de traduire un ensemble de données en

4. Vers une modélisation de la réponse acoustique d'un environnement à des phénomènes physiques

un signal sonore non verbal à des fins d'intelligibilité [128] et d'une façon objective, systématique et reproductible [97].

Cependant, il est assez difficile de rapprocher la démarche de ce genre de travail de l'une ou l'autre technique de sonification telle qu'elles sont traditionnellement décrites dans la taxonomie (voir le chapitre 1). Peut-on encore considérer qu'il s'agit d'une sonification par parameter mapping si les opérations donnant les paramètres sont définies implicitement par un algorithme d'intelligence artificielle plutôt que par un concepteur humain ?

Il semble ainsi qu'une nouvelle famille de techniques de sonification ait commencé à faire son apparition, rendue possible par les progrès récents (et toujours en cours) de l'apprentissage machine, en particulier de la synthèse audio par réseaux de neurones, et que cette famille ne soit pas actuellement comptabilisée dans la taxonomie des techniques de sonification. Pourrait-on s'attendre à ce qu'elle y fasse son entrée à l'avenir sous le nom, par exemple, de *learning-based sonification* ?

Applications envisagées

Métaphores sonores apprises Étant donné un modèle d'apprentissage entraîné à faire correspondre un certain type de son à un certain type de données, il serait possible d'utiliser ce modèle comme une « métaphore » pour représenter des données d'un autre type. Pour l'exemple de notre cas d'étude pour la modélisation du son de l'environnement en fonction de l'évolution du vecteur 3D de vitesse du vent, on pourrait vouloir représenter avec le même type de son d'autres données à trois dimensions évoluant au cours du temps. Par exemple, la diffusion d'une particule au cours du temps (voir chapitre 3) pourrait ainsi être représentée par un son de vent spatialisé. Il suffirait pour cela de transposer ses coordonnées de vitesse instantanée dans les échelles de grandeurs tolérées par le modèle, puis de les lui passer en paramètres.

Mappings personnalisés Une autre application envisageable de l'apprentissage machine en sonification serait pour la création interactive de sonifications personnalisées par les utilisateurs, sans nécessiter de leur part de connaissances en synthèse sonore. De la même façon que, pour un grand nombre de services en ligne, des algorithmes d'apprentissage analysent les préférences et les habitudes d'utilisateurs afin de leur proposer des contenus personnalisés, des techniques d'apprentissage des préférences pourraient être mises en place pour la sonification. Une sonification ainsi personnalisée pourrait offrir une plus grande implication personnelle des utilisateurs, une meilleure compréhension des mappings, et une meilleure durée de vie de l'outil (si le mapping devient lassant, on peut demander à le changer).

Ce type d'application de l'apprentissage pour la sonification a déjà été partiellement exploré par Hermann, Bunte et Ritter en 2007 [101]. Leur méthode consistait à présenter, itération par itération, des évolutions de la dernière sonification préférée, auxquelles l'utilisateur attribuait des notes afin d'affiner ses préférences de paramètres pour une tâche d'exploration des données d'Iris de Fisher.

Han et al. ont ensuite proposé en 2014 une autre expérience de sonification apprise [90]. Dans leur cas, il s'agissait d'effectuer la sonification de caractéristiques visuelles calculées à partir de photographies de peau. Une première phase de préparation consistait à définir manuellement les paramètres de synthèse FM (l'enveloppe de modulation et l'enveloppe d'amplitude) pour un certain nombre d'images d'exemple. Ensuite, un apprentissage par *Support Vector Regression* (SVR) déterminait automatiquement comment attribuer des paramètres de synthèse FM à de nouvelles données de caractéristiques visuelles.

4.5 Conclusion

Dans ce chapitre, nous proposons l'idée qu'une étude des corrélations entre des mesures d'anémomètres et l'évolution d'un son au cours du temps pourrait ensuite permettre d'effectuer une synthèse sonore paramétrée par des mesures d'anémomètre, une technique qu'on considère comme une sorte de sonification par « parameter mapping appris » de ces données d'anémomètres.

Cette approche relativement rare de la sonification présente ici un défi particulier de non linéarité temporelle, liée aux interactions spatiales complexes auxquelles le vent est soumis dans un environnement naturel. Pourtant, certaines réussites récentes en traduction par IA, ainsi que leurs applications en synthèse vocale, semblent encourageantes quant à la faisabilité de l'opération proposée.

Nous avons effectué en février 2021 une première tentative d'acquisition de données synchrones pour cette étude de la corrélation mesures-sons. Cependant, la faible qualité du son enregistré n'a pas permis de produire des analyses viables. Certaines hypothèses quant au rôle mécanique de la structure des arbres dans l'acoustique de l'environnement ont pu être formulées, mais elles nécessiteront d'autres données plus soigneusement acquises pour être confirmées ou infirmées. De nouvelles sessions d'acquisition sont ainsi prévues pour un futur proche.

Troisième partie

Vers un modèle générique pour la sonification de données

Chapitre 5

Description exhaustive d'un outil de sonification

Le principal objectif de la présente thèse est de proposer une modélisation du principe de la sonification permettant d'encadrer la description de n'importe quelle instance de ce type de représentation.

La taxonomie usuelle des techniques de sonification, reprise dans la section 1.4, liste un ensemble de catégories de représentations définies par les choix effectués sur les types de sons employés (auditory icons ou earcons) ou sur la façon dont le son est construit par rapport à l'organisation des données (MBSon, PMSon ou audification).

Ces catégories offrent déjà un cadre pour la conception de sonifications selon les techniques connues et acceptées, mais afin de permettre une expérimentation plus souple pour la représentation auditive dans sa généralité, il est utile de pouvoir définir à un plus bas niveau les différentes caractéristiques d'une sonification.

Dans ce chapitre, nous proposons ainsi une liste des aspects variables de l'objet sonification, pouvant être vue comme une série de questions auxquelles un concepteur répondrait pour décrire ses choix de façon exhaustive :

- Quelles sont les données que la sonification doit permettre de représenter ?
- Quelles propriétés sous-jacentes de ces données doivent être mises en évidence par la représentation ?
- Quels sons doivent être employés afin de traduire ces données ?
- Quelles propriétés perceptives de ces sons doivent servir de variables pour la représentation ?
- Selon quelles opérations les propriétés des données doivent-elles être associées aux propriétés perceptives des sons ?
- Quel rapport temporel y a-t-il entre le déroulement du son et l'organisation des données ?

- Quelle interactivité est offerte à l'utilisateur de l'outil de sonification ?
- Quels paramètres de la sonification peuvent être réglés au cours de l'utilisation de l'outil ?

Dans les section qui suivent, nous précisons la teneur de chacune de ces questions, tout en illustrant la diversité des réponses pouvant y être apportées d'après différents exemples tirés de l'existant.

5.1 Données d'entrée

La sonification consistant invariablement à représenter des données par le biais de sons non verbaux, il semble indispensable avant tout de définir de quels types de données il s'agit, c'est-à-dire selon quelles structures celles-ci sont organisées, et quel rapport au temps elles ont.

Organisation

En informatique, il est inévitable que les données se présentent sous une forme numérique, mais pour les besoins d'un algorithme qui accéderait aux différentes propriétés des données pour les associer à des paramètres sonores, il est utile d'explicitier la façon dont ces valeurs forment des structures dont les différents membres correspondent aux éléments du système à représenter.

Dans des applications pour les courbes auditives ces structures sont simplement des tranches temporelles correspondant à des états successifs de l'ensemble des dimensions des données.

Puis, selon les domaines, les structures de données à représenter peuvent prendre des formes plus complexes. Par exemple, dans le cas d'étude du chapitre 3 sur les trajectoires de protéines, les données sont structurées comme des tranches temporelles contenant les points présents dans la capture à l'instant correspondant, eux-mêmes caractérisés par leurs propriétés de fluorescence, diffusivité et position dans l'espace (qui est elle-même un vecteur de deux coordonnées).

Les structures de données peuvent alors être constituées de sous-structures imbriquées les unes dans les autres. Ainsi par exemple, si l'on part du principe qu'une information de couleur est encodée sur trois canaux numériques rouge, vert et bleu, alors une structure d'image est un tableau constitué de lignes et de colonnes de pixels dont chacun est caractérisé par ces trois canaux de couleur, puis une structure de vidéo est un enchaînement temporel de ces tableaux de pixels.

Si les données sont forcément communiquées sous une forme numérique, elles ne représentent cependant pas toujours des nombres. Pour des besoins de représentation auditive de la structure d'un fichier **HTML** [171] ou d'expressions mathématiques [203], par exemple, les données d'entrée sont lues et structurées comme du texte. Dans le cas d'un outil comme *Sonic Browser*, conçu par Fernström et al. [68] pour l'exploration sonore d'une bibliothèque de musique, les données d'entrée de la sonification sont elles-mêmes des fichiers sonores.

Rapport au temps

La deuxième propriété des données pouvant varier selon les cas est celle de leur rapport au temps. En effet, les données d'entrée peuvent provenir de sources statiques, comme des fichiers préalablement enregistrés, ou de sources dynamiques comme des flux émis par des appareils de mesure ou d'autres programmes fonctionnant simultanément à l'outil de représentation. Puis ces données peuvent elles-mêmes correspondre à des systèmes dynamiques, tels que les états successifs du système se produisent selon un pas de temps donné, ou statiques tels que les données représentent un seul état du système fixé dans le temps.

5.2 Pré-traitement des données

Si, par défaut, une sonification se contente souvent de représenter les données telles qu'elles sont, il est parfois nécessaire d'en faciliter l'interprétation en extrayant de ces données les propriétés sous-jacentes que la représentation doit faire ressortir.

Ces propriétés peuvent être des valeurs statistiques assez simples extraites de l'ensemble des données, comme par exemple leurs valeurs minimales ou maximales, leur moyenne, etc.

Pour des applications de courbes auditives par exemple, il est ainsi souvent utile d'ajouter des repères sonores à la représentation, afin d'informer l'auditeur du contexte quant à la teneur générale des données [199]. On peut alors adjoindre à la représentation de chaque point successif une tonalité constante paramétrée par la valeur maximale, minimale, médiane ou moyenne de la dimension.

Les opérations de pré-traitements peuvent aussi prendre des formes plus complexes. Ainsi par exemple, dans un des cas d'application du sonogramme des données présenté dans [107], Thomas Hermann propose de paramétrer le modèle pour exprimer la diversité des classes présentes dans le voisinage de chaque point. Une propriété des points préalablement calculée est alors leur entropie locale de

classe, donnée par un comptage des représentants de chaque classe parmi les k plus proches voisins du point.

La représentation peut encore nécessiter d'extraire de données brutes certaines propriétés liées au domaine d'application. Par exemple, dans le cas d'étude du chapitre 3, le taux de diffusivité de chaque trajectoire est calculé à partir de la quantité de déplacement effectué entre chaque position successive du point.

Outre ces analyses préalables des données dans leur globalité, les opérations de pré-traitement peut aussi consister à calculer à chaque instant l'écart de valeur d'une dimension des données par rapport à un seuil fixé, de sorte que la valeur à représenter soit cet écart plutôt que la valeur d'entrée elle-même.

Par exemple, dans le cas de la sonification de critères de fabrication pour la supervision d'une fabrication additive (telle que décrite dans le chapitre 2), les valeurs associées aux paramètres sonores sont en fait leurs divergences par rapport aux valeurs cibles définies pour le bon déroulement de la fabrication.

On admet que de tels pré-traitements des données pourraient être effectués préalablement à l'aide d'outils externes, de sorte que les données d'entrée puissent être directement traitées par le programme de sonification. Cependant, il est toujours plus pratique d'utiliser un outil qui ne nécessite pas d'aller et venir entre différentes applications pour produire un résultat. Il est donc important de pouvoir définir des outils de sonification qui acceptent directement les sources de données telles qu'elles sont produites par les outils propres au domaine d'application, et s'occupent eux-mêmes d'en extraire les informations à représenter.

5.3 Son en sortie

La recherche en psychoacoustique pour la sonification semble suggérer qu'il n'existe pas une seule « bande-son miracle », dans laquelle un ensemble donné de sons paramétrés pourrait être représentatif de n'importe quel système de données [70, 184].

Au contraire, les modèles cognitifs actuels encouragent plutôt à sélectionner des assemblages de métaphores très spécifiques en fonction des données à représenter, de la tâche liée à l'observation [14], du domaine d'application, du contexte d'écoute, de la culture [184], ou des capacités musicales [157], cognitives [221] ou sensorielles [220] du public visé.

Il peut alors être préférable, selon les besoins, de plutôt employer des métaphores inspirées par des sons naturels [81] ou par l'écoute musicale, comme par exemple dans le cas des earcons [19]. Roddy et Bridges [184] indiquent ainsi que le choix

de métaphores se fait selon deux sélections : celle des « **complexes sonores** », c'est-à-dire des flux sonores représentatifs des différentes entités présentes dans les données, et celle des « dimensions » sonores représentant l'évolution des dimensions de ces entités. Ce concept de dimensions, ou paramètres, des complexes sonores est exploré plus précisément dans la section suivante.

Le signal sonore peut être construit selon différentes combinaisons de principes de synthèse additive, soustractive, par modulation (AM/FM), par modèles physiques ou à partir de sons pré-enregistrés. Une autre possibilité est que le rendu sonore soit géré par un ou plusieurs signaux **MIDI** communiqués à d'autres modules qui seraient chargés de les jouer. Le son (ou **MIDI**) produit peut encore être destiné à être joué en temps réel de l'utilisation de l'outil, ou enregistré dans un fichier pour un usage ultérieur.

Il est donc important de considérer les sons pour la représentation, ainsi que les techniques de synthèse qui les implémentent, comme des paramètres de la définition de tout outil de sonification.

5.4 Paramètres sonores

Une fois que les différents éléments de la sonification sont sélectionnés, il faut encore définir quels paramètres perceptifs de ces sons sont variables. Cette définition correspond à la deuxième partie du modèle perceptif de Roddy et Bridges [184] : après avoir choisi les « complexes » sonores, on choisit les « dimensions » de ces complexes qui doivent représenter métaphoriquement l'évolution de propriétés des données.

Pour beaucoup de techniques de synthèse du signal, les paramètres prennent une forme qui se rapporte assez directement à la fonction qui génère le signal. Ainsi, en synthèse additive, la fréquence d'une forme d'onde périodique donne la hauteur tonale, un facteur de l'amplitude donne la sonie, une enveloppe de volume peut être contrôlée par une série de valeurs donnant son temps d'attaque, de dégradation, de soutien et de relâchement, une fréquence de modulation de l'amplitude ou de la fréquence peut donner un timbre, etc.

Dans le cas de la synthèse par modélisation physique, les paramètres sont en fait les propriétés des éléments qui constituent le modèle. Par exemple, dans le sonogramme des données de Hermann [107], le modèle physique est construit comme un ensemble de ressorts mis en mouvement oscillatoire par l'interaction

de l'utilisateur, et les paramètres de la synthèse sonore sont ceux du système physique : les raideurs de chacun des ressorts.

Dans le cas du format **MIDI**, la paramétrisation d'un module de synthèse sonore passe par l'envoi d'un message formaté qui peut gérer la durée, la hauteur, la vitesse, la sélection d'un timbre instrumental préprogrammé, ou divers effets.

Même dans le cas d'un son pré-enregistré, il est encore possible de paramétrer, entre autres, le gain de volume qui doit lui être appliqué, sa vitesse de lecture, ou des filtres qui pourraient être appliqués à son signal (passe haut, bas, bande, une ou plusieurs fréquences de coupure).

Il est ainsi nécessaire, en plus de pouvoir définir les techniques de synthèse sonore, de préciser quels paramètres de ces techniques doivent être liés aux données ou laissés constants. Tous les paramètres de la génération du son ne sont en effet pas nécessairement variables, et certains pourront plutôt être choisis comme constants au moment de la conception de la sonification.

5.5 Fonctions de mapping

L'un des principes de base de la sonification est le fait d'effectuer des connexions, ou mappings, entre les valeurs numériques extraites des données et les valeurs numériques servant de paramètres au rendu sonore. On veut alors pouvoir indiquer quelles opérations mathématiques effectuent cette transformation.

La solution la plus courante est d'utiliser des fonctions de mapping affines, de sorte qu'à une valeur x entre x_{min} et x_{max} soit associé un paramètre y entre y_{min} et y_{max} , selon une droite de la forme $y = ax + b$.

Cependant, la perception optimale pour beaucoup de paramètres perceptifs ne se fait pas nécessairement selon une échelle linéaire. Par exemple pour la détection d'informations transmises par un paramètre de sonie, il est recommandé d'appliquer des changements par paliers plutôt que continus [158]. Afin de communiquer le passage d'un seuil, ou de faire sélectionner le comportement d'un son en fonction d'une condition dans les données, il est aussi nécessaire d'utiliser des fonctions de mapping booléennes, donc non linéaires.

Dès lors qu'on cherche à optimiser le temps d'apprentissage nécessaire pour qu'un utilisateur comprenne le rapport entre les dimension perceptives du son et les dimensions des données, il est nécessaire d'affiner les choix de métaphores sonores

mises en œuvre. Ces choix de métaphores peuvent porter sur les paramètres, les polarités ou les pentes des courbes de mapping, selon les besoins [219, 220, 215, 216, 67]. Ainsi par exemple, une température plus élevée a tendance à être associée à une hauteur tonale plus élevée, mais un objet plus volumineux est plutôt associé à une hauteur plus basse [216].

Une étude de Peres et al. [170] suggère par ailleurs que des mappings redondants, tels que plusieurs paramètres représentent une même dimension des données, peuvent être avantageux pour la détection de changements. Inversement, on peut chercher à compresser la quantité d'information à détecter à l'écoute en faisant en sorte que plusieurs dimensions des données soient représentées par un même paramètre perceptif. Par exemple, dans nos travaux sur la supervision d'un processus d'impression 3D [242, 243, 241], nous avons représenté par un même changement de sonie l'apparition d'anomalies à la fois de largeur et de hauteur dans la géométrie du dépôt de matière.

En plus de l'association de paramètres perceptifs à des dimensions des données, il est donc utile de laisser une certaine liberté quant à la description des fonctions qui effectuent ces calculs, au delà de simples fonctions affines définies par leurs bornes minimales et maximales pour ne relier qu'une seule dimension des données à un seul paramètre sonore.

5.6 Temporalité du son

Le son est un phénomène intrinsèquement temporel, et il est inévitable de devoir tenir compte de cette propriété lors de la création d'une représentation auditive de données. Comme noté à la fin de la section 5.1, les données peuvent elles-mêmes se dérouler – ou non – selon un axe temporel, et il est alors nécessaire de définir la façon dont la construction du son au cours du temps se rapporte à l'organisation des données.

Dans les applications les plus simples de la sonification, souvent pour des tâches de supervision, les données se déroulent d'elles-mêmes selon un pas de temps, ou ce déroulement peut être reproduit par leur lecture depuis un enregistrement. Alors, leur observation par la modalité auditive peut simplement se faire par translation du temps des données au temps du son.

Cependant, on peut vouloir concevoir une sonification pour un système de données statique, ne possédant pas de temporalité intrinsèque. Une stratégie simple

est alors de décider d'un axe « des abscisses », par analogie avec la représentation visuelle sur un diagramme, jouant le rôle d'axe temporel le long duquel la représentation se déroule (à une vitesse qui peut elle aussi être paramétrée).

Ce contrôle peut aussi être laissé à la discrétion de l'utilisateur, en lui offrant une tête de lecture interactive qui gère le positionnement de la représentation sur l'axe du temps (ou des abscisses) du système de données. C'est le cas par exemple dans l'outil de sonification interactive conçu par Pauletto et al. pour *Pure Data* [167], ou dans la *Sonification Sandbox* de Walker et al. [217].

Une variante de ce principe sur plusieurs dimensions peut consister à offrir à un utilisateur de balayer un plan à deux dimensions pour en activer les différents éléments sonores, comme le proposait par exemple l'outil de représentation multimodale EXVIS [87].

La technique de sonification model-based [96], est spécialement destinée à des données statiques. Son objectif est de produire une représentation de leur organisation spatiale, sans a priori sur un quelconque axe des abscisses. Dans ces cas, la temporalité du son est complètement détachée des données, de sorte à plutôt être régie par le pas temporel de la simulation physique à l'origine du son.

5.7 Interaction

À partir des années 1990, les progrès de l'informatique ont commencé à permettre la synthèse sonore en temps réel, donc la production de sonifications pouvant tenir compte de paramètres modifiés manuellement au cours de leur déroulement.

Une approche renforcée par ces progrès, et mise en avant notamment par Hermann et Hunt [103], consiste alors à créer des sonifications dont les modalités d'interaction rappellent à l'utilisateur son rapport, en tant qu'être humain, avec le monde physique. On peut ainsi « consulter » un système de données en se situant dans l'espace que celles-ci constituent, en sélectionnant des zones d'intérêt à isoler dans la représentation ou, dans le cas de la model-based sonification, en créant des points d'excitation à partir desquels un modèle physique se déroulerait.

L'interaction utilisateur constitue alors une source de données à part entière pour la paramétrisation du son, mais au lieu d'être les données que le son est censé représenter à proprement dit, les données de l'interaction servent à offrir à l'utilisateur un contrôle sur l'un ou l'autre aspect de la représentation. Pour une analogie : là où les données à sonifier constituent l'instrument de musique, les données d'interactions constituent la façon dont un musicien le fait sonner.

Une interaction peut typiquement être gérée à la souris et au clavier, comme par exemple dans le *Toolkit for Interactive Sonification* de Pauletto et al. [167], mais d'autres auteurs ont aussi entrepris de construire des appareils spécialement conçus pour l'interaction immersive avec un outil de sonification.

Par exemple, l'outil audio-haptique conçu par Hermann et al. [105] est tel que différentes actions de la main sur l'objet produisent différents effets de zooms, déplacements et sélections sur l'espace de données étudié.

Le modèle de balayage tangible sonifié proposé par Bovermann et al. [24] emploie un autre exemple de périphérique dédié, dans lequel un utilisateur balaye un espace 3D de projection des données à l'aide d'un objet planaire, de sorte que chaque point rencontré produise un son propre, selon un modèle similaire au sonogramme des données de Hermann [107].

Enfin, et beaucoup plus simplement, le fait de cliquer sur un bouton *Play* pour démarrer une sonification peut être considéré comme une forme d'interaction.

Un modèle pour la description de n'importe quel outil de sonification doit donc permettre de tenir compte de cette autre source de données qu'est l'interaction en temps réel de l'utilisateur.

5.8 Personnalisation

L'intervention de l'utilisateur sur le comportement d'une sonification peut aussi se faire à un autre niveau : celui de la modification personnalisée de certains paramètres volontairement laissés contrôlables par les concepteurs de l'outil.

À l'issue d'une évaluation de son outil *ShareMon* en 1993 [44], Jonathan Cohen notait le potentiel d'offrir aux utilisateurs un choix de l'écologie sonore qui régit la sonification, à la fois d'après leurs préférences esthétiques et leurs propres associations culturelles entre les événements et les sons. L'auteur n'a jamais, à notre connaissance, effectué d'expérience confirmant l'intérêt de cette approche, mais cette stratégie a notamment été appliquée plus tard par Barra et al. [10] dans leur projet *WebMelody*. Dans cet outil, l'utilisateur pouvait choisir une bande-son personnelle dans laquelle étaient intégrées les notifications pour la supervision d'un serveur web. D'après ces auteurs, une sonification personnalisable permet d'offrir un outil moins lassant et moins envahissant que si celle-ci était constituée d'un seul ensemble de sons fixés par les concepteurs : si les sons deviennent lassants, on peut les remplacer à volonté.

La personnalisation peut aussi permettre à chaque utilisateur de choisir les méthodes optimales pour une bonne détection des propriétés recherchées. Ainsi

dans les outils pour la création de courbes auditives, il est souvent proposé de donner aux utilisateurs l'accès à des réglages de la vitesse de lecture, ou de leur permettre de régler eux-mêmes les échelles de paramètres perceptifs dans lesquelles sont projetées les valeurs des données. Pour des cas plus complexes de sonification par paysages sonores, il a même été proposé que l'utilisateur puisse définir sa propre collection de sons à employer [230].

Certains travaux sur des outils de sonification employant un paramètre spatial ont par ailleurs proposé de laisser à l'utilisateur un contrôle sur la fonction de transfert de la tête (**HRTF**) afin de simuler au mieux le positionnement virtuel de sources sonores [30, 240]. La **HRTF** est cependant une propriété spécifique à la morphologie de chacun et nécessite, pour être réaliste, d'être mesurée individuellement. Une approximation souvent jugée suffisante consiste à donner à l'utilisateur le choix entre un nombre limité de **HRTF** standardisées correspondant à quelques morphologies types.

La gestion de la personnalisation peut paraître assez similaire à celle de l'interaction, mais elle se situe à un plus bas niveau. En effet, là où l'interaction offre une couche métaphorique qui donne à l'utilisateur l'impression de se servir d'un outil manuel ou d'un instrument de musique, la personnalisation donne plutôt un contrôle explicite sur certains paramètres choisis, lui donnant ainsi accès à une partie du processus de conception sonore. Pour reprendre l'analogie avec les instruments de musique : si les données sont les cordes et l'interaction le fait de les gratter, la personnalisation est l'action de tourner les clés de l'instrument pour modifier son accordage.

5.9 Conclusion

Dans ce chapitre, nous avons présenté une série de huit aspects variables des différentes techniques pour la représentation auditive. Celles-ci se rapportent aux choix d'un concepteur quant à la construction d'un son au cours du temps (temporalité, mappings, synthèse et paramètres pour le son), aux informations sous-jacentes que la représentation doit aider à détecter, ainsi qu'aux différents acteurs pouvant influencer la représentation auditive (l'ensemble des données elles-mêmes, ou les actions et préférences de l'utilisateur).

Dans le prochain chapitre 6, cette liste de variables informe notre étude de l'existant, en particulier pour une discussion de l'exhaustivité et de la souplesse des divers outils informatiques permettant la conception de sonifications.

Chapitre 6

Étude de l'existant

Nous effectuons dans ce chapitre une étude préalable des approches proposées par le passé pour modéliser de façon générique le processus de sonification. Ces modèles portent sur différentes étapes de la conception, mais leur but est dans tous les cas de généraliser le principe de la sonification au delà d'un départ ad-hoc systématique. Nous commençons ainsi par résumer les réflexions menées par des auteurs se basant sur des pratiques tirées du domaine du design dans la section 6.1. Ensuite, nous étudions des propositions pour modéliser le principe et les étapes d'un mapping de données à un signal sonore dans la section 6.2. Dans la dernière section 6.3, nous listons un certain nombre d'outils existants permettant de concevoir des sonifications personnalisées, tout en portant un intérêt particulier à la souplesse qu'ils apportent pour la description de sonifications, quelles qu'elles soient.

6.1 Modèles de design

Cette sous-section porte sur les travaux existants qui ont proposé une modélisation de la sonification selon une approche design. Cette approche a pour but d'aider à décider, dans les grandes lignes, d'un type de solution à appliquer et adapter face à une problématique de représentation donnée, dans un contexte donné. L'intérêt est alors dans la possibilité d'itérer et d'affiner des choix de représentation à partir d'un corpus d'existant plutôt généraliste, de sorte à ensuite contribuer à ce même corpus avec une nouvelle solution adaptée et plus spécifique.

Dans cette catégorie d'étude, on place les design patterns pour la sonification par Barrass et Adcock [15, 1], puis la description d'un espace de contexte par Frauenberger et al. [74, 75, 73], puis encore d'autres espaces de design proposés par Barrass [14], de Campo [49], puis Ludovico et Presti [137].

6.1.1 Design patterns

La description des solutions de sonification sous la forme de design patterns, initialement proposée par Stephen Barrass en 2003 [15], vise à impliquer des techniques du domaine du design dans la recherche en sonification en proposant une adaptation du langage normalement utilisé pour décrire des designs. Le but de ce format est alors de permettre la communication de résultats de recherche d'une façon facile à comprendre, reprendre et à adapter.

Selon ce format, on définit un design de sonification comme une « solution à un problème dans un contexte donné ». Le langage employé consiste à remplir une phrase à trous décrivant le contexte et la problématique, puis une solution, sa justification et ses conséquences. Traduite en français, cette démarche donne la formulation suivante :

```
[Nom du pattern]
Si l'on se trouve dans [un contexte],
par exemple [au moins trois exemples],
avec [une problématique]
impliquant [des forces et contraintes],
alors
pour [des raisons]
appliquer [des règles de design]
pour construire [une solution]
menant à [un nouveau contexte]
et [un autre pattern].
```

Le format a ensuite été réorganisé par Adcock et Barrass en 2004 [1] pour ne plus imposer la formulation en phrase à trous. À la place, un design est décrit par une suite structurée de paragraphes : contexte, problématique, forces, solutions, raisonnement, exemples.

À partir de cette proposition pour la communication de sonifications, les auteurs ont tenté de centraliser la collaboration de chercheurs souhaitant décrire leurs techniques en proposant un site de type Wiki¹. La section de ce site dédiée à la sonification n'a pas été mise à jour depuis 2010, et ne semble pas avoir réussi à réunir un nombre particulièrement grand de contributions.

1. <http://c2.com/cgi-bin/wiki?PatternIndex> – erreur 404 au 4/1/23

Entre 2007 et 2009, Frauenberger et al. [74, 75, 73] ont ensuite travaillé sur une extension de cette adaptation des design patterns pour la sonification. Partant du constat que la contribution de Barrass et al. n'avait pas eu le succès espéré, leur idée était d'y apporter un cadre méthodologique qui puisse faciliter la découverte, la création, l'application et l'évolution de designs sonores. Ils proposent ainsi de ranger les design patterns dans un espace de six dimensions donnant les caractéristiques du contexte :

- Quatre dimensions ordinales (échelles entre 0 et 5) qui caractérisent la spécificité du public visé, de l'environnement d'utilisation, de l'appareil sur lequel la sonification doit avoir lieu et du domaine d'application.
- Ces quatre dimensions peuvent être renforcées par des tags textuels pour apporter plus de précisions sur le contexte (par exemple : « utilisateurs aveugles », « environnement bruyant » ou « pour smart phone »).
- Deux dimensions supplémentaires, purement textuelles, peuvent contenir des tags portant sur l'expérience utilisateur prévue (par exemple : « amusante », « immersive ») et le contexte social (par exemple : « application de rencontres », ou « public jeune ») de l'outil.

Une application en ligne s'occupait alors de relier les design patterns entre eux, à la fois par proximité d'après leurs dimension ordinales et par recoupement des tags décrivant les différents critères du contexte. Les utilisateurs pouvaient s'en servir pour rechercher des design patterns se rapportant à leur problématique, ou du moins s'en approchant. Un utilisateur ayant produit un nouveau pattern plus ou moins spécifique à un contexte était encouragé à l'ajouter à la base de données à l'aide d'un formulaire interactif des descripteurs du contexte.

6.1.2 Espaces de design

TaDa, proposé par Barrass en 1996 [14] consiste à organiser les choix de design pour la sonification selon d'une caractérisation des données à représenter et de la tâche associée à la représentation. Cette modélisation est accompagnée d'un cadre logiciel, *Personify* [11], qui permet au designer de décrire sa problématique comme un scénario et de sélectionner les mots-clés associés à ses besoins. Le scénario est recherché dans la base de données *EarBenders* [12], qui contient des exemples de situations du quotidien où l'écoute est source d'information. Un prototype de représentation correspondant au scénario le plus pertinent est alors généré et l'utilisateur peut explorer l'espace des paramètres perceptifs (un autre travail de Barrass, conçu par analogie aux espaces de couleurs [13]) afin de trouver une représentation à sa convenance.

Le travail de Frauenberger et al. entre 2007 et 2009 [74, 75, 73], déjà cité plus haut, consistait initialement à construire un espace à 9 dimensions ordinales, avec en plus 3 critères nominaux, pour la classification des contextes dans une base de données dont ils pourraient ensuite être extraits en réponse à de nouvelles descriptions de contextes. Cependant, l'idée a ensuite évolué vers le système de 6 critères nominaux – dont 4 sont par ailleurs quantifiés sur une échelle – décrit plus haut. Ce compromis semble traduire une certaine difficulté à synthétiser d'une façon représentative les critères pour la description d'un contexte sans poser des questions ouvertes. L'approche rappelle finalement, dans une certaine mesure, celle de Barrass pour *TaDa*, qui recherchait déjà des solutions dans une base de données à partir d'un certain nombre de descripteurs textuels sur la problématique de représentation.

Alberto de Campo a proposé en 2007 de classer les designs de sonification dans un espace à trois dimensions [49]. Un premier axe de cet espace donne la quantité de points des données nécessaires à la détection d'une gestalt (une tendance, une particularité, un point d'intérêt perceptif), c'est-à-dire à quel point il est important de distinguer chacun des points individuellement ou seulement de repérer des tendances dans l'ensemble qu'ils forment. Le deuxième axe donne le nombre de propriétés, ou dimensions, des données qui doivent être traduites par la représentation. Le troisième axe donne le nombre de flux sonores attribués aux différentes propriétés à représenter.

Sur la carte visuelle de cet espace, quelques approches usuelles de sonification (notes de musique individuelles, textures sonores, parameter mapping, audification) sont positionnées, encadrées par une nouvelle classification des techniques en trois grandes catégories proposées pour l'occasion par de Campo et portant sur la granularité du son construit par rapport aux données : model-based (intégration des données dans un modèle physique) pour une forte dimensionnalité des données, event-based (des points sonores distincts) pour de petites quantités de points, ou stream-based (des flux sonores continus) pour plus de points.

Il est ainsi possible de positionner une problématique sur cette carte, de repérer la technique de sonification la plus proche et d'itérer sur un design qui s'en inspire en changeant les critères associés au positionnement dans l'espace. Par exemple en changeant de résolution temporelle, en compressant les dimensions des données, ou en ajoutant des flux sonores.

Ludovico et Presti ont proposé en 2016 un espace de design qui tente d'offrir plus de précision quant aux choix de représentations, au delà de la simple sélection

d'une technique [137]. Leur espace se déroule ainsi selon deux dimensions : un axe quantifiant la granularité temporelle, de façon assez similaire à ce que proposait de Campo (ci-dessus), et un axe quantifiant l'abstraction des sons employés (plus ou moins naturels ou synthétiques). Cependant, de leur propre aveu, le rangement des techniques usuelles dans cet espace paraît assez contre-intuitif et leurs méthodes de calcul automatique des critères proposés à partir des signaux sonores ne suffisent pas à bien rendre compte des mécanismes cognitifs mis en œuvre.

6.1.3 Discussion

Si ces modèles facilitent bien la sélection d'une solution en fonction d'un problème-contexte, ils ne semblent pas spécialement prévus pour permettre à leur utilisateur de détailler l'implémentation de son outil par la suite. Au delà de quelques généralités liées à la famille de techniques la mieux associée à un type de problème (par exemple : « Pour superviser une série de données multidimensionnelles au cours du temps, on construit un paysage sonore traduisant par l'évolution de paramètres perceptifs les divergences de ces dimensions par rapport aux seuils tolérés. »), il est plus difficile de décrire sous un format de design toutes les subtilités d'un outil (« Quel genre de données multidimensionnelles ? », « À quel(s) paramètre(s) perceptif(s) associer chacune ? », « Selon quelles opérations ? »).

La modélisation que nous recherchons, qui permettrait de détailler les aspects listés dans la section 5, s'inscrit en réalité à la suite méthodologique de la réponse apportée par ces approches dans le domaine du design : une fois qu'une solution plus ou moins générique a été trouvée pour le contexte auquel on cherche à appliquer une sonification, on peut effectuer une description plus précise des choix pour l'application spécifique basée sur cette solution.

6.2 Modèles de mapping

Dans cette sous-section, nous nous intéressons aux modélisations existantes de la sonification dont la démarche consiste à formaliser la description des mappings mis en place entre les données et les sons.

Nous étudions ainsi quelques débuts de formulations de la sonification comme la fonction de transfert des données vers le son, puis certaines considérations, empruntées au domaine des instruments musicaux informatiques, pour l'expression des étapes successives d'un mapping.

6.2.1 Fonction de transfert données-son

D'après la définition de la sonification par parameter mapping donnée par Hermann dans sa thèse [96], la sonification peut s'exprimer de façon algébrique comme une fonction de transfert donnant l'échantillon du signal s à l'instant t en fonction d'un ensemble de N points de données d'entrée $\{X_1, \dots, X_N\}$:

$$s(t) = \sum_{i=1}^N f(g(X_i), t) \quad (6.1)$$

Avec f la fonction qui synthétise un flux sonore en fonction d'un certain nombre m de paramètres, eux-mêmes donnés par la fonction de mapping g appliquée à l'ensemble des d dimensions variables $x_{i,j}$, ($j \in [1; d]$) de chaque point X_i . Cette fonction g consiste en un vecteur de m fonctions h_k , ($k \in [1; m]$), chacune traitant une dimension x_{i,j_k} , ($j_k \in [1; d]$), et dont les sorties consistent chacune en une seule dimension.

Cette expression est intrinsèquement liée à une forme très basique de parameter mapping selon laquelle chaque paramètre dépend d'une seule dimension des données, de sorte à mixer une série de N signaux tous générés selon une même fonction f . Elle ne permet alors pas, pour donner un exemple simple, d'exprimer une représentation selon laquelle un certain groupe de dimensions serait représenté par des modulations en fréquence d'ondes sinusoïdales, tandis qu'un autre correspondrait à des filtres appliqués à des bruits blancs.

Dans un article de 2007 présentant un outil d'optimisation évolutive du parameter mapping [101], Hermann, Bunte et Ritter ont ensuite proposé une nouvelle expression générique de la sonification comme une fonction $s(X, \theta, T)$, cette fois prenant comme variables des ensembles de données d'entrée X , de paramètres de mappings réglables θ et des données d'interaction T . Un exemple d'application est alors donné pour une sonification par parameter mapping assez simple :

$$s(X, \theta, T) = \sum_{\alpha=1}^N \phi(t, \vec{p}(x^{\vec{\alpha}}, \theta)) \quad (6.2)$$

Dans l'expression ci-dessus, le son est construit à chaque instant t comme la superposition de sons générés par une fonction de synthèse unique ϕ , paramétrée par chacune des N lignes des données d'entrée. Le vecteur de paramètres est calculé par une fonction \vec{p} appliquée à la ligne $x^{\vec{\alpha}}$ en fonction de θ . Il s'agit dans cet exemple d'un mapping affine tel que θ est un tuple (A, \vec{b}) et $\vec{p}(x^{\vec{\alpha}}, \theta) = Ax^{\vec{\alpha}} + \vec{b}$.

L'interaction utilisateur ne jouant pas de rôle dans cet exemple, l'ensemble T est omis du membre de droite de l'expression.

Comme pour l'expression 6.1, l'exemple d'application décrit par 6.2 est très spécifique et ne permet pas la représentation abstraite de n'importe quelle sonification dès l'instant que l'on souhaiterait, par exemple, utiliser plus d'un type de son ou plus d'une opération de mapping pour les différentes dimensions des données.

On note cependant que l'idée d'exprimer le transfert données-sons sous une forme algébrique semble présenter un fort potentiel pour la modélisation généraliste de la sonification, celle-ci pouvant, après tout, toujours être vue comme une fonction de synthèse du signal paramétrée par le temps et les données d'entrée.

En 2010, Rohrhuber propose un nouveau formalisme générique de la fonction de transfert pour la sonification, notée \mathring{S} [185]. Dans cette formulation, le temps du son est un paramètre noté \mathring{t} qui, selon les choix de la représentation, peut dépendre ou pas du temps des données.

La fonction donnant un signal sonore sur n canaux au cours d'une variable temporelle réelle $\mathring{t} \in \mathbb{R}$, et en fonction d'un ensemble de données \mathbb{A} , a alors une signature de la forme :

$$\mathring{S} = \mathbb{A} \times \mathbb{R} \rightarrow \mathbb{R}^n \quad (6.3)$$

On peut ainsi, par exemple, adapter la fonction de transfert du parameter mapping donnée dans 6.1 pour en faire :

$$\mathring{S}(\mathring{t}) = \sum_{i=1}^N f(g(X_i, t), \mathring{t}) \quad (6.4)$$

Selon cette écriture, les fonctions de mappings g peuvent, ou pas, tenir compte de l'évolution temporelle des données selon leur variable de temps t , tandis que le son est construit selon sa variable temporelle propre \mathring{t} . Cette variable \mathring{t} peut par exemple être exprimée en fonction de t , ou peut dépendre d'une interaction de l'utilisateur.

Il reste à notre sens un certain nombre de degrés de liberté du design à prendre en compte dans l'expression du mapping données-sons. Pour cette raison, nous étudions plus avant l'adaptation de l'expression 6.2 donnée par Hermann et al. [101], ainsi que les conséquences d'une variable temporelle propre à la sonification comme proposée par Rohrhuber [185], dans le chapitre 7 dédié à la description du modèle générique.

6.2.2 Mappings entre espaces perceptifs

À son tour, une telle représentation algébrique du mapping données-sons se prête à un recoupage en fonctions plus simples, combinées les unes aux autres et gérant les différentes étapes du mapping. On peut ainsi indiquer explicitement les calculs de propriétés sous-jacentes des données, puis le mapping de ces propriétés sur des échelles perceptives appropriées. Un domaine auquel il semble intéressant d'emprunter des méthodes pour ces considérations est celui, cousin de la sonification, de la programmation d'instruments musicaux contrôlés par les gestes.

Ainsi Arfib et al. [7] ont proposé pour ce genre de tâches de compartimenter l'enchaînement des opérations de mapping selon trois couches de calcul passant par des espaces qui tiennent compte de la perception humaine :

- Perception des gestes : interprétation des mesures physiques brutes en données caractéristiques de la gestuelle.
- Perception du son : traduction de ces caractéristiques en propriétés perceptives quantifiables, pour leur représentation dans un son.
- Synthèse du son : traduction des propriétés perceptives en paramètres de synthèse sonore, selon les modalités de génération de son employées.

Ce modèle s'inspire d'un autre, proposé par Wanderley et al. en 1998 et employé dans l'outil de mapping gestuelle-sons *ESCHER* [223]. Cette version antérieure du modèle ne comprenait que deux couches pour l'interprétation des données en paramètres de gestuelle et la traduction de cette gestuelle en paramètres de synthèse sonore, sans passer par la couche de paramètres perceptifs.

L'approche d'Arfib et al. offre le double avantage de représenter l'opération de sonification sous une forme algébrique visuelle, donc claire, et de passer par une couche de mapping qui rend les choix effectués explicites selon une optique psychoacoustique.

Les auteurs présentent exclusivement cette modélisation dans un cadre de création d'instruments musicaux informatiques et interactifs, mais elle semble pouvoir être généralisée à la sonification en remplaçant le premier espace de la « perception des gestes » par un espace plus général de « propriétés extraites des données ». Une telle généralisation est au moins triviale dans les cas où la temporalité du son suit celle des données puisque, à l'instar de la gestuelle musicale, le son est alors généré en temps réel de l'évolution des données.

6.3 Outils généralistes de sonification

En 1999, les auteurs du *Sonification Report* soulignent le besoin, dans le domaine de la sonification, d'un cadre logiciel générique pour la création de sonifications [128]. La recherche d'un modèle théorique encadrant n'importe quelle sonification va de pair avec ce besoin, puisqu'un tel outil sera forcément conçu pour permettre plus ou moins de liberté dans la description des différents aspects d'une représentation auditive.

Le *Sonification Report* énonce trois principaux besoins liés à cet outil hypothétique. En plus de la portabilité (l'outil est facilement adapté à toute plateforme informatique) et l'intégrabilité (possibilité d'adjoindre l'outil à des environnements externes de visualisation ou d'interaction) qui sont à notre sens des besoins plutôt liés à l'outil en tant que programme, le besoin de flexibilité (adaptabilité à n'importe quel cas de figure) paraît le plus important à satisfaire dans la théorisation d'une boîte à outils idéale pour la sonification.

La création d'un tel outil a fait l'objet de recherches actives pendant de nombreuses années, et de nombreuses équipes se sont attelées à la tâche de produire un méta-outil de sonification pour l'expérimentation, le prototypage ou le développement d'applications plus spécifiques. En effectuant une étude systématique de la flexibilité offerte par chacun de ces outils, nous cherchons à comprendre ce qui manque toujours pour faire qu'aucun, à notre connaissance, n'a encore été largement adopté par la communauté.

Nous avons donc répertorié dans le tableau 6.3 18 outils généralistes de sonification publiés depuis les années 80 et régulièrement cités dans la recherche. Les critères d'acceptation pour cette liste ont été ceux décrits dans les quatre paragraphes suivants.

L'outil ne part d'aucun a priori sur la teneur des données à représenter. Par opposition à des outils spécifiques qui présupposent que les données d'entrée comportent un certain nombre de dimensions portant sur un système propre à un domaine d'étude, les outils auxquels nous nous intéressons ici traitent de données qui n'ont pas besoin de se rapporter à quoi que ce soit. *xSonify* [32] a été toléré dans notre liste, bien qu'il ne traite que des données de recherche spatiale, parce qu'il est très souvent cité parmi les outils de sonification personnalisée dans la littérature, et que ces données spatiales sont avant tout de simples séries temporelles. Par ailleurs, si un certain nombre d'outils génériques pour la conception d'interfaces utilisateur auditives a été proposé par le passé [171, 122, 48], nous considérons que ce type d'application constitue un domaine encore trop spécifique pour notre étude.

L’outil permet à l’utilisateur de personnaliser tout ou partie des modalités et paramètres de la synthèse sonore. C’est-à-dire qu’au lieu de fournir une représentation préalablement programmée pour tenir compte de considérations psychoacoustiques liées à un domaine d’étude, l’outil propose un certain nombre d’options pour définir soi-même la paramétrisation d’un son synthétisé en fonction des données d’entrée.

L’outil a fait l’objet d’une publication dont il est le sujet principal, que ce soit dans un article scientifique ou sur le web. Ainsi, si certains outils comme *OctaveSC* [100] ou *Real-time Audio Toolkit for Sonification (RATS)* [105] sont évoqués en passant dans des articles sur des applications spécifiques qu’ils ont permises, il est tout de même nécessaire de pouvoir se référer à une notice explicite pour comprendre dans le détail ce qu’ils permettent de faire ou pas.

L’outil est spécialement prévu pour la sonification. Ainsi, même si des outils de synthèse sonore comme *SuperCollider* ou *Pure Data* permettent entre autres la sonification de données, on s’intéresse ici aux outils (d’ailleurs parfois intégrés à ces environnements) dont les fonctionnalités sont tout particulièrement destinées à encadrer et faciliter cette activité.

Ces outils ont été notés selon des échelles de 0 à 5 qui permettent de juger de leur souplesse pour la construction de sonifications d’après les différents degrés de liberté énoncés dans la section 5. Pour chacun de ces aspects, les 6 degrés de l’échelle se rapportent aux propriétés suivantes :

- 5 → Les fonctionnalités sont extensibles grâce au cadre fourni par l’outil lui-même.
- 4 → Les fonctionnalités sont extensibles parce que l’outil est intégré à un environnement facilitant leur implémentation.
- 3 → Les fonctionnalités sont extensibles parce que le code de l’outil est libre d’accès et de modification.
- 2 → L’outil propose un nombre limité d’options parmi lesquelles l’utilisateur peut choisir.
- 1 → L’outil ne traite cet aspect que d’une seule façon.
- 0 → L’outil ne permet pas de traiter cet aspect, ou l’information n’est pas donnée dans la publication.

Il est important de noter que les valeurs plus ou moins élevées ne sont pas nécessairement représentatives de la qualité objective de l’outil, mais plutôt de sa souplesse pour la création d’outils de sonification spécifiques, quels que soient le domaine d’application et les modalités de représentation.

Les degrés 3 et 4 sont en particulier représentatifs de l'extensibilité de l'outil, soit parce qu'il est open source, soit parce qu'il est intégré à un environnement plus large dont le fonctionnement permet une certaine souplesse sur les fonctionnalités en question. Par exemple des outils intégrés à *Excel* [205] ou *Matlab* [149] fournissent de facto un grand nombre de fonctionnalités pour le pré-traitement des données, même si l'outil de sonification lui-même apporte peu ou pas de nouvelles fonctionnalités sur ce front. De même, des outils intégrés à *SuperCollider* [50] ou *Pure Data* [167] sont bien équipés pour des extensions ultérieures des techniques de synthèse sonore, même s'ils n'apportent eux-mêmes que quelques fonctionnalités de base.

Ainsi dans le tableau, certains des outils ont reçu deux notes pour certains critères : la première pour l'étendue réelle des options offertes par l'outil tel quel et la deuxième, entre parenthèses, pour l'étendue potentielle des options qui pourraient être ajoutées à l'outil sous réserve de « regarder sous le capot ». On note par exemple le cas de *SonART* [16], initialement proposé par Ben-Tal et al. pour lire des fichiers de données, puis étendu par Yeo et al. pour l'interaction en temps réel par communication *OSC* [228].

Le degré 3 sur le statut open source ne permet pas seul de juger de l'extensibilité d'un outil, mais en tout cas de la disponibilité publique de son code. Ainsi par exemple, le code de la *Sonification Sandbox* [217] étant publiquement disponible, il serait techniquement possible d'étendre cette application pour que les sonorités utilisées soient plus souples que le seul format *MIDI* actuellement implémenté, mais rien ne garantit que cette modification ne nécessite pas une refonte complète de son architecture.

Les lignes du tableau dans lesquelles on observe les « meilleures » notes sont souvent dues à un cadre de contrôle assez bas niveau, lié au fait que l'outil est intégré dans un environnement de programmation existant. C'est par exemple le cas des fonctionnalités de gestion de l'audio dans le langage *R* [108], qui est déjà largement adapté par ailleurs aux traitements des données et des signaux.

On remarque le cas particulier de *SoniPy*, pour lequel il est assez difficile de déterminer si les notes devraient être 4 ou 5, et qui semblerait en tout cas particulièrement adapté à tous les besoins énoncés, mais qui consiste essentiellement en une compilation et encapsulation de diverses fonctionnalités propices à la sonification dans le langage *Python*. Par ailleurs, à notre connaissance, ce projet semble être resté à un stade essentiellement théorique depuis sa description initiale en 2007 [232].

NOM (année)	Données	Pré-traitement données	Synthèse sonore	Paramètres synthèse	Mappings	Temporalité	Interaction
EXVIS (1989) [87]	5	5	1 (3)	1 (3)	2 (3)	2 (3)	2 (3)
Personify (1995) [11]	2	0	2	2	2	2	1
PorSonify (1995) [139]	4	4	4	4	5	1	4
LISTEN (1996) [227]	4	4	2 (3)	2 (3)	2 (3)	1 (3)	4
MUSE (1997) [135]	4	4	4	2 (3)	2 (3)	1 (3)	4
R. sound package (2002) [108]	4	4	5	5	4	4	4
SonART (2002) [16]	3	3	4	5	3	2	3
SKDTools (2003) [149]	4	4	1 (3)	1 (3)	1 (3)	1 (3)	1 (3)
Sonification Sandbox (2003) [217]	2 (3)	2 (3)	1 (3)	1 (3)	2 (3)	1 (3)	1 (3)
PD Toolkit for ISon (2004) [167]	1 (3)	2 (3)	2 (4)	2 (4)	2 (4)	2 (3)	2 (3)
SonEnvir (2004) [51, 50]	2 (3)	2 (3)	4	4	2 (3)	2 (3)	2 (3)
Excel sonification (2005) [205]	4	4	1	1	1	1	1
MathTrax/MDE (2005) [196]	2 (3)	2 (3)	2 (3)	1 (3)	1 (3)	1 (3)	2 (3)
XSonify (2006) [32]	2 (3)	2 (3)	2 (3)	2 (3)	2 (3)	1 (3)	2 (3)
SoniPy (2007) [232]	4-5	4-5	4-5	4-5	4-5	4-5	4-5
SoniYer (2008) [58]	2	2	2	2	2	1	2
Rotator (2017) [41]	2 (3)	2 (3)	2 (3)	2 (3)	2 (3)	2 (3)	2 (3)
Sonif. Workstation (2019) [172]	2 (3)	0	2 (3)	2 (3)	2 (3)	2 (3)	2 (3)

TABLE 6.1 – Étude des degrés de liberté offerts par divers outils généralistes de sonification

PorSonify semblait aussi, à en croire la publication de Madhyastha et al. [139], suffisamment souple pour permettre de faire à peu près n'importe quoi en termes de sonification. Cependant, d'après une publication ultérieure dans laquelle Wilson et al. décrivent leur propre outil *LISTEN* [227], la prise en main de *PorSonify* serait particulièrement difficile. Entre autres, sa portabilité se faisait au prix de devoir systématiquement écrire à la main une interface avec tout nouveau matériel audio.

Les scores les moins élevés sont souvent dûs à des objectifs assez spécifiques comme l'exploration sonifiée de données (outil *Excel* pour utilisateurs mal voyants [205], ou *MathTrax* [196]), où l'emphase est sur la généricité vis-à-vis des données tandis que les options de paramétrisation de la synthèse sonore sont plus restreintes.

La plupart des outils qui se présentent sous la forme d'interfaces (*Personify* [11], *Sonification Sandbox* [217], *Pure Data Toolkit for Interactive Sonification* [167], *xSonify* [32], *SonifyYer* [58], *Rotator* [41], *Sonification Workstation* [172]) ont aussi tendance à avoir des notes dans les 1 ou 2, puisqu'ils proposent toujours au moins quelques fonctionnalités essentielles pour permettre à un sonificateur novice de produire des prototypes, et c'est leur caractère open source ou non qui détermine leur extensibilité.

6.3.1 Remarques finales sur les outils généralistes de sonification

Comme le font très justement remarquer Phillips et Cabrera dans leur article sur la *Sonification Workstation* [172], les outils de sonification généralistes ont tendance à se placer sur un spectre qui force un compromis entre la facilité de prise en main et la flexibilité. D'un côté se trouvent les interfaces utilisateurs qui proposent seulement quelques fonctionnalités basiques de parameter mapping, éventuellement d'audification, et de l'autre côté se trouvent les langages de programmation, qui nécessitent un apprentissage plus poussé mais permettent de produire virtuellement n'importe quelle sonification envisageable.

Dans la présente thèse nous soutenons cependant que, sous réserve de réduire le procédé de sonification à ses composantes les plus élémentaires, d'organiser l'architecture de ces composantes sous une forme concise mais exhaustive et de considérer comme abstraits, réglables et extensibles les différents aspects variables de la chaîne de traitement, il est possible de construire un modèle de représentation de la sonification englobant n'importe quelle tâche de traduction de données en sons.

Nos réflexions sur le format d'un tel modèle sont rapportées dans le chapitre 7, puis des perspectives sur l'implémentation d'un outil généraliste basé sur ce modèle sont rapportées dans le chapitre 9.

6.4 Conclusion

Dans ce chapitre, nous avons étudié les différentes approches déjà proposées par la littérature pour la création d'un cadre générique de conception de la sonification.

Si les approches inspirées du domaine du design semblent prometteuses pour la classification des techniques de sonification par rapport aux besoins et contexte, elles présentent cependant un certain manque de lisibilité de par un format souvent exclusivement textuel, et de reproductibilité de par leur manque de précision concrète sur la mise en œuvre des solutions de sonification proposées.

D'après l'étude rapportée dans la section 6.3, les outils informatiques pour la conception générique de sonifications semblent pour la plupart restreints en réalité à un nombre assez limité d'usages.

Cela semble dû, d'une part, à un compromis qui semble inévitable entre l'ergonomie et le contrôle, tel qu'un outil à la prise en main facile n'offre souvent que quelques fonctionnalités basiques tandis qu'un outil plus puissant requiert plus d'apprentissage ou des notions de programmation plus ou moins avancées.

D'autre part, on note une certaine tendance des auteurs à considérer qu'une sonification est nécessairement synchronisée avec l'évolution des données, de sorte que les techniques proposées par les outils se rapportent très souvent à une **PMSon** ou une audification. Certains outils plus rares peuvent proposer de plutôt paramétrer des modèles interactifs pour la représentation de données réparties dans un espace et statiques dans le temps. Cependant, à notre connaissance, ces deux approches n'ont encore jamais coexisté dans un même outil.

Pour notre propre modèle de description de la sonification, dont nous posons les bases dans le chapitre 7 suivant, nous tentons de palier à ces manques de lisibilité, de précision et de souplesse en nous inspirant de notions tirées des travaux étudiés dans la section 6.2 : la temporalité de la sonification pourrait gagner à être considérée du point de vue de la synthèse sonore plutôt que des données, et une attention particulière devrait être accordée à l'écriture explicite des fonctions qui expriment les sons en fonctions des données.

Chapitre 7

La Cellule de Sonification

Dans ce chapitre, nous présentons élément par élément le modèle générique proposé pour la description d'outils de sonification.

Ce modèle consiste en une représentation de l'enchaînement d'opérations qui contrôlent les paramètres de fonctions de synthèse du signal, à partir d'informations en provenance d'un environnement qui gère l'interface utilisateur et les sources de données.

Nous avons nommé ce modèle la « Cellule de Sonification » (*Sonification Cell* en anglais) en référence à l'expression « *sonification shell* » employée par les auteurs du document fondateur *Sonification Report* [128] pour décrire un hypothétique outil de sonification qui soit générique, portable, extensible et facilement intégré à des environnements existants. Au moment de la rédaction de ce document, le modèle est encore théorique, mais nous anticipons que sa structure puisse être adaptée en architecture logicielle par la suite. Par ailleurs, cette structure dont le contour (les sources de données et le rendu sonore) assure l'interface avec l'extérieur et dont le contenu (les opérations de mappings) gère l'information propre à sa fonction se prête assez bien à une analogie avec la structure d'une cellule, elle-même dotée d'une membrane et d'un noyau.

7.1 Critères pour le modèle

L'une des principales motivations citées par Stephen Barrass [15] pour appliquer à la sonification des principes liés au design est le besoin de pouvoir facilement transmettre, d'un concepteur à un autre et dans un langage commun, les techniques employées pour répondre à des problématiques données. De même, notre but lors de la recherche d'une modélisation de la sonification est d'offrir la possibilité de transmettre, puis découvrir, reproduire et adapter un design de sonification. Comme détaillé dans la liste ci-dessous, ces objectifs correspondent

à une série de qualités attendues de notre modélisation : la lisibilité pour la découverte de solutions, la modularité pour l'adaptation de solutions et l'exhaustivité pour la transmission.

- **Lisibilité pour la découverte** Afin de permettre qu'une technique de sonification soit reconnue et comprise facilement, il est nécessaire que le modèle en offre une représentation claire et lisible.
- **Modularité pour l'adaptation** Pour la conception de représentations auditives, qui est une science encore largement expérimentale, il est utile de pouvoir reprendre une partie d'un design existant et d'en modifier une autre afin de l'adapter à des besoins proches mais distincts. Un concepteur peut aussi avoir besoin de développer un design itérativement, en modifiant tout ou partie de ses choix de représentation au fur et à mesure des retours d'utilisateurs. Ces différentes parties du procédé de représentation gagnent donc à être explicitement séparées, sous la forme de modules, dans le cadre apporté par le modèle.
- **Exhaustivité pour la transmission** Le modèle doit permettre de tenir compte de toutes les informations propres à la création d'une représentation auditive. Ainsi, une personne extérieure souhaitant s'inspirer de choix de représentation pour une problématique proche de la sienne peut en tirer une compréhension complète d'après sa lecture de la modélisation. Le chapitre 5 de la présente thèse donne une liste détaillée des aspects variables pour les diverses approches à la sonification de données.

7.2 Vue d'ensemble

Le principe de la Cellule de Sonification est de décrire comme une grande fonction de transfert le calcul d'un signal sonore en fonction d'un ensemble comprenant les données d'entrée, les paramètres réglables et les données d'interaction de l'utilisateur. Dans cette sous-section, nous commençons par décrire le cheminement, d'après l'existant, qui a mené à la formulation de ce modèle, puis nous présentons la forme graphique sous laquelle il sera le plus souvent utilisé par la suite.

7.2.1 Fonction de transfert de la sonification

En 2007, Hermann, Bunte et Ritter [101] ont décrit la fonction de sonification, quelle qu'elle soit, comme une fonction $s(X, \theta, T)$ pour un ensemble de données X , un ensemble de paramètres de mapping θ et un ensemble de données d'interactions T . Dans ce même article, les auteurs donnent ensuite un exemple pour l'expression

d'une sonification assez simple, de type parameter mapping. Cette expression est recopiée dans la formule 6.2 de la section 6.2 de l'étude de l'existant. Ci-dessous, nous décrivons comment les différents aspects de cette expression spécifique à un usage peuvent être adaptés à une description plus générale de la sonification. Pour rappel :

$$s(X, \theta, T) = \sum_{\alpha=1}^N \phi(t, \vec{p}(x^{\vec{\alpha}}, \theta)) \quad (6.2)$$

Membre de gauche Le paramètre temporel t apparaît seulement dans le membre de droite, mais rien ne semble empêcher de considérer s elle-même comme une fonction de synthèse du signal au cours du temps, donc de la formuler selon un paramètre temporel supplémentaire : $s(t, X, \theta, T)$.

On considère alors que X , θ et T sont des fonctions du temps, de signatures : $\mathbb{R}^+ \rightarrow \mathbb{R}^d$, avec d le nombre de dimensions des données, réglages ou interactions respectivement. On écrit ainsi $X(t)$, $\theta(t)$ et $T(t)$ pour accéder à l'état de ces sources de données à un instant t .

Membre de droite En représentation auditive, il est rarement judicieux de représenter tous les éléments par un seul et même timbre [70]. L'unique fonction de synthèse sonore ϕ donnée dans l'expression 6.2 devrait donc plutôt être une collection de fonctions de synthèse correspondant à un nombre donné M de complexes sonores. On écrit ces fonctions de synthèse ϕ_i , indexées par i entre 1 et M .

Le fait de donner le signal sonore comme la somme $\sum_{\alpha=1}^N$ des sons synthétisés pour chaque vecteur ligne $x^{\vec{\alpha}}$ fonctionne dans l'exemple de [101] puisque l'idée est de synthétiser autant de sons simultanés qu'il y a de lignes de données, mais de façon plus générale c'est surtout l'ensemble des M complexes sonores synthétisés à chaque instant que l'on superpose dans le signal. La somme principale de l'expression devrait ainsi être $\sum_{i=1}^M$. Cette forme est compatible avec le cas particulier susmentionné, puisque celui-ci consiste à considérer que chacune des N lignes des données est représentée par son propre complexe sonore, de sorte que tous ces complexes sont synthétisés par une fonction $\phi_i = \phi$ pour tout i entre 1 et M , avec $M = N$.

Chacune des fonctions ϕ_i prend potentiellement des paramètres différents, calculés en fonction de différents sous-ensembles des données, paramètres et interactions. Ainsi, comme pour les complexes sonores, on indexe par i le vecteur des paramètres obtenus par mapping : \vec{p}_i .

Ainsi, l'alternative générique que nous proposons pour l'expression initialement donnée par Hermann et al. a la forme suivante :

$$s(t, X, \theta, T) = \sum_{i=1}^M \phi_i(t, \vec{p}_i(t, X(t), \theta(t), T(t))) \quad (7.1)$$

Dans cette expression, $s(t, X, \theta, T)$ est la valeur du signal sonore en fonction du temps t de déroulement de la sonification, des données d'entrée X , des paramètres réglables θ et des données d'interaction T . Pour un même instant t , $X(t)$, $\theta(t)$ et $T(t)$ sont les états instantanés de ces sources de données.

M est le nombre total de complexes sonores constituant la représentation. Pour i entre 1 et M , ϕ_i est la fonction de synthèse sonore produisant le $i^{\text{ème}}$ complexe sonore de la sonification et \vec{p}_i est la fonction donnant le vecteur des paramètres pour cette fonction.

7.2.2 Représentation graphique de la fonction de transfert

S'il est techniquement possible de donner la construction de la totalité des complexes sonores et les calculs de chacun de leurs paramètres sous la forme d'une fonction de transfert, ainsi que de décrire les sources de données sous une forme textuelle détaillée, une telle représentation peut rapidement compliquer la lecture humaine en formant de longues listes d'expressions mathématiques rallongées par l'accumulation d'opérations pour le traitement du flux de données. Pour des sonifications particulièrement complexes, une forme textuelle peut par ailleurs compliquer la tâche de planification en ne permettant pas d'obtenir une vue d'ensemble sur les choix effectués.

Nous proposons donc que, en complément d'une description textuelle complète, l'écriture de la sonification en tant que fonction de transfert passe par une représentation graphique explicitant les sources des différents types de données pour la paramétrisation, ainsi que leurs passages dans des blocs d'opérations successifs et leurs rôles dans la synthèse sonore. La figure 7.1 donne une vue d'ensemble de la structure d'une telle représentation graphique du modèle.

Dans l'optique future de l'adaptation de ce modèle en un outil informatique pour la création de sonifications, qui serait dédié à un public (relativement) large de concepteurs ne bénéficiant pas toujours de connaissances en programmation textuelle, une telle analogie graphique à un « circuit électrique » pourra offrir une plus grande liberté et une prise en main plus facile pour la planification des mappings, en plus d'une vue d'ensemble plus lisible.

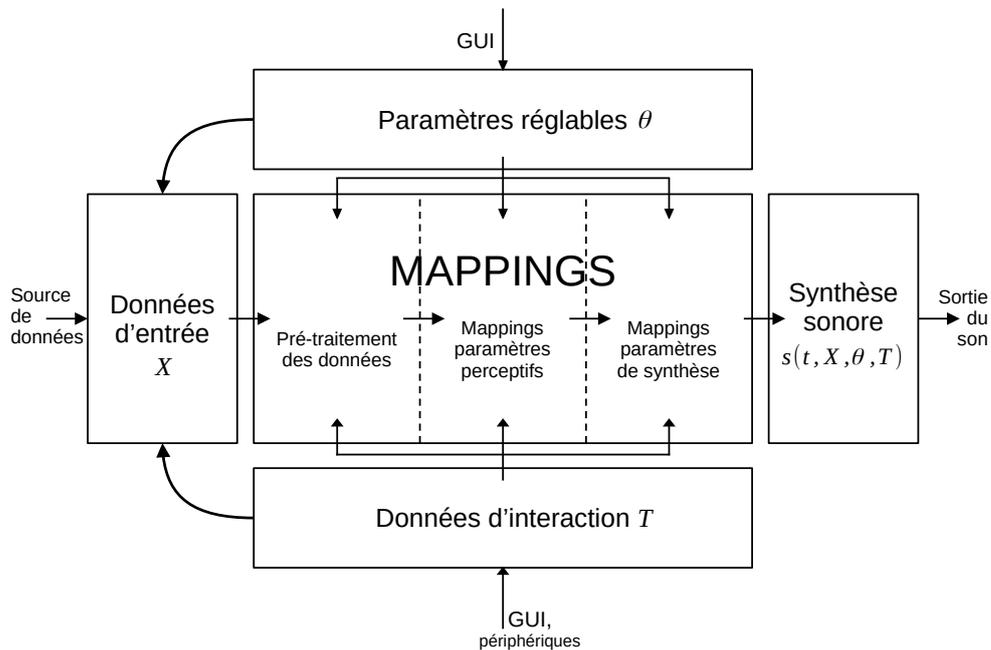


FIGURE 7.1 – Une vue d'ensemble du modèle Cellule de Sonification

Trois blocs d'entrée situés à gauche, au dessus et en dessous du schéma représentent les différentes sources de données qui peuvent servir à la paramétrisation du son : respectivement les données d'entrée à représenter, les paramètres de mapping explicitement réglables par l'utilisateur, et les données d'interaction selon les modalités offertes par l'interface de l'outil.

Le bloc central représente l'enchaînement des opérations de pré-traitements, puis de mappings, qui calculent les paramètres de synthèse sonore à partir des différentes sources de données.

Le bloc à droite du schéma représente l'interface de l'outil avec le rendu sonore. Il permet de décrire les fonctions qui synthétisent le signal sonore en fonction des paramètres calculés par le bloc central.

Dans les sous-sections qui suivent, nous donnons une description plus détaillée des différents éléments du schéma et de leur rôle vis-à-vis des aspects variables d'une sonification.

7.2.3 Rôle dans le design d'une sonification

Dans le processus de design de la sonification tel qu'il est étudié dans la section 6.1, le présent modèle s'inscrit à la suite de la sélection d'un type générique de représentation d'après les besoins, les données, la tâche et le contexte. Son rôle est d'offrir un cadre structuré et modulaire pour la description détaillée des différents aspects propres à un outil de sonification.

7.3 Bloc des données d'entrée

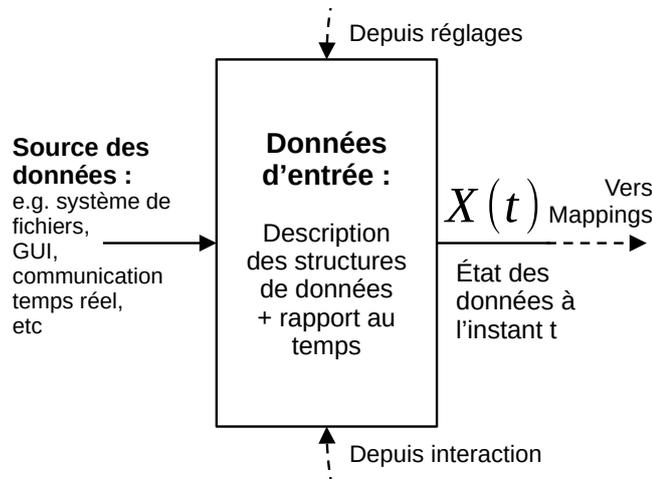


FIGURE 7.2 – Détail du bloc des données d'entrée

- Le bloc des données d'entrée, dont le détail est donné par la figure 7.2, permet :
- d'indiquer la source des données d'entrée (lecture d'un fichier, éléments d'une interface utilisateur, protocole de communication *OSC*, etc).
 - d'explicitier les structures de données auxquelles cette source donne accès. Il peut par exemple s'agir de séries temporelles dans lesquelles les lignes donnent le temps et les colonnes donnent les dimensions, ou l'inverse. Il est aussi possible de définir des objets plus complexes dotés de données membres pour leurs différentes caractéristiques.
 - d'offrir l'accès à ces données membres dans le bloc des mappings.
 - d'indiquer si les données sont statiques ou dynamiques. Si elles sont statiques, X est constant pour tout t : il existe un vecteur de valeurs X_0 tel que $\forall t \in \mathbb{R}^+, X(t) = X_0$. Si elles sont dynamiques, $X(t)$ dépend d'une période d'échantillonnage $\Delta t_d \in \mathbb{R}^+$ donnant le pas de temps entre chaque état successif des données : $\forall t \in \mathbb{R}^+, X(t) = X(\Delta t_d * \lfloor t/\Delta t_d \rfloor)$.

- de préciser, pour des ensembles de données entièrement connus a priori, des opérations d'analyse permettant d'en extraire certaines informations globales (entre autres : le minimum, le maximum, la moyenne). Ces opérations peuvent éventuellement dépendre de valeurs lues depuis les autres sources de données, décrites dans les sections 7.4 et 7.5 plus bas, pour l'interaction et les réglages. Un exemple de cet usage est visible dans le chapitre 8 de validation, pour le cas d'étude du sonogramme des données : à chaque nouvelle définition du point d'interaction (déclaré dans le bloc des interactions), on calcule à nouveau la plus grande distance (déclarée dans le bloc des données d'entrée) entre ce point et chaque point de l'espace de données représenté.

7.4 Bloc d'interaction utilisateur

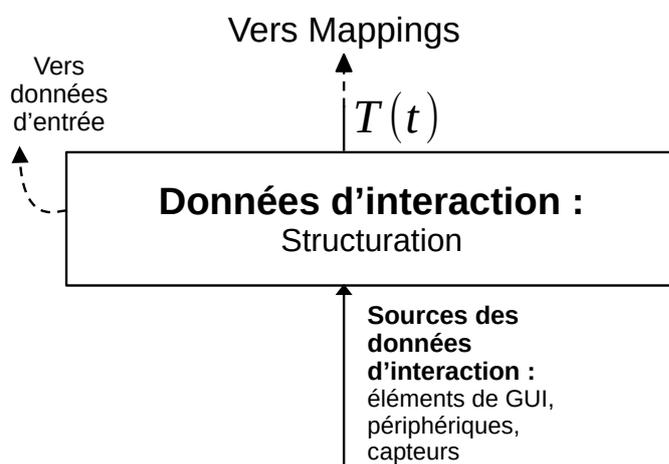


FIGURE 7.3 – Détail du bloc d'interaction utilisateur

Le bloc d'interaction utilisateur, dont le détail est donné par la figure 7.3, permet :

- d'indiquer, en entrée en dessous, les modalités selon lesquelles l'utilisateur peut interagir avec la sonification. Il peut s'agir d'une interface graphique, d'un contrôle à la souris ou au clavier, d'un périphérique dédié, etc.
- d'indiquer, en plusieurs sorties au dessus, les dimensions des données qui sont générées par l'interaction, et à quels niveaux du mapping au son ces contrôles se situent. Le contrôle offert peut ainsi porter sur le traitement préalable des données, sur l'exploration des échelles de mappings, ou sur la sélection de différents types de sons.

7.5 Bloc des paramètres réglables

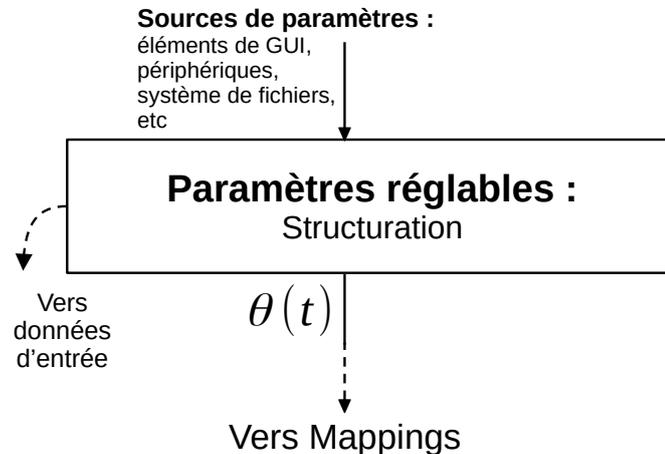


FIGURE 7.4 – Détail du bloc des paramètres réglables

Le bloc des paramètres réglables, dont le détail est donné par la figure 7.4, permet :

- d’indiquer, en entrée au dessus, les modalités de contrôle offertes à l’utilisateur pour gérer les paramètres de mapping réglables. Il peut s’agir de la sélection d’un fichier d’entrée dans lequel une configuration aurait été préalablement sauvegardée (pour les **HRTF** par exemple), ou de divers éléments d’une interface utilisateur (curseurs numériques, potentiomètres, boutons, etc).
- d’indiquer, en plusieurs sorties en dessous, les valeurs qui sont ainsi modifiables par l’utilisateur, et les niveaux du mapping au son auxquels elles se rapportent. On peut ainsi offrir la possibilité, entre autres, de sélectionner des zones d’intérêt dans l’espace des données, de placer des seuils sur les dimensions, de contrôler manuellement les échelles perceptives dans lesquelles les mappings se situent, ou de sélectionner différents sons synthétiques.

7.6 Bloc de la synthèse sonore

Le bloc de la synthèse sonore, dont le détail est donné par la figure 7.5, permet :

- de lister les différents complexes sonores à synthétiser et superposer dans le rendu. Le détail des fonctions de synthèse impliquées peut ensuite être

donné dans d'autres schémas plus spécifiques, selon la même notation donnée dans la section 7.7 pour les mappings.

- de lister, en entrées à gauche, les paramètres variables de ces fonctions pour un contrôle du rendu sonore au cours du temps.
- d'indiquer, en sortie à droite, la destination du son synthétisé (fichier enregistré, périphérique audio, bus pour la communication avec un autre module de traitement du son, etc).

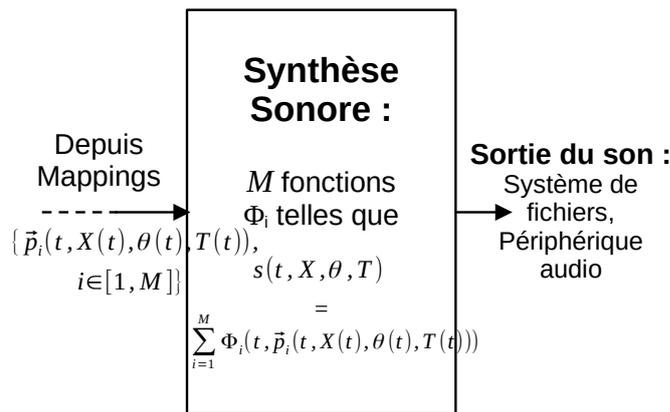


FIGURE 7.5 – Détail du bloc de synthèse sonore

Les définitions des complexes sonores dans ce bloc peuvent par ailleurs être factorisées de sorte à définir un nombre donné de complexes d'une seule et même façon à un indice près. Cette factorisation peut être utile pour des représentations où toutes les dimensions des données sont représentées à un même instant par le même type de son (par exemple un mix d'ondes sinusoïdes dont les hauteurs sont régies par les colonnes de la matrice de données). Une factorisation peut aussi être nécessaire dans des situations où les données correspondent à des points et l'on veut associer chaque point à son complexe sonore propre, synthétisé en fonction de ses propriétés.

La factorisation est représentée dans le bloc de la synthèse sonore par un cadre entourant le complexe à factoriser, accompagné d'un texte en dessous décrivant l'indexation. Par exemple pour décrire le type de complexe sonore à construire pour un ensemble de M points numérotés de 0 à $M - 1$, l'indexation s'écrit : $\forall i \in [0; M - 1]$.

Des exemples de cette écriture sont donnés dans le chapitre 8 dans le cadre d'une retranscription des cas d'études de *FluoSim-sonif* (voir le chapitre 3) et du sonogramme des données (voir [96]) sous le format du modèle.

7.7 Bloc des mappings

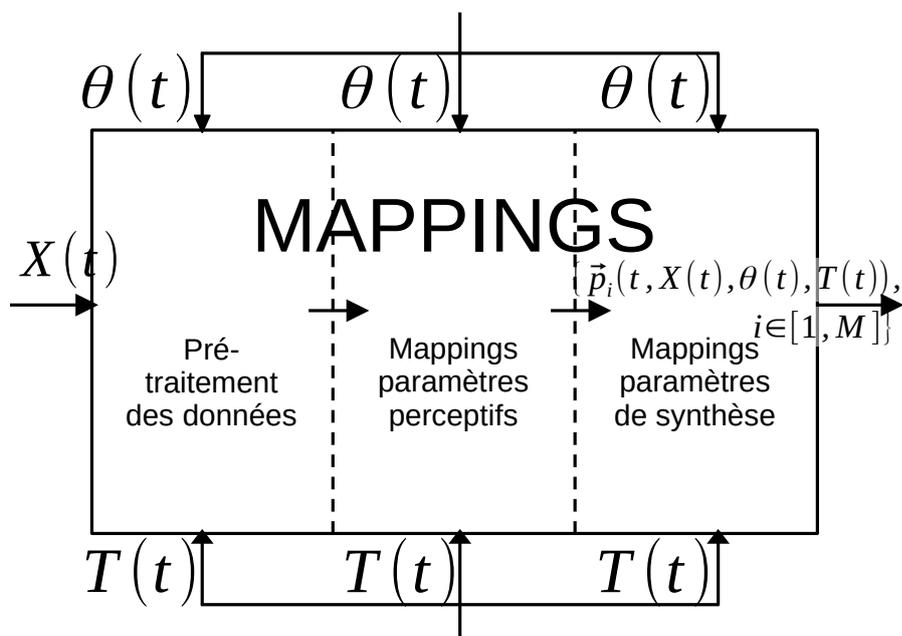


FIGURE 7.6 – Détail du bloc des mappings

Le bloc des mappings est représenté par la figure 7.6. Il prend comme entrées, à gauche, en haut et en bas, les flux de données provenant respectivement de la source de données à représenter, des modalités de réglages des paramètres et des outils d'interaction. À l'intérieur du bloc, on détaille les opérations qui produisent les paramètres à communiquer au module de synthèse sonore, en sortie à droite.

7.7.1 Trois couches de mappings

Les trois couches successives de mappings, séparées dans la figure 7.6 par des lignes verticales en pointillés, sont inspirées du modèle de mappings entre espaces perceptifs proposé par Arfib et al. [7], initialement prévu pour le design d'instruments de musique dirigés par des données de gestuelle.

À l'issue de quelques adaptations lexicales généralisant son application à n'importe quel type de données d'entrée, ce découpage permet d'offrir un cadre clair pour structurer les étapes de transformation du flux de données :

- 1^{ère} couche : pré-traitement des valeurs d'entrée afin d'en extraire des informations et relations sous-jacentes.
- 2^{ème} couche : quantification des paramètres perceptifs associés aux propriétés ainsi extraites.
- 3^{ème} couche : transposition, là où elle est nécessaire, des paramètres perceptifs vers des échelles et unités compatibles avec les techniques de synthèse sonore mises en place.

La couche du milieu pourrait être considérée comme facultative, puisqu'il peut suffire de directement associer des propriétés à des paramètres de synthèse sonore. Elle est cependant recommandée afin d'ajouter un niveau d'abstraction explicite aux choix d'échelles perceptives, une telle modularité permettant ensuite d'anticiper d'éventuels besoins de compatibilité pour le passage à la synthèse sonore.

Ainsi, par exemple, on pourrait avoir décidé au cours du prototypage d'une sonification d'associer une grandeur à des hauteurs tonales choisies dans une échelle diatonique. Ce mapping serait décrit dans la couche du milieu, par exemple en sélectionnant un nombre entier identifiant la note en fonction de seuils sur la grandeur d'entrée. La troisième couche servirait alors d'interface avec un module de synthèse sonore qui pourrait être choisi de différentes façons. Par exemple, pour un message **MIDI**, le paramètre de hauteur serait un identifiant de note entre 0 et 127, tandis que pour une fonction de type oscillateur le paramètre serait une fréquence en Hz.

En testant le modèle sur des cas d'études dans le chapitre 8, on remarquera que le positionnement de certaines opérations dans l'une ou l'autre couche de mapping dépend parfois plus des préférences subjectives du concepteur que de règles universelles sur la hiérarchisation des transformations. Le but de ce découpage est en tout cas d'offrir une aide à la conception qui amène, sans imposer de contraintes, à prévoir la modularité des étapes du mapping et à expliciter les choix de métaphores et de paramètres perceptifs.

7.7.2 Description graphique des mappings

Le remplissage du bloc des mappings consiste à décrire les enchaînements d'opérations à partir de modules qui en constituent les étapes successives, connectés entre eux par des flèches indiquant le flot d'informations.

Ces modules prennent, en entrée à leur gauche, un ou plusieurs paramètre(s) provenant de X , de θ , de T ou de modules d'opérations précédents, et produisent une valeur de sortie à leur droite.

Les intitulés de ces modules peuvent décrire des opérations plus ou moins complexes selon les besoins en précision et l'espace qui peut leur être accordé dans la représentation graphique. Il est toujours possible de donner des noms abrégés à certaines fonctions pour en décrire le principe en dehors du schéma, sous forme de texte ou dans un autre graphique qui en donne le détail. Dans ces intitulés, les variables acquises à gauche sont désignées par des noms constitués de la lettre x suivie de la position de la flèche dans l'ordre d'entrée du bloc, de haut en bas. Ainsi, la flèche la plus en haut correspond à une variable x_0 dans le bloc, puis la deuxième à une variable x_1 , etc. Si un bloc ne comporte qu'un seul port d'entrée, la variable correspondante s'appelle simplement x .

Les modules d'opérations du bloc mapping peuvent contenir n'importe quelle opération, tant que leurs valeurs de sortie sont d'un type numérique. Cela inclut les variables logiques, qu'on peut considérer comme numériques en tant que membres de l'ensemble $\{0, 1\}$, ainsi que les vecteurs de deux valeurs ou plus, dénotés en tant que variables par des flèches au dessus de leurs noms.

7.7.3 Fonctions récurrentes

Un certain nombre de fonctionnalités, notamment de changement d'échelle, sont inévitablement employées en sonification. Nous proposons dans cette sous-section de formaliser leurs syntaxes et noms afin de les rendre facilement réutilisables et reconnaissables sans leur consacrer trop d'espace dans un schéma.

Mapping Affine $M_A(x, [x_{min}; x_{max}], [y_{min}; y_{max}])$

La fonction de mapping affine donne la projection de la valeur d'entrée x sur une droite définie par la fonction :

$$M_A(x, [x_{min}, x_{max}], [y_{min}, y_{max}]) = y_{min} + \frac{(x - x_{min})(y_{max} - y_{min})}{(x_{max} - x_{min})} \quad (7.2)$$

x_{min} et x_{max} sont les bornes de l'intervalle des valeurs d'entrée, telles que $x_{min} < x_{max}$.

y_{min} et y_{max} sont les valeurs de sortie associées respectivement aux bornes x_{min} et x_{max} , mais elles ne constituent pas nécessairement les bornes respectivement minimale et maximale des valeurs de sortie. En effet, si le mapping est de polarité inverse, $y_{min} > y_{max}$.

Mapping Affine Discret $M_{AD}(x, [x_{min}; x_{max}], \{y_{min}..y_{max}\})$

La fonction de mapping affine discret est une fonction de mapping affine pour laquelle l'ensemble des valeurs de destination est constitué de valeurs numériques ordinales. C'est-à-dire que cet ensemble de valeurs possède un ordre et correspond à un ensemble de catégories individuelles. Dans le cas du format **MIDI** par exemple, il s'agit alors d'identifiants de notes, qu'on choisit d'inscrire dans le schéma sous les noms donnés par la notation française, suivis du numéro de l'octave à laquelle ils appartiennent ($Do_1, Re_1, Mi_1, Fa_1, Sol_1, La_1, Si_1$, etc), mais qui sont concrètement gérés comme des nombres entiers entre 0 et 127.

Cette opération commence par calculer le mapping $M_A(x, [x_{min}; x_{max}], [y_{min}; y_{max}])$ donné par l'expression 7.2 précédente, puis renvoie le membre de l'ensemble de destination dont la valeur est la plus proche. Par exemple si l'ensemble de destination est un intervalle entier, il s'agit de l'arrondi.

7.7.4 Expressions conditionnelles

Les blocs d'opérations peuvent calculer et renvoyer des valeurs booléennes. Les conditions qui dictent leurs valeurs de sortie sont écrites selon la syntaxe :

- Pour comparer des valeurs : $\neq, ==, \leq, \geq, >, <$.
- Pour combiner des conditions : $||$ pour « ou » et $\&\&$ pour « et ».
- Des parenthèses peuvent servir à expliciter la priorité des opérations.
- La fin d'une condition est marquée par un « ? ».

Ainsi, par exemple, « $x_0 \neq x_1 \&\& x_0 \geq 0$ » est la vérification que deux variables x_0 et x_1 sont différentes l'une de l'autre et que x_0 est supérieure ou égale à zéro.

Une expression écrite selon la syntaxe décrite ci-dessus peut aussi être utilisée pour conditionner le comportement d'une fonction. Ainsi, un bloc peut contenir une condition suivie de deux valeurs de retour alternatives selon que la proposition est vraie ou fausse, séparées par « : ».

Ainsi l'expression « $x_0 \geq 0?x_1:x_2$ » signifie que si x_0 est supérieure ou égale à zéro, le bloc renvoie la valeur x_1 , et sinon le bloc renvoie la valeur x_2 .

7.8 Temporalité

De la même façon que dans les modélisations de la sonification précédemment décrites par Hermann et al. [101] et Rohrhuber [185], l'expression de la sonification donnée par notre modèle est régie par l'écoulement du temps pour le rendu sonore, celui-ci constituant une constante inévitable du concept même de la sonification. On s'abstrait ainsi des différents rapports possibles des données d'entrée au temps, qu'elles soient statiques, dynamiques selon un pas, ou générées en temps réel.

7.8.1 Synchronicité

L'ordonnement des opérations est entièrement régi par l'écoulement du temps pour la synthèse sonore. À chaque pas de temps, le bloc de synthèse « consulte » les valeurs des paramètres qui serviront à rendre le prochain échantillon de son. La consultation est transmise aux blocs d'opérations précédents, qui consultent à leur tour leurs prédécesseurs, jusqu'à arriver à l'une des sources de données X , θ , T , ou une variable mémoire. Les opérations sont alors effectuées et les résultats sont renvoyés en cascade jusqu'à ce que les complexes sonores aient tous leurs paramètres.

On suppose que toutes les opérations se déroulent exactement à un même instant et que, dans une implémentation logicielle du modèle, cette synchronicité sera correctement simulée. Le but est qu'au moment d'effectuer l'opération d'un bloc, toutes ses entrées doivent avoir été calculées pour la même itération temporelle.

7.8.2 Paramètre temporel

Dans le modèle, le temps correspond à un paramètre lisible depuis tous les blocs. Sous sa forme la plus générique, il s'agit d'un nombre réel donnant la quantité de temps écoulé depuis l'initialisation du programme, par exemple en secondes.

Comme tous les exemples étudiés par la suite dans cette thèse consistent à synthétiser un signal sonore échantillon par échantillon, nous considérons pour simplifier que le paramètre temporel dans tout ce qui suit est en fait une valeur entière représentant le nombre de périodes d'échantillonnage du son écoulées depuis l'initialisation du programme¹. Dans les représentations, on suppose que les opérations de mappings sont toujours effectuées suffisamment vite pour que cette valeur puisse être incrémentée de 1 à chaque itération des calculs de paramètres.

1. Pour un autre exemple, si la sonification produisait un enregistrement **MIDI**, ce temps serait donné en ticks à la place.

La période d'échantillonnage est un temps qui correspond à l'inverse du taux d'échantillonnage du son Fe , lui aussi global dans le modèle. Le temps total, en secondes, écoulé depuis l'initialisation du programme est ainsi donné par t/Fe . Pour des opérations où il est nécessaire de convertir un nombre de périodes d'échantillonnage en temps on effectue une division par Fe , et pour la conversion d'un temps en un nombre de périodes d'échantillonnage on effectue une multiplication par Fe .

Dans les blocs d'opérations qui génèrent un signal, notamment utilisés pour détailler le contenu des complexes sonores, les appels à des fonctions qui dépendent du temps sont notées $Fonction(paramètres)(t)$, de sorte à séparer explicitement le fait d'appeler un générateur de signal en fonction des *paramètres* d'une part, et le fait de récupérer la valeur de ce signal qui correspond à l'instant t d'autre part.

7.8.3 Variables mémoires

Les variables mémoires sont des champs de valeurs qui peuvent être ajoutés au modèle afin de mémoriser l'état d'un paramètre à un instant donné pour des consultations ultérieures. Elles peuvent ainsi être utilisées comme des paramètres globaux dans la chaîne de traitement. Par exemple, et le plus souvent, une variable mémoire sert à comparer un paramètre à une valeur qu'il avait prise un ou plusieurs pas de temps en arrière.

On indique l'affectation initiale d'une variable mémoire dans son bloc de source : un bloc sans prédécesseur ni successeur, reconnaissable dans le schéma par son contour plus épais.

Ensuite, au cours du déroulement des calculs dans le modèle, des affectations notées $var \leftarrow expression$ peuvent être provoquées par des conditions. Dans la représentation graphique des opérations, ce type de liaison particulier est indiqué par une double flèche entre la condition et l'affectation. Comme le bloc d'initialisation, le bloc d'affectation se distingue graphiquement par son contour plus épais.

On considère que les opérations d'affectation des variables sont prioritaires sur les utilisations qui en sont faites au cours du traitement d'un instant donné. Tout usage d'une variable mémoire est considéré comme ayant lieu après son éventuel changement de valeur, à l'exception de sa condition d'affectation si celle-ci nécessite de consulter son dernier état enregistré.

7.9 Conclusion

Dans ce chapitre, nous avons décrit la progression d'après l'existant qui nous a amenés au modèle de Cellule de Sonification, ainsi que les différents éléments qui constituent sa représentation graphique, une forme qui est à notre sens indispensable pour sa lisibilité et potentiellement celle sous laquelle il sera présenté dans le cadre d'un outil de conception interactive.

Afin de valider sa compatibilité avec différentes approches pour la sonification, nous montrons dans le prochain chapitre 8 comment les cas d'études de la présente thèse, ainsi que quelques autres exemples tirés de la littérature, peuvent être retranscrits sous le format proposé.

Chapitre 8

Validation d'après cas d'études

Afin de valider au moins en partie l'exhaustivité du modèle de Cellule de Sonification, nous proposons dans ce chapitre de reprendre sous le format du modèle un certain nombre de cas particuliers de sonification, caractérisés par diverses spécificités.

Dans une première section, nous revenons donc sur les solutions de sonification proposées pour les cas d'études de cette thèse, puis dans la suivante nous traitons de deux cas d'études supplémentaires tirés de l'existant afin de montrer la souplesse du modèle quant à la représentation de techniques de sonification au-delà de la simple **PMSon**.

Pour tous ces exemples, la description des solutions de sonification sera donné à la fois sous une forme textuelle précise structurée selon les différents éléments du modèle, et sous la forme graphique détaillée dans le chapitre 7.

8.1 Cas d'études de la thèse

Nos réflexions sur la description d'un modèle de la sonification ont eu lieu dans le cadre de cas d'études réalisés en collaboration avec divers acteurs de la recherche. Ces cas d'études, décrits dans les chapitres 2 et 3, sont représentatifs d'un certain nombre de contraintes et spécificités dont nous vérifions dans cette section la compatibilité avec notre modèle de représentation.

Comme il n'a encore abouti à aucune solution concrète, le sujet d'apprentissage et restitution du son du vent présenté dans le chapitre 4 de la présente thèse sera pour le moment exclu de cette étude de la validité du modèle.

8.1.1 Supervision du procédé de fabrication additive wire-arc

Au cours de notre étude pour la supervision sonifiée du procédé wire-arc, par ailleurs rapportée en détail dans le chapitre 2 de la présente thèse, nous avons construit quatre versions successives d'une représentation auditive permettant de suivre l'évolution de cinq mesures physiques caractérisant le bon déroulement d'une impression 3D.

Ces quatre versions étant assez similaires, on choisit ici de ne reproduire que *Mixte1*, qui est la plus représentative des choix effectués. Un intérêt de cette représentation en particulier pour la validation du modèle est qu'on y fait coexister des complexes sonores tirés de sons pré-enregistrés d'une part et de sons synthétisés d'autre part, et qu'on cherche ici à tous les décrire selon un format commun. Dans son état actuel, la solution de sonification dépend uniquement des données d'entrée et ne propose ni interaction, ni paramètres réglables à l'utilisateur.

Ce cas d'étude est l'occasion de montrer l'application du modèle à une sonification assez classique. En effet, comme celle-ci est destinée à la supervision d'un procédé, le rapport temporel entre l'acquisition des données et le rendu sonore est direct. Cependant, on remarque aussi quelques particularités vis-à-vis, notamment, des seuils d'activation : la paramétrisation du flux *Mélodie* ne se fait que si l'un des paramètres récemment calculés est suffisamment différent du dernier état enregistré, et l'alarme *Friture* ne peut se réactiver que 10 secondes au minimum après son dernier déclenchement.

La figure 8.1 montre l'application du format de la Cellule de Sonification à la représentation *Mixte1*.

Vue d'ensemble Dans *Mixte1*, les deux critères *HC* et *LC* donnant la géométrie du dépôt de matière (ou cordon) sont représentés par un même complexe nommé *Mélodie*, le critère de hauteur de la pièce *HP* est représenté par un complexe nommé *Bourdon* donnant sa comparaison avec la valeur attendue HP_0 , le critère de température du cordon *TC* est représenté par des sons d'*Ébullition* ou de *Craquèlement* selon qu'il est trop haut ou trop bas, et le critère de température de la pièce *TP* provoque un son de grésillement, nommé *Friture*, en cas de dépassement du seuil toléré.

Pour les besoins de cette sonification, quatre variables mémoire ont été ajoutées aux données d'entrée. L'instant t_{seuil} sert à vérifier que la dernière activation de l'alarme pour *PT* a eu lieu au moins 10 secondes plus tôt. Les variables p_0 , p_1 et p_2 servent à garder en mémoire le dernier état accepté des trois paramètres de *Hauteur*, *Gain* et *Durée* pour le complexe *Mélodie*, afin de vérifier que la différence est suffisamment marquée pour une mise à jour.

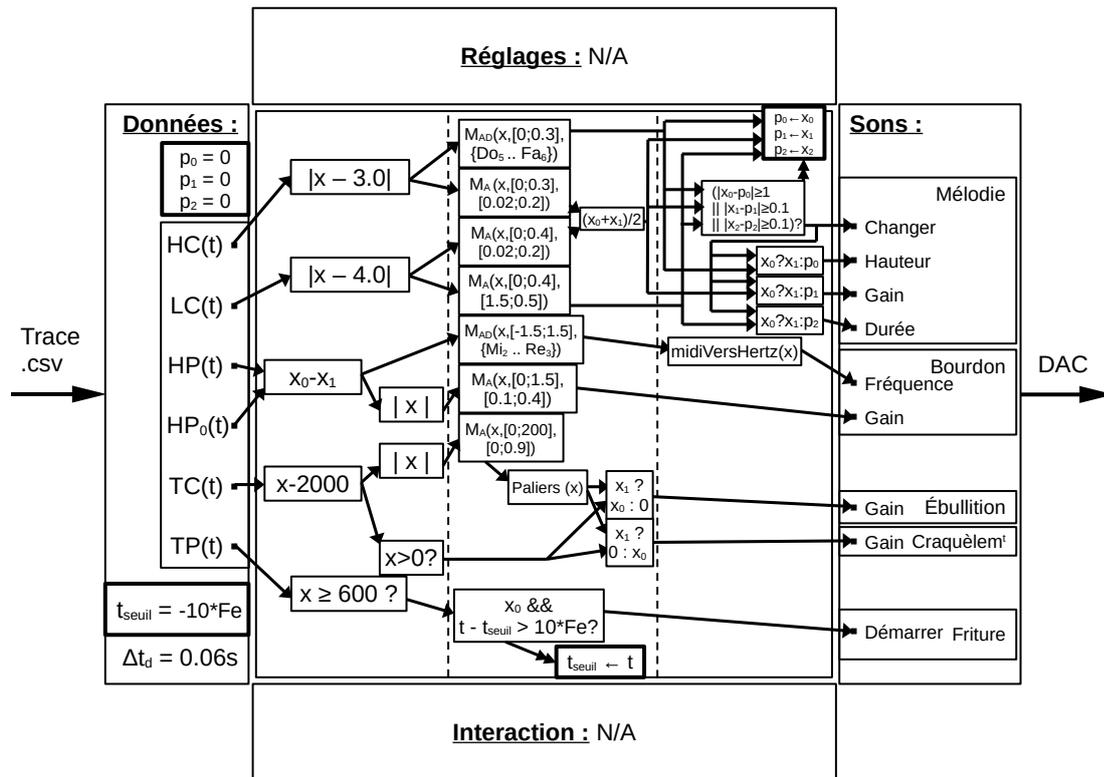


FIGURE 8.1 – La sonification *Mixte1* du chapitre 2 décrite selon le format du modèle

Données d'entrée Pour ce cas d'étude, les données sont acquises depuis un fichier **CSV** statique contenant une trace de fabrication pré-enregistrée. Le pas de temps du déroulement des données Δt_d est de 0.06s. Les dimensions des données HC , LC , HP , HP_0 , TC et TP sont lues dans cet ordre d'après les colonnes 4 à 9 du fichier (les trois premières colonnes étant occupées par des méta-données sur l'avancement de la fabrication) et structurées telles-quelles en autant de ports de données.

À l'initialisation, l'instant t_{seuil} prend la valeur $-10 * Fe$ afin que sa vérification ne soit pas bloquante si un déclenchement devait être nécessaire dans les 10 premières secondes de fonctionnement. Les variables p_0 , p_1 et p_2 sont simplement initialisées à 0, puisqu'on admet que les premières valeurs prises par les paramètres de *Mélodie* seront forcément assez différentes de 0 pour être prises en compte.

Mappings Le bloc des mappings, au milieu de la figure 8.1, montre ici assez nettement la progression des paramètres selon les trois couches de traitement,

marquées par des traits pointillés. Les transformations effectuées couche après couche sont décrites dans les trois paragraphes qui suivent.

Pré-traitements Dans la zone à gauche du bloc des mappings, on décrit les pré-traitements à appliquer aux données pour les besoins de la représentation. Ici il s'agit de calculer les écarts par rapport aux objectifs pour les différents critères de fabrication. De haut en bas :

- $HC(t) : |x - 3.0|$ donne l'écart absolu par rapport à 3.0 mm.
- $LC(t) : |x - 4.0|$ donne l'écart absolu par rapport à 4.0 mm.
- $HP(t) : x_0 - x_1$ donne l'écart relatif par rapport à $HP_0(t)$ et $x_0 - x_1 \rightarrow |x|$ donne l'écart absolu.
- $TC(t) : x - 2000 \rightarrow |x|$ donne l'écart absolu par rapport à 2000°C et $x - 2000 \rightarrow x > 0?$ donne le signe de la différence sous la forme d'une valeur booléenne, telle que Vrai signifie positif et Faux signifie négatif.
- $TP(t) : x \geq 600?$ donne une valeur booléenne indiquant le dépassement du seuil 600°C .

Paramètres perceptifs Dans la zone centrale du bloc des mappings, on décrit ensuite la transformation des données pré-traitées en paramètres perceptifs. Ici, les paramètres sont assez basiques et traitent de la sonie, de la hauteur tonale et du rythme des différents éléments du son. Une application notable de cette couche pour ce cas d'étude est la sélection de l'un ou l'autre complexe sonore pour l'alerte de TC , selon que la température est trop basse ou trop haute. De haut en bas les mappings aux paramètres perceptifs sont :

- l'écart absolu pour $HC(t)$ donne un identifiant de note selon le mapping affine discret $M_{AD}(x, [0; 0.3], \{Do_5 \dots Fa_6\})$
- l'écart absolu pour $HC(t)$ donne un facteur de gain entre 0.02 et 0.2 selon le mapping affine $M_A(x, [0; 0.3], [0.02; 0.2])$
- l'écart absolu pour $LC(t)$ donne un facteur de gain entre 0.02 et 0.2 selon le mapping affine $M_A(x, [0; 0.4], [0.02; 0.2])$
- la moyenne des deux facteurs de gain est calculée selon une opération $(x_0 + x_1)/2$
- l'écart absolu $LC(t)$ donne une durée en secondes selon le mapping affine (dont on remarque au passage la polarité inverse) $M_A(x, [0; 0.4], [1.5; 0.5])$
- l'écart relatif de $HP(t)$ donne un identifiant de note selon le mapping affine discret $M_{AD}(x, [-1.5; 1.5], \{Mi_2 \dots Re_3\})$

- l'écart absolu de $HP(t)$ donne un facteur de gain entre 0.1 et 0.4 selon le mapping affine $M_A(x, [0; 1.5], [0.1; 0.4])$
- l'écart absolu de $TC(t)$ donne un facteur de gain entre 0 et 0.9 selon le mapping affine $M_A(x, [0; 200], [0; 0.9])$, qui est ensuite passé à la fonction *Paliers* (décrite plus bas) pour prendre l'une de ses trois valeurs caractéristiques
- les deux blocs $x_1?x_0:0$ et $x_1?0:x_0$ effectuent ensuite la sélection du complexe sonore à activer ou désactiver entre *Ébullition* et *Craquèlement* à partir du signe de l'écart pour $TC(t)$ préalablement calculé. Si l'écart est positif, *Ébullition* prend le paramètre de gain et *Craquèlement* prend 0, sinon c'est l'inverse.
- la condition de dépassement de $TP(t)$ est combinée à une vérification $t - t_{seuil} > 10 * Fe?$ qui indique si le temps écoulé depuis le dernier déclenchement est de 10 secondes ou plus. À ce moment, si la condition est vraie, on affecte la valeur actuelle du paramètre de temps t à la variable t_{seuil} .

La fonction $Paliers(x)$ renvoie l'une ou l'autre valeur en fonction de l'appartenance de x à l'un ou l'autre intervalle, de sorte que pour une valeur d'entrée $x \in [0; 0.9]$:

$$Paliers(x) = \begin{cases} 0 & \text{si } x \in [0; 0.4] \\ 0.2 & \text{si } x \in]0.4; 0.6] \\ 0.5 & \text{si } x \in]0.6; 0.9] \end{cases} \quad (8.1)$$

Paramètres de synthèse Dans la zone à droite du bloc des mappings, enfin, on effectue les dernières transformations et vérifications nécessaires pour la bonne communication des paramètres aux fonctions de synthèse. Ici il s'agit de respecter certaines contraintes de formats et de seuils, telles qu'elles sont détaillées ci-dessous :

- les paramètres de hauteur, gain et durée, destinés à *Mélodie* passent par une vérification de leur différence par rapport au dernier état accepté pour le changement de paramètre : $(|x_0 - p_0| \geq 1 \ || \ |x_1 - p_1| \geq 0.1 \ || \ |x_2 - p_2| \geq 0.1)?$ renvoie Vrai si au moins l'un des paramètres est suffisamment différent du précédent (de 1 pour la hauteur et de 0.1 pour le gain ou la durée).
- si cette condition est vraie, on affecte les nouvelles valeurs des paramètres aux états sauvegardés en passant par le bloc d'affectation en haut à droite : $p_0 \leftarrow x_0; p_1 \leftarrow x_1; p_2 \leftarrow x_2$

- puis, selon que la condition est vraie ou fausse, les paramètres communiqués à *Méلودie* par les trois blocs de la forme $x_0 ? x_1 : p_i$ (i entre 0 et 2) sont soit les dernières valeurs calculées, soit les valeurs sauvegardées dans les variables.
- la valeur de sortie de la condition sert aussi pour le paramètre *Changer* dans *Méلودie*, en tant qu'indicateur d'une mise à jour des paramètres qui impose de redémarrer la séquence de notes
- le paramètre de hauteur calculé à partir de $HP(t)$ et $HP_0(t)$ passe par une fonction *midiVersHertz*, décrite ci-dessous, afin de prendre la forme du nombre de périodes par seconde, en Hertz, qui correspond à son identifiant de note.

La fonction *midiVersHertz*(x) donne la fréquence fondamentale en Hertz associée à un identifiant de hauteur $x \in [0; 127]$ selon le format **MIDI**. L'opération typiquement effectuée pour obtenir ce nombre se base sur l'association du 69^{ème} identifiant de hauteur à la note La_4 de fréquence fondamentale 440 Hz :

$$midiVersHertz(x) = 440 * 2^{(x-69)/12} \quad (8.2)$$

Complexes sonores Les quatre prochains paragraphes donnent le détail de la synthèse sonore pour les complexes *Méلودie*, *Bourdon*, *Ébullition*, *Craquèlement* et *Friture* mis en œuvre dans la représentation.

Méلودie est un motif de trois notes (tonique-tierce-quinte sur une gamme majeure variable) jouées par le synthétiseur par défaut de *SuperCollider*¹. Les paramètres de ce complexe sont la hauteur de la note tonique, la durée de chaque note et un facteur d'amplitude donnant une sonie plus ou moins forte. Le changement de paramètres ne fait effet que si la variable booléenne supplémentaire *Changer* est vraie. Dans ces cas, la synthèse du motif redémarre à sa première note.

Dans notre implémentation de cette sonification sous *SuperCollider*, ce complexe consiste en une structure de type *pattern*, dont on se sert comme une boîte noire à laquelle les paramètres sont communiqués (à l'aide de *PBind*), et qui se charge elle-même de gérer le déroulement des notes au cours du temps. Si l'on reconstitue cependant le comportement du *pattern* sous le même format modulaire que le reste du modèle, on obtient le schéma donné sur la figure 8.2 :

- Les paramètres de *Hauteur*, *Gain* et *Durée* sont mis à jour si la variable *Changer* est vraie, et laissés à leur dernière valeur sinon.

1. Le code de ce synthétiseur peut être retrouvé à la ligne 166 du document `/SCClassLibrary/Common/Collections/Event.sc` à partir du répertoire d'installation de *SuperCollider*.

8. Validation d'après cas d'études

- Le repère de temps $t_{changer}$ permet de mémoriser le dernier instant où la variable *Changer* a été vraie.
- La note (tonique, tierce ou quinte) à jouer est sélectionnée en fonction du positionnement de l'instant actuel ($t - t_{changer}$) dans la première, deuxième, ou troisième partie de la séquence de notes, chacune de longueur *Durée*, puis convertie en une fréquence fondamentale pour le synthétiseur.
- Les paramètres de fréquence et de gain, ainsi que le temps propre de la note en cours (t_{synth}), sont communiqués à la fonction de synthèse sonore *Default*, dont le détail est donné dans la figure 8.3.

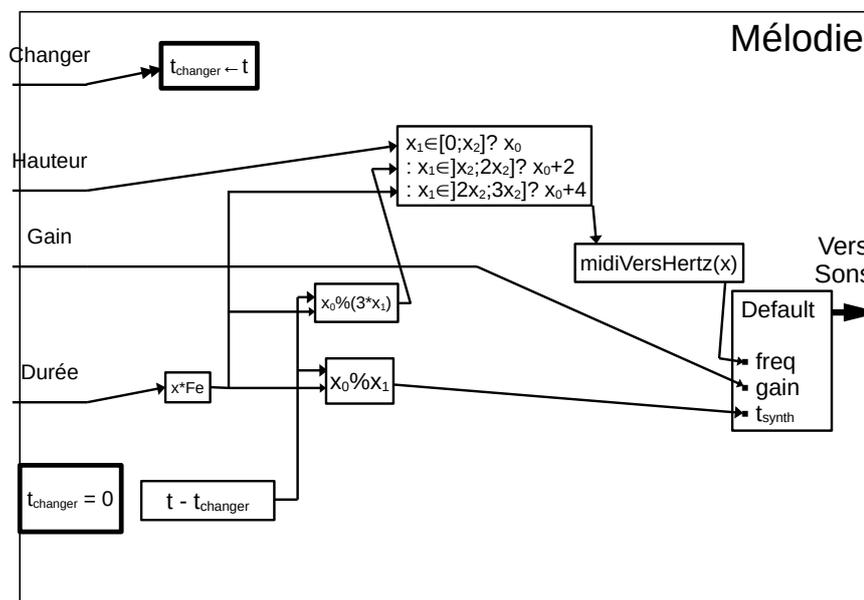


FIGURE 8.2 – Détail de la fonction de synthèse donnant le complexe sonore *Mélodie*

Les blocs d'opérations utilisés dans le schéma pour décrire *Default* portent les noms des *UGen* initialement utilisés sous *SuperCollider* pour définir ce synthétiseur.

VarSaw génère une onde en dents de scie dont la fréquence est donnée par son premier argument. *XLine* crée une courbe exponentielle qui évolue de la valeur de son premier argument à la valeur de son deuxième argument, pour une durée donnée en secondes par le troisième argument. *Rand* renvoie une valeur aléatoire entre ses deux arguments. *LPF* applique un filtre passe-bas sur le signal donné en

premier argument, selon la fréquence de seuil donnée par le deuxième argument. *LinEn* définit une enveloppe d'amplitude selon des arguments donnant ses temps successifs d'attaque, de dégradation, de soutien et de relâchement.

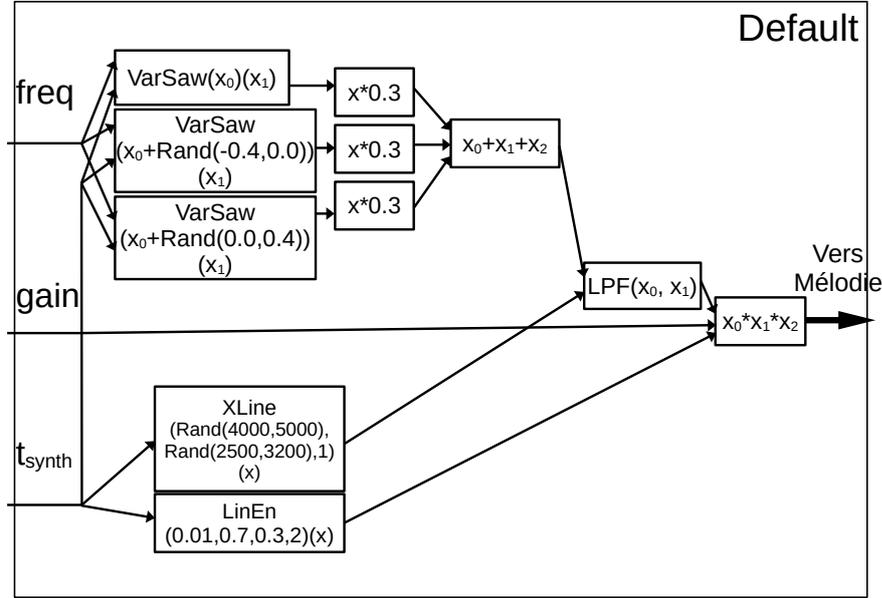


FIGURE 8.3 – Détail du synthétiseur *Default* impliqué dans le complexe *Mélodie*

Bourdon est un son continu constitué d'une onde en dents de scie filtrée selon deux filtres passe-bande répartis sur les deux canaux stéréo. Pour $FPB(s, f, q)$ la fonction appliquant au signal s un filtre passe-bande de ratio q autour de la fréquence f , et pour $Scie(f)(t)$ la fonction du signal en dents de scie de fréquence f à l'instant t (comme celle donnée par l'équation 1.2 du chapitre 1), le complexe *Bourdon* à l'instant t pour une fréquence F et un gain G est donné par l'expression suivante, aussi représentée sous le format du modèle dans la figure 8.4 :

$$Bourdon(F, G)(t) = \begin{cases} FPB(G * Scie(F)(t), F, 100) \text{ à gauche} \\ FPB(G * Scie(F)(t), 2 * F, 100) \text{ à droite} \end{cases} \quad (8.3)$$

Sur la figure, la fonction *Stéréo* effectue la concaténation des deux signaux d'entrée dans un vecteur de deux valeurs, la première donnant le canal de gauche et la deuxième le canal de droite.

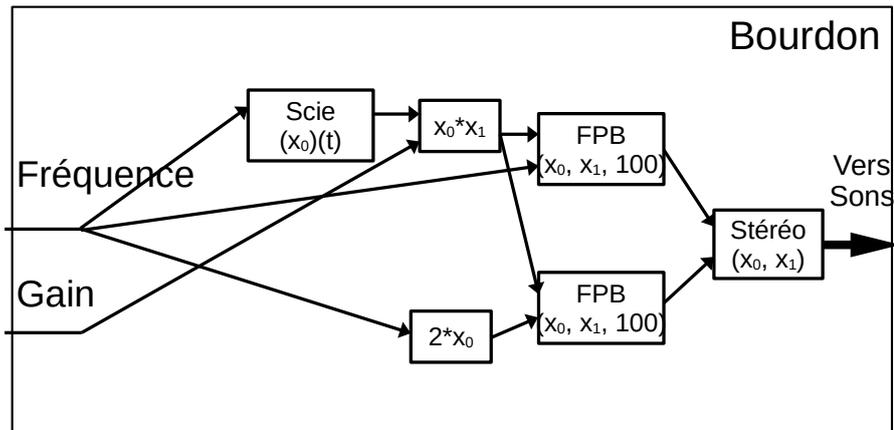


FIGURE 8.4 – Détail du complexe sonore *Bourdon*

Ébullition/Craquèlement sont deux complexes sonores alternatifs pouvant se produire en cas d'anomalie, respectivement haute ou basse, de la température du cordon. Ces complexes consistent en des enregistrements d'un son d'eau en ébullition ou du craquèlement de glaçons, chargés en mémoire et restitués en boucle à un volume plus ou moins élevé selon leurs paramètres de gains respectifs. Dans la chaîne de traitement, une vérification de la direction de l'anomalie $TC(t) - 2000 > 0?$ indique lequel doit recevoir le paramètre de gain et lequel doit être rendu silencieux.

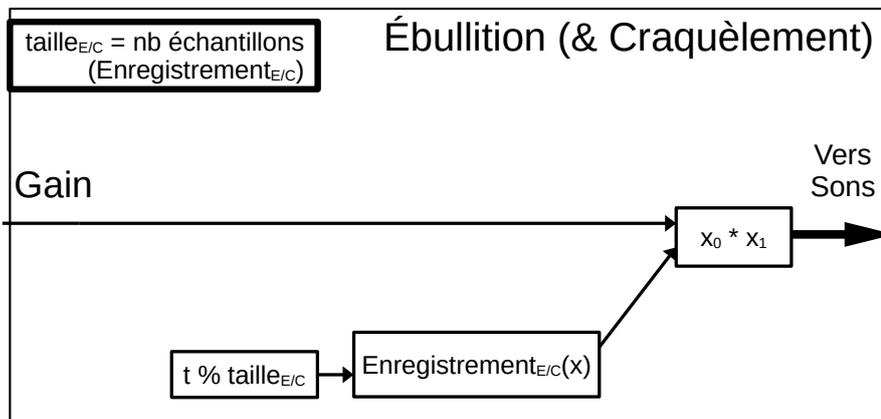


FIGURE 8.5 – Détail de la construction des complexes sonores *Ébullition* et *Craquèlement*

La figure 8.5 donne une représentation graphique de la construction des complexes sonores *Ébullition* et *Craquèlement*. Assez simplement, la synthèse sonore consiste à lire l'échantillon du son point par point dans un tableau de valeurs, nommé $Enregistrement_{E/C}$ (respectivement E pour *Ébullition* et C pour *Craquèlement*) sur le schéma. L'indice de l'échantillon d' $Enregistrement_{E/C}$ à lire à chaque instant est calculé comme le nombre t de pas de temps effectués depuis l'initialisation, modulo le nombre $taille_{E/C}$ d'échantillons qui constituent $Enregistrement_{E/C}$. Ainsi, la lecture du son effectue automatiquement une boucle quand elle atteint la fin de l'enregistrement.

Friture est un enregistrement d'un son d'eau grésillant sur une surface brûlante. La lecture de ce son est déclenchée par la valeur booléenne notée $Démarrer$ dans le schéma, qui est vraie quand le seuil de température de la pièce est dépassé, sous réserve que son dernier déclenchement ait eu lieu plus de 10 secondes plus tôt.

La figure 8.6 donne une représentation graphique de la construction du complexe sonore *Friture*. Comme pour *Ébullition* et *Craquèlement*, la synthèse de ce complexe consiste à lire un son échantillon par échantillon depuis un tableau, noté $Enregistrement_F$ sur le schéma. Ici, la lecture ne boucle pas, mais doit commencer quand $Démarrer$ est vraie et s'interrompre quand l'indice de l'échantillon à lire atteint la $taille_F$ du tableau de son. L'indice de lecture du son est calculé à partir du temps écoulé depuis le dernier instant $t_{démarrer}$ où $Démarrer$ était vraie. La variable permettant de mémoriser ce repère temporel est initialisée à la valeur $-taille_F$ de sorte que, tant que $Démarrer$ n'a pas été vraie au moins une fois, l'indice de lecture donné par $t - t_{démarrer}$ sera toujours supérieur ou égal à $taille_F$, bloquant ainsi le rendu sonore.

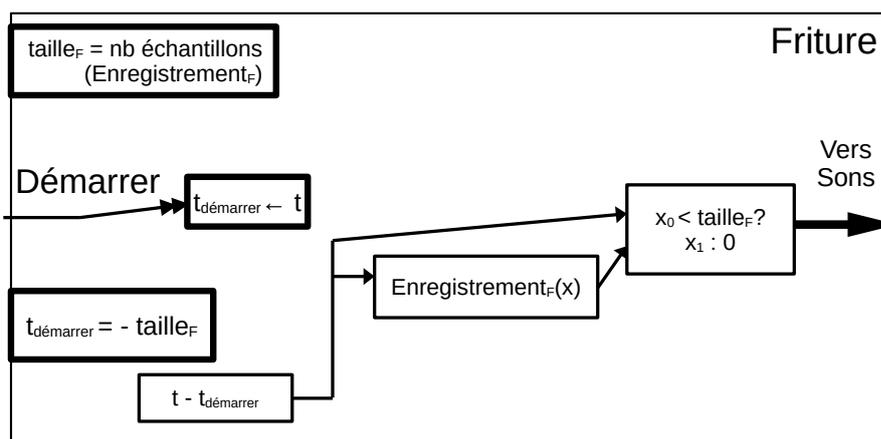


FIGURE 8.6 – Détail de la construction du complexe sonore *Friture*

Remarques sur les écologies *Synth* et *Nature*

Comme rapporté dans le chapitre 2 dédié à ce cas d'étude, une expérience réalisée à partir de ce sujet a consisté à tester et comparer trois assemblages différents de stimuli sonores, dits « écologies », nommés *Mixte1*, *Synth* et *Nature*.

L'écologie *Mixte1*, représentée ci-dessus, consiste en deux sons naturels et deux sons synthétiques. L'écologie *Synth* consiste uniquement en des sons synthétiques et *Nature* uniquement en des sons naturels, chacune réutilisant les éléments respectivement synthétiques et naturels déjà présents dans *Mixte1*.

Ce développement de plusieurs représentations les unes à partir des autres est l'occasion de s'interroger sur l'apport de la modularité du modèle pour l'adaptation à différents choix de sonification.

Synth Les complexes sonores *Ébullition* et *Craquèlement* sont remplacés par un seul complexe *Sonnerie* produisant une tonalité de fréquence 220 ou 660 Hz selon que l'anomalie de *TC* est négative ou positive. Les paramètres de gain pour ce complexe sont les mêmes que précédemment. Le complexe sonore *Friture* est remplacé par un complexe *Cloche*, déclenché exactement de la même façon en cas de dépassement du seuil pour *TP*.

La transformation du schéma de *Mixte1* en un schéma pour *Synth* peut consister à remplacer *Ébullition* et *Craquèlement* par un seul complexe *Sonnerie* prenant la fréquence à produire comme paramètre en plus du gain. Dans la couche des mappings aux paramètres perceptifs, on remplace alors la sélection d'*Ébullition* ou *Craquèlement* en fonction de la condition $TC(t) - 2000 > 0?$ par une sélection de la fréquence 220 ou 660 à passer en paramètre à *Sonnerie*.

Une autre option plus simple peut consister à modifier seulement le bloc « Sons », de sorte à remplacer *Ébullition* par un complexe *Sonnerie660* jouant la fréquence 660 Hz, *Craquèlement* par un autre complexe *Sonnerie220* jouant 220 Hz. Alors, aucune autre modification du schéma n'est nécessaire.

Nature Pour cette écologie, comme il n'est plus possible d'utiliser des paramètres de hauteur tonale dans un contexte de sons naturels pré-enregistrés, les paramètres perceptifs doivent être modifiés de façon plus drastique. Ainsi, l'adaptation du schéma nécessite des changements plus importants aussi. Les données d'entrée et les analyses préalables ne changent pas, puisque les informations à représenter sont toujours les mêmes, mais le bloc « Sons » et les deux couches de mapping aux paramètres perceptifs et de synthèse sont tous remplacés.

Plus généralement, on note que le bloc des données d'entrée et la couche d'analyse n'ont pas besoin d'être modifiés quels que soient les choix de représentation, tant que le but de la représentation est bien de permettre la supervision des mêmes critères pour le même type de fabrication.

8.1.2 Sonification immersive du suivi de protéines individuelles

Le but de ce cas d'étude est la réalisation d'un outil interactif pour l'étude des trajectoires de particules individuelles acquises par imagerie à super résolution. Ici ces trajectoires sont celles des protéines assurant la motilité cellulaire. La solution de sonification proposée à l'occasion de cette étude a été nommée *FluoSim-sonif* d'après l'outil de gestion, simulation et analyse des trajectoires *FluoSim*, auquel cette sonification a été intégrée. Le détail et la justification des choix de représentation pour ce cas d'étude sont donnés dans le chapitre 3 de la présente thèse.

La représentation de ce cas d'étude sous le format du modèle présente un certain nombre de points d'intérêt pour sa validation. En effet, toutes les sources de données proposées dans le modèle sont utilisées ici, puisqu'on choisit de laisser beaucoup de liberté à l'utilisateur dans les réglages de la représentation et qu'on offre une interaction pour le positionnement au sein des données. D'autre part, il s'agit d'une représentation pour laquelle le pas de temps du déroulement des données est réglable depuis l'interface. Enfin, les complexes sonores sont générés en temps réel d'après les points présents dans la trajectoire à chaque instant, ce qui donne l'occasion de tester la représentation factorisée de complexes sonores.

Vue d'ensemble La figure 8.7 donne le schéma, selon le format du modèle, du traitement des données pour la représentation *FluoSim-sonif*.

Ici, les données d'entrée sont partiellement pré-traitées et organisées par *FluoSim*. Il s'agit de tranches temporelles successives ($Plan(t)$) contenant chacune un ensemble de points ($Point_i(t)$) dans un espace en deux dimensions. Chaque point est caractérisé par sa position $Pos_i(t)$, son taux de fluorescence $L_i(t)$, et deux taux de diffusivité $D_i(t)$ et D_i calculés respectivement pour le mouvement du point le plus récent et pour sa trajectoire entière.

L'ensemble des points pour une tranche temporelle donnée est représenté comme la superposition de sons synthétisés individuellement et spatialisés dans le rendu stéréo autour d'une position virtuelle contrôlée par l'utilisateur (la position $UPos(t)$ et le point de visée $UVis(t)$).

Des réglages en plus, proposés dans l'interface utilisateur, permettent de choisir le pas de temps $\Delta t_d(t)$ pour le déroulement des plans, le rayon $R(t)$ du cercle d'audibilité autour de la position d'écoute, les bornes minimale ($min_F(t)$) et maximale ($max_F(t)$) pour le mapping de la diffusivité à des notes de musique, la sélection de la forme d'onde selon laquelle synthétiser les sons ($osc(t)$) et la sélection du fichier depuis lequel sont lues les données de **HRTF** pour la spatialisation.

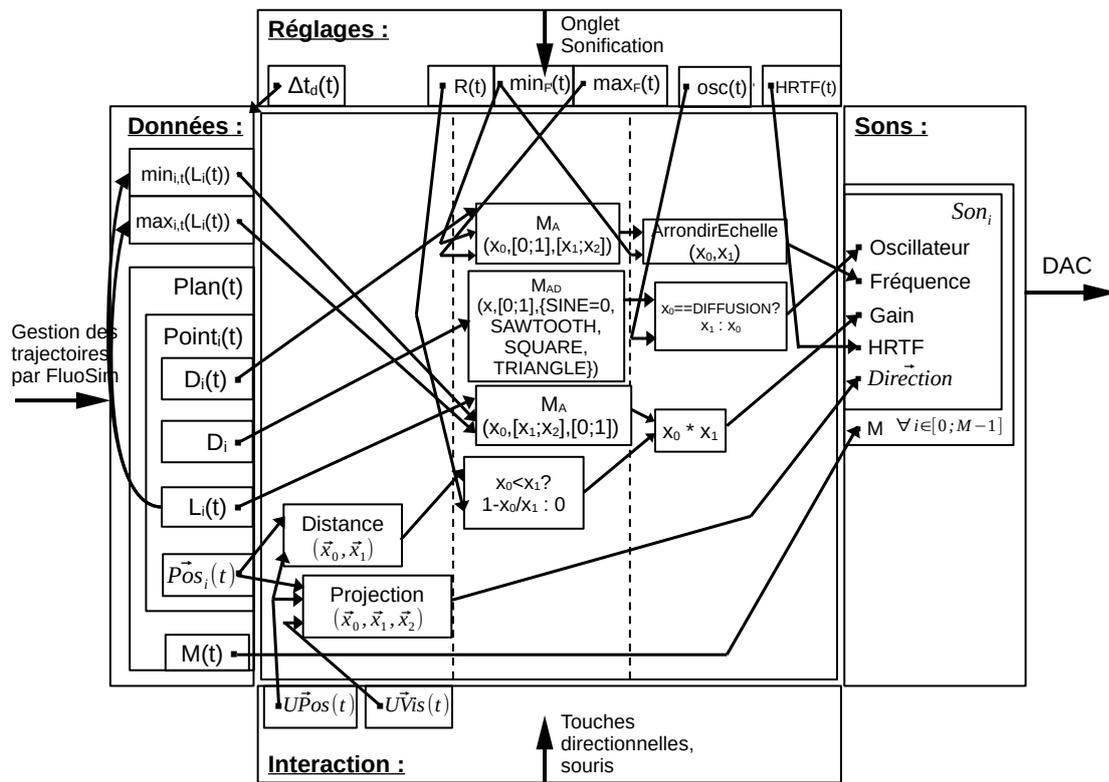


FIGURE 8.7 – La sonification interactive *FluoSim-sonif* du chapitre 3 décrite selon le format du modèle

Données d'entrée Dans les fichiers *.trc* gérés par *FluoSim* pour sauvegarder et charger les trajectoires, les différentes colonnes correspondent aux champs des données et les lignes correspondent à leur enchaînement temporel. Pour chaque ligne, les colonnes donnent, dans cet ordre : le numéro de la trajectoire, le numéro du plan, la position horizontale du point, la position verticale du point, un champ laissé vide pour le débogage, et la luminosité du point.

Les lignes sont ordonnées plan par plan pour chaque trajectoire l'une après l'autre, de sorte qu'une trajectoire est écrite chronologiquement de son début à sa fin, puis la trajectoire d'identifiant suivant est décrite à son tour du début à la fin, quitte à ce que son début soit plus tôt dans la chronologie que la fin de la précédente, et ainsi de suite.

FluoSim effectue une analyse préliminaire des positions successives dans chaque trajectoire afin d'en extraire les taux de diffusivité globaux D_i . Pour deux positions successives d'un même point, le programme calcule aussi un taux de diffusivité instantané $D_i(t)$. Ces deux types de grandeurs sont mis à l'échelle logarithmique

et normalisés, de sorte à ne prendre que des valeurs réelles entre 0 et 1.

Un parcours complet des taux de luminosité $L_i(t)$ pour tout instant t et pour toute trajectoire i permet d'en extraire la valeur la plus petite ($\min_{i,t}(L_i(t))$) et la plus grande ($\max_{i,t}(L_i(t))$) à la phase d'initialisation précédant le début de la représentation.

Les coordonnées horizontales et verticales des points à chaque instant sont concaténées dans des vecteurs de deux dimensions $\vec{Pos}_i(t)$.

Le nombre de trajectoires actives, donc de points à sonifier, pour chaque instant est communiqué au système de représentation sous la forme d'un nombre entier $M(t)$.

La structure de données ainsi construite pour la représentation prend la forme d'un enchaînement des plans au cours du temps, chacun étant caractérisé par son ensemble de points actifs $Point_i(t)$ et le nombre d'éléments de cet ensemble $M(t)$. Chaque structure $Point_i(t)$ est caractérisée par ses taux de diffusivité globale D_i et instantanée $D_i(t)$, son taux de luminosité $L_i(t)$ et sa position $\vec{Pos}_i(t)$. En dehors de ces structures, les valeurs pré-calculées et statiques du minimum et du maximum de la luminosité sont aussi communiquées à la représentation.

Paramètres réglables Cet outil devant permettre une exploration assez souple de l'ensemble de données, un certain nombre de réglages sont laissés à la discrétion de l'utilisateur dans l'interface. Dans la modélisation de cette représentation cela donne lieu aux variables suivantes, visibles dans le bloc du haut de la figure 8.7 :

- $\Delta t_d(t)$ représente le choix du pas de temps (en secondes) séparant un *Plan* du suivant dans les représentations synchronisées auditive et visuelle.
- $R(t)$ représente le rayon du cercle d'audibilité (selon une proportion de la taille de l'écran) formé autour de l'auditeur virtuel.
- $\min_F(t)$ représente le choix de borne minimale (choisi en tant qu'identifiant de hauteur **MIDI** et converti en Hertz) pour la sélection des notes dans la représentation.
- $\max_F(t)$ représente le choix de borne maximale (choisi en tant qu'identifiant de hauteur **MIDI** et converti en Hertz) pour la sélection des notes dans la représentation.
- $osc(t)$ représente le choix de forme d'onde (sous la forme d'un identifiant entier arbitraire entre 0 et 4) à appliquer dans la synthèse des sons. Les identifiants 0 à 3 correspondent aux formes d'ondes sinusoïdale, en dents de scie, carrée et triangulaire. L'identifiant 4 correspond à la sélection automatique de l'une de ces formes d'ondes en fonction du taux de diffusion global de la trajectoire.

- $HRTF(t)$ représente le choix du fichier .sofa à partir duquel sont lues les données de fonctions de transfert de la tête pour la spatialisation.

Données d'interaction L'interaction de l'utilisateur avec l'affichage visuel lui permet de se positionner et s'orienter en tant qu'auditeur virtuel dans la représentation auditive. Les mouvements peuvent être effectués avec les touches du clavier et, optionnellement, avec la souris. Dans ces deux cas, les données générées pour la représentation et reportées dans le bloc du bas de la figure 8.7 sont : le vecteur de deux valeurs $UPos(t)$ qui donne la position de l'auditeur dans le repère global de l'affichage, et l'autre vecteur de deux valeurs $UVis(t)$ pour le point vers lequel la tête de l'auditeur est orientée.

Mappings Le bloc des mappings se situe au milieu de la figure 8.7, et contient les opérations pour la traduction des structures de données en paramètres pour la synthèse des sons correspondants. Les trois paragraphes qui suivent donnent le détail de ces opérations pour chacune des couches de la progression données-perception-paramètres, en pointillés sur la figure.

Pré-traitements Les propriétés des données extraites en temps réel pour cette représentation sont représentées par les deux blocs d'opérations dans la zone de gauche de la figure 8.7. Il s'agit ici, pour chaque point des données, de déterminer sa position dans le repère défini interactivement par l'utilisateur.

La fonction $Distance(\vec{x}_0, \vec{x}_1)$ renvoie, pour deux points $\vec{x}_0 = (x_{x_0}, y_{x_0})$ et $\vec{x}_1 = (x_{x_1}, y_{x_1})$, la valeur :

$$Distance(\vec{x}_0, \vec{x}_1) = \sqrt{(x_{x_0} - x_{x_1})^2 + (y_{x_0} - y_{x_1})^2} \quad (8.4)$$

Ici, $Distance$ est utilisée pour calculer la distance entre la position $UPos(t)$ choisie par l'utilisateur et la position $Pos_i(t)$ d'un point donné.

La fonction $Projection(\vec{x}_0, \vec{x}_1, \vec{x}_2)$ renvoie la position du point \vec{x}_0 dans le repère orthonormé donné par l'origine \vec{x}_1 et l'axe $\vec{x}_2 - \vec{x}_1$. La figure 8.8 illustre le principe de cette projection.

Ici, $Projection$ est utilisée pour projeter les coordonnées $Pos_i(t)$ d'un point dans le repère formé par la position $UPos(t)$ et l'axe $UVis(t) - UPos(t)$, choisis par l'utilisateur.

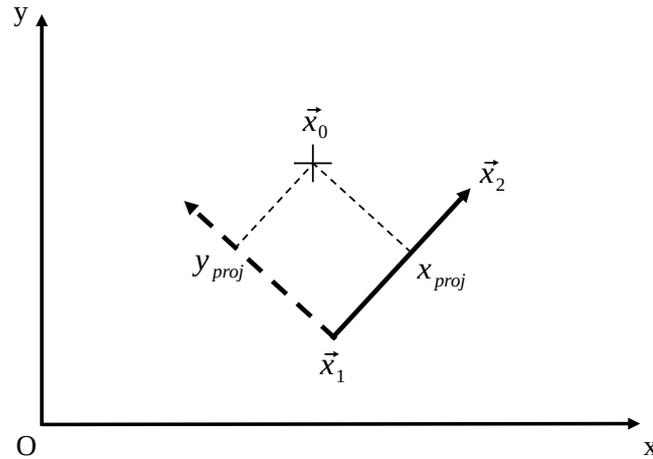


FIGURE 8.8 – La projection du point \vec{x}_0 dans le repère orthonormé donné par le point d'origine \vec{x}_1 et l'axe $\vec{x}_2 - \vec{x}_1$. La sortie est alors le vecteur (x_{proj}, y_{proj})

Paramètres perceptifs Dans la zone du milieu du bloc des mappings sur la figure 8.7 sont indiquées les opérations qui associent des paramètres perceptifs aux données d'entrée. Dans le cas de cette représentation, il s'agit d'associer une hauteur tonale à la diffusivité instantanée, un timbre à la diffusivité globale (en option) et deux facteurs d'atténuation de la sonie en fonction de la luminosité d'un point et de sa distance à l'auditeur virtuel :

- la diffusivité instantanée $D_i(t)$ est associée à une hauteur tonale entre $min_F(t)$ et $max_F(t)$ selon un mapping affine $M_A(x_0, [0; 1], [x_1; x_2])$.
- pour les cas où la diffusivité est associée à la sélection de la forme d'onde, c'est à ce niveau que cette sélection est effectuée, selon une fonction de mapping affine discret $M_{AD}(x, [0; 1], \{SINE = 0, SAWTOOTH, SQUARE, TRIANGLE\})$. Les noms « SINE », « SAWTOOTH », « SQUARE » et « TRIANGLE » se rapportent à des identifiants entiers entre 0 et 3.
- la luminosité est associée à un facteur réel d'atténuation entre 0 et 1 selon une fonction de mapping affine $M_A(x_0, [x_1; x_2], [0; 1])$, où x_1 et x_2 sont les extrema préalablement calculés pour les valeurs de luminosité.
- si la distance du point à l'auditeur virtuel est en dessous du rayon du cercle d'audibilité (condition $x_0 < x_1?$), un facteur d'atténuation $1 - x_0/x_1$ (avec x_0 la distance et x_1 le rayon) est calculé. Sinon, le point est considéré comme hors de portée et le facteur de sonie vaut zéro.

Paramètres de synthèse Dans la zone de droite du bloc Mapping sur la figure 8.7, on décrit trois opérations encore nécessaires pour adapter les résultats à la synthèse sonore :

- la fonction $ArrondirEchelle(x_0, x_1)$ renvoie la première fréquence fondamentale supérieure à x_0 dans l'échelle diatonique majeure de la note x_1 . Elle est utilisée ici pour arrondir, d'après l'échelle sélectionnée par l'utilisateur, la fréquence préalablement associée à la diffusivité instantanée.
- si la sélection de forme d'onde osc a pour identifiant la valeur DIFFUSION (4 dans le programme), on récupère la valeur préalablement calculée par le mapping affine discret. Sinon on utilise directement l'identifiant de forme d'onde.
- les deux facteurs d'atténuation calculés pour la luminosité et pour la distance sont multipliés entre eux (bloc $x_0 * x_1$) afin de donner un seul facteur d'atténuation représentatif des deux dans la synthèse sonore.

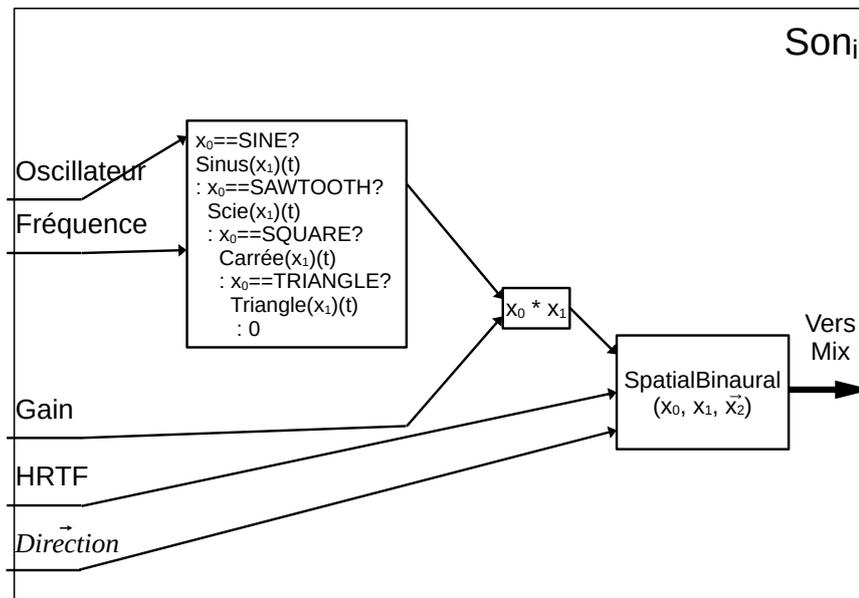


FIGURE 8.9 – Détail de la construction des complexes sonores individuels Son_i pour chaque point

Complexes sonores La sonification pour ce cas d'étude consiste en un nombre variable M^t de complexes sonores, tous synthétisés pour chaque tranche temporelle selon le même modèle Son_i ($i \in [0; M(t) - 1]$) annoncé dans le bloc de droite de la figure 8.7 et détaillé dans le schéma de la figure 8.9. Ces sons sont des tonalités produites selon différentes formes d'ondes nommées SINE, SAWTOOTH,

SQUARE et TRIANGLE dans l'application. Celles-ci sont identifiées par des nombres entiers entre 0 et 3, et correspondent aux oscillateurs décrits par les expressions 1.1 à 1.4 dans le chapitre 1 de la présente thèse.

Chaque complexe sonore est filtré sur les deux canaux d'un rendu stéréo par une fonction de spatialisation binaurale (*SpatialBinaural* sur la figure 8.9) employant les ressources de **HRTF** désignées par l'utilisateur dans l'interface, selon un vecteur *Direction* déterminé par le positionnement et l'orientation choisis par l'utilisateur à l'écran.

8.2 Cas d'études externes

Les cas d'études de la présente thèse correspondent à des besoins de représentation finalement assez similaires : des données évoluent au cours du temps, selon un pas Δt_d relativement grand, et l'écoute doit permettre de détecter des tendances au cours de cette évolution. Dans tous ces cas, la sonification proposée en solution revient à une variation sur le principe de la parameter mapping sonification, pour lequel le déroulement temporel du son est calqué sur celui des données.

Il existe pourtant des techniques de sonification alternatives, notamment dans leur gestion du rapport temporel entre les données et les sons, qu'un modèle générique de la sonification devrait aussi permettre de représenter. Nous étudions ainsi, dans cette section, deux autres approches pour la sonification tirées de l'existant : l'audification et la sonification model-based (à travers l'exemple du sonogramme des données).

8.2.1 Audification

L'audification est une technique de sonification particulièrement simple qui consiste en la transposition d'un signal de données (c'est-à-dire une seule dimension évoluant selon un axe temporel), dans un domaine de fréquences et d'amplitudes permettant leur intelligibilité à l'écoute. La perception d'informations ne se fait alors plus sur chaque point individuellement, mais sur un signal sonore constitué par ces points.

Vue d'ensemble La représentation du principe de l'audification, dans le cas où les valeurs sont connues a priori et enregistrées dans un fichier trace, est donnée dans la figure 8.10.

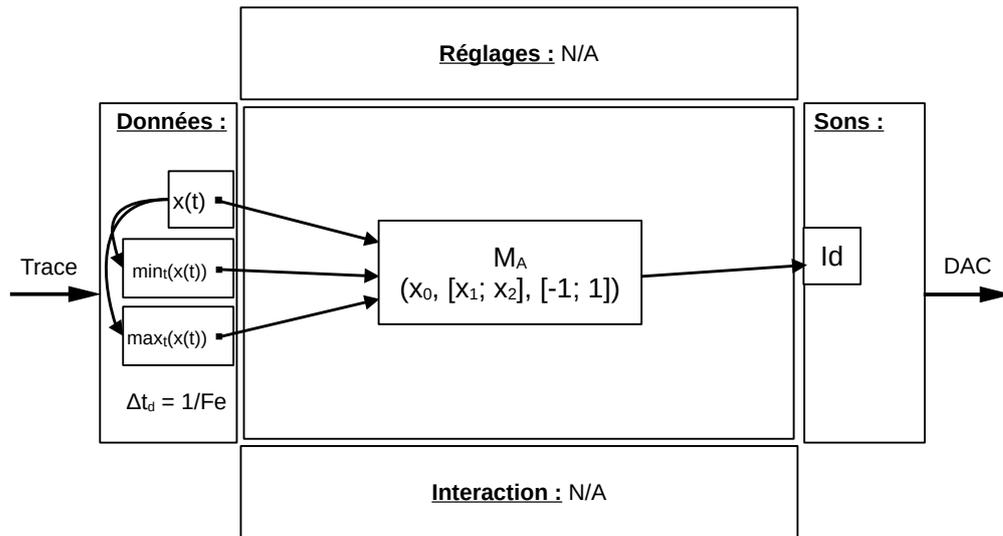


FIGURE 8.10 – Une représentation du principe de l'audification d'après le format du modèle

L'axe temporel du son se rapporte directement à celui des données, à un facteur d'accélération (ou ralentissement) près, et la seule transformation des valeurs d'entrée consiste à les associer à des valeurs de pression d'air (typiquement pour le son numérique : $[-1; 1]$).

Cette compression ou dilatation de l'axe temporel est donnée dans le modèle en faisant varier la fréquence d'échantillonnage Fe du son selon les besoins, et en indiquant que le pas de temps Δt_d pour le déroulement des données correspond à l'inverse de cette fréquence.

Données d'entrée On suppose pour cet exemple que le signal de données à transformer est entièrement connu d'avance et lu à partir d'un fichier de trace, par exemple sous un format de type **CSV**. On note x ce signal de données et $x(t)$ chaque valeur successive du signal pour chaque instant t . À l'initialisation, les deux valeurs minimale ($x_{min} = \min_t(x(t))$) et maximale ($x_{max} = \max_t(x(t))$) sont calculées pour l'ensemble des données.

Mappings La transposition des valeurs dans l'échelle des pressions d'air est effectuée par une fonction de mapping affine M_A allant de l'intervalle donné par les bornes x_{min} et x_{max} à l'intervalle $[-1; 1]$.

On note que le découpage du bloc des mappings en trois couches est ici difficile à mettre en œuvre puisque, selon l'interprétation qu'on choisit d'en faire, l'unique opération peut être placée dans n'importe laquelle des trois couches de mapping.

En effet, elle peut être vue à la fois comme un changement d'échelle des données dans la première couche, comme un mapping à un paramètre perceptif haut niveau de pression d'air dans la deuxième, ou comme un paramètre de synthèse du son échantillon par échantillon dans la troisième.

Comme ce type de représentation porte en fait sur la réalité objective de l'évolution des données, on pourrait argumenter que la représentation d'une progression des données vers des métaphores puis des paramètres de synthèse n'a finalement pas lieu d'être dans ce cas.

Complexe sonore Le complexe sonore de sortie pour une opération d'audification est directement constitué des données transformées. Symboliquement, on utilise dans le modèle une fonction de synthèse Id qui transfère les valeurs à la sortie DAC sans aucune transformation supplémentaire.

La retranscription de cette audification en tant que fonction de synthèse du signal s'écrit donc :

$$\begin{aligned}
 s(t, X, \theta, T) &= s(t, x, x_{min}, x_{max}) \\
 &= M_A(x(t), [x_{min}; x_{max}], [-1; 1]) \\
 &= -1 + \frac{(x(t) - x_{min})(1 - (-1))}{(x_{max} - x_{min})} \\
 &= 2 * \frac{(x(t) - x_{min})}{(x_{max} - x_{min})} - 1 \\
 &= 2 * \frac{(x(t) - x_{min})}{(x_{max} - x_{min})} - 1
 \end{aligned} \tag{8.5}$$

8.2.2 MBSon : le sonogramme des données

La sonification de type model-based se distingue des autres techniques par le fait que sa temporalité est indépendante du déroulement des données, le son étant à la place construit selon un temps imposé par les interactions de l'utilisateur avec l'outil de représentation. Pour que notre modélisation du principe de sonification puisse être considérée comme exhaustive, nous devons donc confirmer qu'elle permet aussi la représentation de ce type de techniques.

Dans cette section, l'exemple de sonification model-based que nous avons choisi de représenter est celui du **sonogramme des données**, décrit par Thomas Hermann dans [107] puis [96]. Cet exemple en particulier a été choisi pour sa relative facilité de compréhension et de prise en main par rapport à la plupart des approches model-based.

En effet, le rapport entre les données et le modèle physique est assez intuitif (les points de données servant d'objets physiques) et l'interaction à l'origine du son est simple et directe (placer une position de départ dans l'espace des données, autour de laquelle un rayon d'action augmente au cours du temps), ce qui laisse la possibilité de concentrer nos explications sur les rôles particuliers du pas de temps, du mapping et de la synthèse sonore dans cette famille de techniques.

On s'intéresse ici à un cas d'application du sonogramme pour lequel chaque point est caractérisé par son appartenance à une classe, ainsi que par sa position dans un espace à deux dimensions, obtenue par exemple par projection depuis un espace de plus grande dimensionnalité.

Vue d'ensemble Dans cette représentation, chaque point de données \vec{V}_i est associé à une particule de masse m attachée à un ressort dont la raideur k_i est proportionnelle, selon un facteur C , à son entropie locale de classes (*Local Class Entropy* en anglais, ou *LCE*). Le calcul de cette grandeur est donné par l'expression 8.9 : il s'agit d'une valeur qui représente le nombre de classes différentes dans le voisinage du point.

L'interaction consiste pour l'utilisateur à placer un point de départ \vec{V}_l au sein de l'espace des points, à partir duquel une onde d'excitation se propage à une vitesse v_0 et applique cette même vitesse initiale à chacune des particules rencontrées. Ces particules effectuent alors un mouvement oscillatoire, progressivement atténué par un facteur de résistance R jusqu'à retourner à un état d'équilibre immobile. À chaque instant, les quantités de déplacement des points de ce système sont considérées comme des valeurs de pression d'air, spatialisées en fonction de la position du point par rapport au point de départ (qui sert ainsi de position à l'auditeur) et superposées dans le rendu sonore pour former le son de la représentation.

Le théorème de Newton appliqué à un point du système donne l'équation différentielle suivante, dont la solution x_i correspond à la quantité de déplacement au cours du temps de la particule d'indice i :

$$\ddot{x}_i(t) + \frac{R}{m}\dot{x}_i(t) + \omega_0^2 x_i(t) = 0 \quad (8.6)$$

Dans l'équation 8.6, R est la résistance, m est la masse commune à toutes les particules, et $\omega_0^2 = k_i/m$, pour k_i la raideur du ressort associé au point \vec{V}_i .

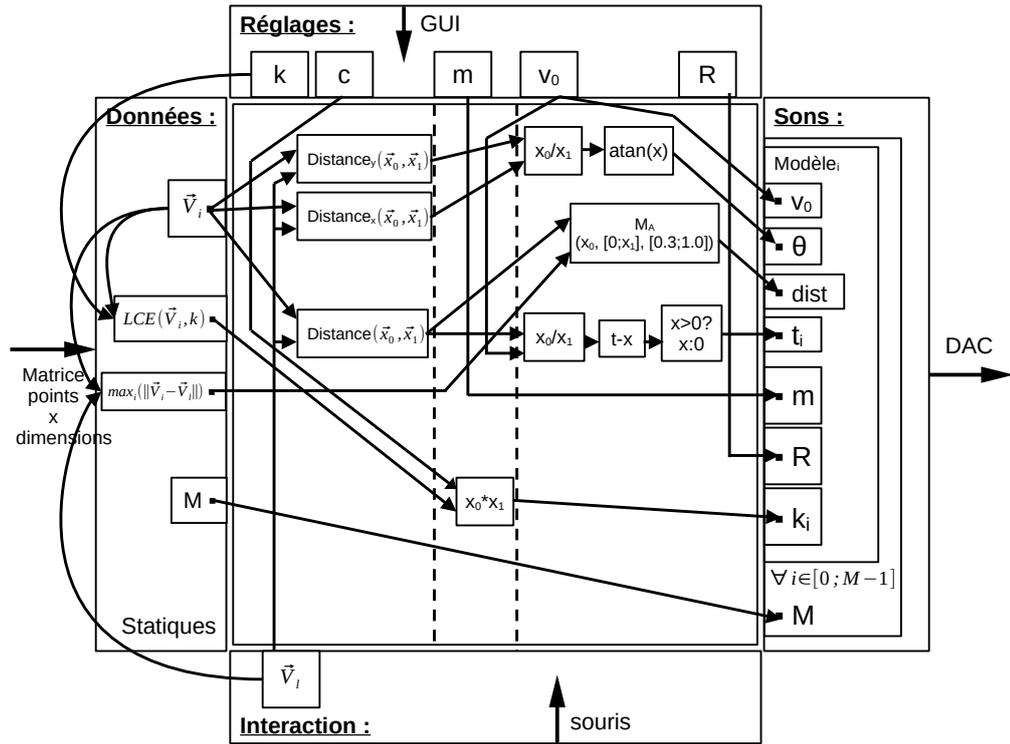


FIGURE 8.11 – Représentation du sonogramme des données d'après le format du modèle

La solution donnée pour cette équation dans [96] et utilisée pour la synthèse sonore est la suivante :

$$x_i(t) = A * \sin(\omega_f t) * \exp(-t/\tau) \quad (8.7)$$

Avec :

$$\begin{aligned} A &= v_0/\omega_0 \\ \omega_f &= \omega_0 \sqrt{1 - \left(\frac{R}{2m\omega_0}\right)^2} \\ \tau &= 2m/R \end{aligned} \quad (8.8)$$

La figure 8.11 donne la modélisation des réglages, interactions et mappings pour cette représentation, puis la figure 8.12 donne le détail de la construction des sons individuels $Modèle_i$ pour chaque point, avec la résolution du système donnée dans la figure 8.13.

Chaque complexe sonore $Modèle_i$ possède son propre paramètre de temps, noté t_i . Chacun est propre à un point \vec{V}_i et calculé d'après sa distance au point d'interaction \vec{V}_i et la vitesse v_0 de propagation de l'impact, de sorte à démarrer au moment où l'onde atteint le point : $t_i = t - Distance(\vec{V}_i, \vec{V}_i)/v_0$. Chaque t_i reste bloqué à zéro tant que le contact n'a pas eu lieu. On considère que le paramètre global t du modèle est positionné à 0 à chaque nouveau positionnement du point d'impact.

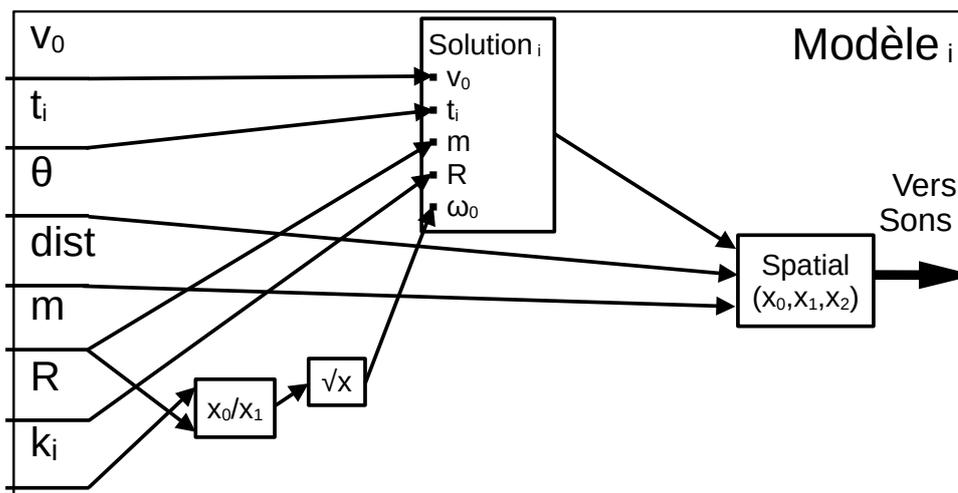


FIGURE 8.12 – Représentation du complexe sonore $Modèle_i$ utilisé dans le sonogramme des données

Données d'entrée Les données d'entrée pour cette représentation sont données par le bloc à gauche du schéma sur la figure 8.11.

On considère que ces données proviennent d'une matrice de M lignes et D colonnes, où les colonnes représentent les dimensions d'un espace et les lignes représentent des points dans cet espace.

Les points sont projetés dans un espace à deux dimensions, par exemple selon deux axes choisis parmi les colonnes de la matrice. On note l'ensemble de point projetés $\{\vec{V}_i, i \in [0; M - 1]\}$.

La fonction LCE correspond à un calcul de l'entropie locale de classes de chaque point, tel qu'il est donné dans [96]. Ce calcul consiste à effectuer la somme :

$$LCE(\vec{V}_i, k) = - \sum_{\alpha=1}^N p_{\alpha} \ln(p_{\alpha}) \quad (8.9)$$

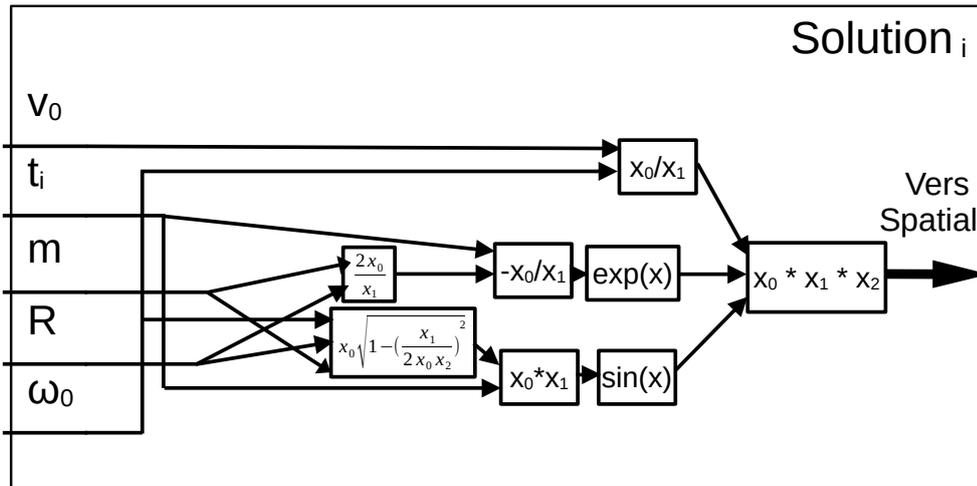


FIGURE 8.13 – Représentation de la fonction $Solution_i$ utilisée dans le complexe sonore $Modèle_i$

Avec N le nombre de classes en présence et p_α la densité de probabilité d'un point de classe α à la position \vec{V}_i . Cette densité de probabilité est calculée approximativement comme la proportion de membres de chaque classe parmi les k plus proches voisins de \vec{V}_i , préalablement détectés par un algorithme *k-nearest neighbours* dont le paramètre k est par ailleurs réglable pour l'utilisateur.

Chaque nouvelle définition de l'emplacement \vec{V}_i de l'auditeur (voir la description des données d'interaction plus bas) donne lieu à un nouveau calcul de la plus grande distance possible entre les points en présence et ce nouveau point d'action : $max_i(||\vec{V}_i - \vec{V}_i||)$.

Le nombre M de lignes lues dans la source des données, qui est aussi le nombre de points à représenter, est communiqué à la chaîne d'opération afin d'informer le nombre de complexes sonores factorisés qui doivent être construits.

Données d'interaction La source de données d'interaction est représentée par le bloc du bas sur la figure 8.11.

L'interaction offerte à l'utilisateur dans le sonogramme des données consiste à définir un point de départ pour la propagation d'une onde qui active chacun des points au fur et à mesure qu'ils sont rencontrés. Dans l'application hypothétique schématisée ici, on considère que la modalité d'interaction offerte à l'utilisateur est de cliquer à la souris sur l'emplacement souhaité dans un affichage visuel de l'espace des données.

La position de cet emplacement au sein de l'espace observé est communiquée à la chaîne de représentation sous le nom de \vec{V}_i .

Paramètres réglables Le modèle de sonogramme des données possède un nombre assez important de variables. Si beaucoup de ces variables pourraient être fixées à des valeurs constantes, ou faire l'objet de divers mappings à des propriétés des données, nous faisons ici le choix de considérer qu'elles sont toutes laissées à la discrétion de l'utilisateur, par exemple par le biais d'une interface dotée de curseurs et de champs numériques.

Le bloc en haut de la figure 8.11 montre ces variables et leurs points d'entrée dans la chaîne de traitement :

- k est le nombre de voisins pour le calcul de la *LCE*. S'il est mis à jour, la *LCE* est à nouveau calculée pour chacun des points.
- c est le facteur qui fait correspondre la *LCE* à une raideur de ressort.
- m est la masse commune à tous les points.
- v_0 est la vitesse de propagation du rayon d'activation, ainsi que la vitesse appliquée aux points rencontrés lors de cette propagation.
- R est la résistance du milieu dans lequel évoluent les particules. Elle influence la vitesse à laquelle les oscillations des points sont atténuées au cours du temps.

Mappings Le bloc du milieu de la figure 8.11 contient les enchaînements d'opérations pour le calcul des paramètres impliqués dans la synthèse sonore.

On note que la progression vers les paramètres de synthèse est plus ouverte à l'interprétation que pour les autres cas étudiés précédemment, puisque les paramètres qui entrent en jeu dans la synthèse du son se situent à un niveau plus métaphorique : au lieu d'être des paramètres psychophysiques ou musicaux comme la sonie, la hauteur ou le timbre, ce sont ici des paramètres qui se rapportent aux propriétés de l'objet simulé à l'origine du phénomène sonore.

Pré-traitements La couche de pré-traitement, représentée par la zone à gauche du bloc central des mappings sur la figure 8.11, consiste pour cet exemple à calculer la distance entre les points d'intérêt (chaque point \vec{V}_i et le point d'interaction \vec{V}_l), ainsi que ses composantes dans le repère de la représentation.

Les fonctions $Distance_x$ et $Distance_y$ donnent les écarts entre deux points, respectivement selon l'axe x ou y du repère de l'espace. Pour $\vec{x}_0 = (x_{x_0}, y_{x_0})$ et $\vec{x}_1 = (x_{x_1}, y_{x_1})$:

$$\begin{aligned} Distance_x(\vec{x}_0, \vec{x}_1) &= |x_{x_0} - x_{x_1}| \\ Distance_y(\vec{x}_0, \vec{x}_1) &= |y_{x_0} - y_{x_1}| \end{aligned} \quad (8.10)$$

La fonction $Distance$ est, quant à elle, la même que celle donnée par l'expression 8.4 dans le cadre du cas d'étude *FluoSim-sonif*.

Paramètres perceptifs Dans la zone centrale du bloc mappings de la figure 8.11, on place l'opération qui construit un paramètre de raideur de ressort en multipliant la LCE du point par le facteur c choisi par l'utilisateur.

Bien qu'il s'agisse, techniquement, d'un mapping à un paramètre physique plutôt que perceptif, on considère que sa contribution à la métaphore centrale des « ressorts » dans cette représentation le place plutôt parmi les mappings aux paramètres perceptifs.

Paramètres de synthèse La zone à droite du bloc des mappings sur la figure 8.11 contient les opérations pour le calcul des paramètres de synthèse du complexe sonore $Modèle_i$.

L'enchaînement des deux blocs x_0/x_1 et $atan(x)$ donne le calcul, à partir des deux composantes de distance précédemment calculées, de l'angle θ pour la spatialisation du complexe sonore.

La fonction de spatialisation nécessite aussi une mesure de la distance entre le point d'impact et le point à représenter. Celle-ci a déjà été calculée dans la couche des pré-traitements ($Distance(\vec{x}_0, \vec{x}_1)$), mais comme recommandé par [96], on la fait encore passer par un changement d'échelle vers un intervalle $[0.3; 1.0]$ afin d'empêcher que la spatialisation des points trop proches de l'auditeur sature le rendu sonore.

À partir de la distance entre le point d'impact et le point à représenter ($Distance(\vec{x}_0, \vec{x}_1)$), et de la vitesse de propagation de l'onde (v_0), l'enchaînement de blocs $x_0/x_1 \rightarrow t - x \rightarrow x > 0 ? x : 0$ donne le calcul du temps écoulé depuis le contact entre l'onde et le point à représenter. Ce temps est borné à zéro si le contact n'a pas encore eu lieu.

Complexes sonores Comme le montre le bloc à droite de la figure 8.11, la définition des complexes sonores pour cette représentation est factorisée pour le nombre de points de données à représenter.

La figure 8.12 donne le détail de la construction du complexe sonore individuel $Modèle_i$ associé à chaque point. Il s'agit du résultat calculé par la fonction $Solution_i$ à partir des paramètres physiques v_0 , m , R , ω_0 et du temps t_i propre au complexe, alors positionné dans le rendu stéréo par une fonction $Spatial$ selon la distance au point d'interaction et l'orientation par rapport aux axes de l'espace.

Le paramètre ω_0 en particulier est calculé à l'intérieur du bloc définissant $Modèle_i$, à partir des paramètres de masse m et de raideur du ressort k_i . Comme indiqué plus haut, $\omega_0^2 = k_i/m$, donc $\omega_0 = \sqrt{k_i/m}$.

La fonction $Solution_i$ est elle-même détaillée sur la figure 8.13. Les opérations consistent simplement à reconstituer l'expression 8.7 pour la résolution du modèle au cours du temps.

Remarques finales sur le sonogramme des données On retient de ce cas d'étude que, pour une sonification de type model-based, les mappings dans le bloc central du schéma donnent en fait le calcul des propriétés physiques d'un modèle sonore, plutôt que le calcul de paramètres pour des fonctions de synthèse additive.

C'est ce modèle sonore, et plus précisément la fonction qui donne la résolution de ses équations différentielles, qui est alors considéré comme un complexe sonore dans le bloc Sons.

Par ailleurs, comme ce genre de représentation est destiné à des données qui ne présentent pas de temporalité propre, on indique dans le schéma que les données d'entrée sont invariantes en fonction de t , et c'est l'évolution du système gérant l'interaction avec l'utilisateur qui donne les paramètres pour un déclenchement du son.

8.3 Conclusion

Dans ce chapitre nous avons testé la représentation, sous le format du modèle, d'une série d'exemples de sonifications représentant différents choix quant à la construction de sons paramétrés et la temporalité du son vis-à-vis des données. Selon l'optique des critères énoncés pour une modélisation dans la section 7.1, nous pouvons remarquer les points suivants.

Exhaustivité et non ambiguïté Pour la quantité relativement petite d'exemples testés, la traduction dans le format du modèle semble effectivement possible. Notamment, le modèle semble bien effectuer la liaison entre les types

de sonifications où le son suit l'évolution des données et celles où les données ne présentent pas d'évolution temporelle. Il reste nécessaire de vérifier, pour une plus grande quantité d'exemples aussi variés que possible, qu'aucune modalité de représentation ne reste ambiguë dans le format du modèle. L'absence d'ambiguïté pourra être vérifiée d'une façon plus concrète une fois que la traduction du schéma en algorithme, décrite dans le chapitre 9 suivant, aura été complétée. Les exemples repris dans ce chapitre sont uniquement des sonifications pour lesquelles le son est produit directement en tant que signal échantillonné, mais il faudra encore tester des exemples utilisant d'autres modalités comme les protocoles **MIDI** ou **OSC**.

Lisibilité Pour les réécritures manuelles des différentes techniques de sonification étudiées dans ce chapitre, il n'a pas toujours été aisé de produire une représentation graphique parfaitement lisible. Il n'est pas rare en effet que des flèches se croisent, ou que la taille de certains blocs doive être fortement réduite afin d'éviter qu'ils se chevauchent.

La répartition des opérations dans les trois couches du bloc des mappings s'est parfois avérée être une contrainte spatiale plutôt qu'un réel avantage pour la lisibilité. À notre sens, pourtant, aucun des schémas n'a fini par être surchargé au point de ne plus permettre de suivre visuellement les flèches pour reconstituer les opérations. Il sera cependant judicieux à l'avenir d'expérimenter plus largement sur la communication d'une représentation schématique de sonification à des testeurs extérieurs, notamment avec ou sans le support textuel.

Modularité L'organisation des opérations de mappings selon trois couches s'est avérée particulièrement utile dans le cas de la fabrication additive sonifiée, pour laquelle nous avons effectivement eu l'occasion d'expérimenter sur différentes évolutions des choix de mappings. Dans ces cas, nous avons pu constater la possibilité de réutiliser une partie des strates d'opérations et de ne modifier que la partie de la chaîne de traitement liée aux nouvelles modalités de représentation. À l'occasion de travaux futurs, il sera intéressant d'expérimenter sur d'autres conceptions itératives afin de constater de façon plus complète à quel point cette modularité facilite l'adaptation de représentations.

Chapitre 9

Perspectives et conclusion

Dans cette partie, nous avons rapporté nos réflexions sur le sujet principal de la présente thèse : la construction d'un modèle générique pour la sonification des données.

Nous avons d'abord présenté les critères d'un modèle de description de la sonification, notamment une série d'aspects variables pour tenir compte exhaustivement de tous les choix de représentation possibles.

Ensuite, notre étude de l'existant nous a permis de constater les difficultés toujours présentes pour la modélisation exhaustive du principe de la sonification, puis d'extraire certaines considérations, à notre sens judicieuses, pour un modèle plus complet de la sonification.

Ce modèle, et en particulier sa représentation graphique, ont ensuite été décrits en détail, puis testés sur un ensemble de cas d'études représentatifs d'un certain nombre d'approches différentes pour la sonification.

Un certain nombre de perspectives, principalement pour l'évaluation expérimentale, ont été notées à l'issue de cette première phase de validation dans le chapitre précédent : il serait judicieux d'enseigner le format de représentation du modèle à un public plus large de chercheurs susceptibles de concevoir des sonifications, afin de collecter leurs retours quant à la transmission de techniques à l'aide du modèle.

En plus de ces perspectives, nous rapportons dans le présent chapitre nos réflexions pour la création hypothétique d'un outil informatique de conception de la sonification basé sur le modèle, ainsi qu'un certain nombre de techniques envisageables pour la sonification qui n'ont pas encore été étudiées dans cette thèse.

9.1 Vers un outil dédié

Le modèle générique de la sonification décrit dans la présente thèse est encore à un stade relativement informel mais son but est d'informer la création, à l'avenir, d'un outil permettant à un utilisateur construire des sonifications. Nous présentons ainsi dans cette section les réflexions déjà engagées quant aux besoins d'un tel outil : d'une part en tant qu'interface graphique, d'autre part en tant que langage visuel pour la sonification.

On garde en tête les besoins déjà exprimés pour un tel outil générique par les auteurs du *Sonification Report* [128] puis étendus par de Campo et al. dans leur article préliminaire sur *SonEnvir* [51], qui peuvent être synthétisés selon les points suivants :

- L'outil doit permettre de gérer la chaîne de traitement et représentation d'un bout à l'autre : de l'acquisition des données au rendu sonore.
- L'outil doit pouvoir être intégré ou synchronisé avec d'autres modules de représentation ou d'interaction.
- La prise en main de l'outil doit être rapide et facile. Il doit ainsi être muni d'une interface graphique permettant la modification interactive de la chaîne de mappings et proposer un certain nombre d'exemples basiques pré-programmés.
- L'outil doit être gratuit, libre, open source, participatif et fonctionner indépendamment de la plate-forme.

9.1.1 Vers une interface de design

Afin d'offrir une modalité ergonomique et intuitive pour la construction de sonifications selon le modèle décrit ici, il est nécessaire que celui-ci soit présenté sous la forme d'un outil graphique.

L'interaction serait similaire à celle offerte par des outils de patching existants, comme *Pure Data* ou *ossia score*. L'interface à la souris permettrait de dessiner un outil de sonification en définissant des ports, des blocs d'opérations et en tirant des traits entre eux. Le clavier permettrait de rédiger le détail des opérations selon une syntaxe simple.

La compartimentation des modules explicitant les sources de données, opérations de mapping et rendus sonores, telle que décrite dans le chapitre 7, constituerait le plus haut niveau de l'affichage dans cette interface afin de clarifier l'interaction des différents éléments dans une sonification et de permettre une vue d'ensemble des ports de données et leurs connexions.

Quelques vérifications syntaxiques seraient automatiquement effectuées en temps réel par l'outil afin de vérifier la validité des expressions inscrites dans les blocs ainsi que l'absence de boucles dans le graphe.

L'outil posséderait un certain nombre de fonctionnalités pré-programmées qui se rapporteraient au moins aux usages les plus fréquents de la sonification : entre autres la lecture depuis un fichier **CSV**, la construction d'une interface utilisateur comprenant des boutons, champs numériques ou curseurs, l'extraction du minimum, du maximum ou de la moyenne des données, les fonctions M_A et M_{AD} décrites dans la section 7.7, et les quelques oscillateurs de base (sinusoïde, dents de scie, triangulaire, carrée, bruit blanc) décrits dans le chapitre 1.

Le but de ces briques de base est de permettre à un utilisateur novice de construire la plupart des outils de sonification basiques (pour des courbes auditives par exemple) uniquement à partir des fonctionnalités offertes par l'outil. Dans la description des mappings, on souhaite aussi que l'utilisateur n'ait pas trop souvent besoin de surcharger de détails certains embranchements logiques et numériques, comme nous l'avons fait par souci d'exhaustivité dans le chapitre 8 d'application aux cas d'études (voir par exemple la gestion des paramètres de *Mélo die* dans le schéma 8.2).

Comme pour *Pure Data*, il serait possible de définir et sauvegarder des modules à partir du langage de l'outil afin de pouvoir les importer librement par la suite. Le caractère open source de l'outil pourrait en plus permettre d'y ajouter des fonctionnalités à un plus bas niveau, notamment pour les modalités d'acquisition des sources de données ou pour le rendu sonore.

Enfin, l'outil devrait permettre d'expérimenter sur un prototype de représentation directement dans l'interface en offrant la possibilité de calculer un rendu sonore à la volée d'après le contenu du schéma (comme sous *ossia score*). Le schéma devra alors pouvoir être compilé sous différentes formes au choix (comme le permet par exemple l'**IDE** en ligne de *FAUST*), soit en tant que programme complet, soit en tant que plug-in pour des intégrations dans divers langages. La sous-section 9.1.2 ci-dessous place un certain nombre de notions de base pour une telle transformation d'un schéma en algorithme de sonification, d'après l'état actuel de nos réflexions sur cette perspective.

9.1.2 Vers une traduction automatique du schéma

Le rôle de la représentation graphique du modèle est de permettre la planification d'un outil de sonification sous un format schématique qui en donne une vue d'ensemble clairement lisible et intelligible. On souhaite par ailleurs que

ce format soit suffisamment exempt d'ambiguïté pour permettre la traduction de l'outil décrit sous la forme d'un programme, ou d'un plug-in.

Nous décrivons donc dans cette section la procédure à suivre à partir d'un schéma pour reconstituer l'expression, ou une collection d'expressions, donnant le signal sonore en fonction du temps et des sources de données : $s(t, X, \theta, T)$.

Le signal sonore à chaque instant est construit comme la somme des signaux générés par chacun des complexes sonores. Soit M le nombre de ces complexes, on rappelle l'expression de la sonification donnée dans 7.1 :

$$s(t, X, \theta, T) = \sum_{i=1}^M \phi_i(t, \vec{p}_i(t, X(t), \theta(t), T(t))) \quad (7.1)$$

Parcours des opérations

Pour exprimer chaque complexe sonore Φ_i en fonction du temps et des sources de données, on effectue un parcours à rebours de la construction du signal en partant à chaque fois du bloc d'opération dont la sortie donne sur la flèche « Vers Sons ».

Chacun de ces derniers blocs d'opérations utilise des variables écrites sous la forme x_k , avec k entre 0 et le nombre d'entrées du bloc moins un. Ainsi, un bloc possédant trois ports d'entrées utilise les variables x_0 , x_1 et x_2 , chacune correspondant à une flèche qui arrive, dans le même ordre de haut en bas, sur le côté gauche du bloc. Si le bloc ne prend qu'une seule valeur d'entrée, la variable est simplement notée x .

Chacune de ces variables peut être remplacée par l'opération du bloc relié à son port d'entrée. Si cette opération implique elle-même des variables en x_k , on peut alors continuer ce parcours en les remplaçant à chaque fois par l'expression inscrite dans le bloc précédent.

Quand le parcours rencontre une source de données (une dimension de X , θ ou T), et si cette source n'a pas été rencontrée au préalable, elle peut rejoindre la liste des valeurs à acquérir à chaque instant. À la fin du parcours, l'expression construite ne devrait plus traiter que des sources de données, des expressions auxiliaires (voir plus bas) ou des variables mémoires, et ne plus contenir aucune variable x_k .

Exemple La figure 9.1 donne le schéma pour un exemple de représentation de quatre dimensions A , B , C et D , évoluant au cours du temps selon un pas $\Delta t_d = 0.1s$, par une paramétrisation de deux complexes sonores Φ_0 et Φ_1 dont le détail est donné par la figure 9.2.

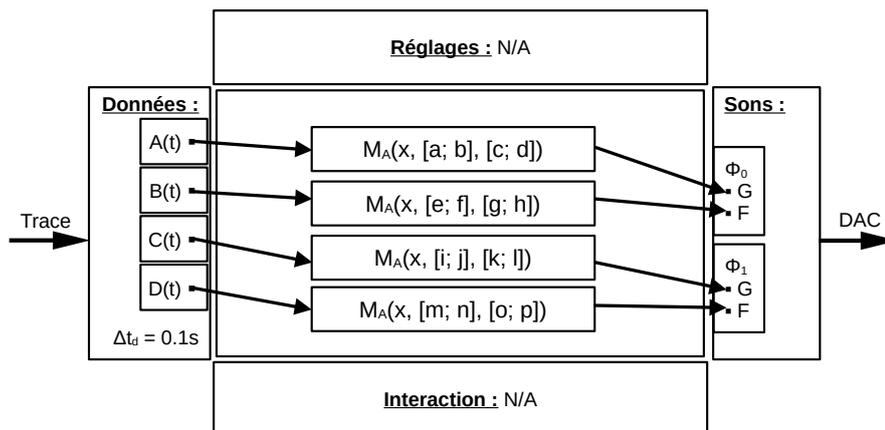


FIGURE 9.1 – Un exemple simple pour le parcours des opérations

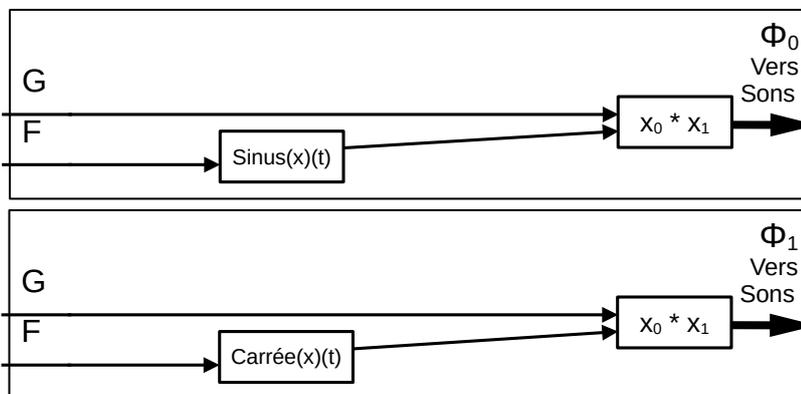


FIGURE 9.2 – Les complexes sonores utilisés dans les exemples des figures 9.1 et 9.3

Avant le parcours, l'expression est de la forme :

$$s(t, X, \theta, T) = \Phi_0(t, X(t), \theta(t), T(t)) + \Phi_1(t, X(t), \theta(t), T(t))$$

Pour la lisibilité, les sources de données θ et T sont omises puisqu'elles ne sont pas utilisées ici :

$$s(t, X) = \Phi_0(t, X(t)) + \Phi_1(t, X(t))$$

On développe d'abord Φ_0 en partant de sa sortie « Vers Sons » en haut de la figure 9.2, qui part d'une opération $x_0 * x_1$:

$$s(t, X) = \mathbf{x}_0 * \mathbf{x}_1 + \Phi_1(t, X(t))$$

Puis on remonte la flèche qui donne sur le premier port du bloc pour développer x_0 :

$$s(t, X) = M_A(\mathbf{x}, [a; b], [c; d]) * x_1 + \Phi_1(t, X(t))$$

Il reste à développer le x utilisé dans la première M_A . Il s'agit de la source de données $A(t)$:

$$s(t, X) = M_A(A(t), [a; b], [c; d]) * \mathbf{x}_1 + \Phi_1(t, X(t))$$

On développe de la même façon la branche qui donne la variable x_1 du dernier bloc de Φ_0 :

$$\begin{aligned} s(t, X) &= M_A(A(t), [a; b], [c; d]) * \text{Sinus}(\mathbf{x})(t) + \Phi_1(t, X(t)) \\ &= M_A(A(t), [a; b], [c; d]) * \text{Sinus}(M_A(\mathbf{x}, [e; f], [g; h]))(t) + \Phi_1(t, X(t)) \\ &= M_A(A(t), [a; b], [c; d]) * \text{Sinus}(M_A(B(t), [e; f], [g; h]))(t) + \Phi_1(t, X(t)) \end{aligned}$$

On peut ensuite développer Φ_1 selon le même processus pour obtenir l'expression complète de la sonification donnée par les figures 9.1 et 9.2 :

$$\begin{aligned} s(t, X) &= M_A(A(t), [a; b], [c; d]) * \text{Sinus}(M_A(B(t), [e; f], [g; h]))(t) \\ &\quad + M_A(C(t), [i; j], [k; l]) * \text{Carrée}(M_A(D(t), [m; n], [o; p]))(t) \end{aligned}$$

Expressions auxiliaires

Il est possible que la sortie d'un bloc d'opération soit utilisée en entrée de plusieurs blocs ultérieurs à la fois. Dans ces cas-là, afin d'éviter de reporter plusieurs fois l'enchaînement d'opérations donnant ce bloc (donc de générer un programme dans lequel du code serait répété), on ajoute une expression auxiliaire à la principale qui traite de cette branche des opérations. On peut par exemple nommer ces fonctions f_l avec l un compteur des expressions auxiliaires rencontrées au cours du parcours.

Exemple La figure 9.3 donne le schéma pour un exemple de représentation de trois dimensions A , B et C évoluant au cours du temps selon un pas $\Delta t_d = 0.1s$, par une paramétrisation de deux complexes sonores Φ_0 et Φ_1 dont le détail est donné par la figure 9.2. La paramétrisation est telle que la fréquence fondamentale calculée à partir de B influence les deux complexes sonores à la fois.

Avant le parcours, l'expression est de la forme :

$$s(t, X) = \Phi_0(t, X(t)) + \Phi_1(t, X(t))$$

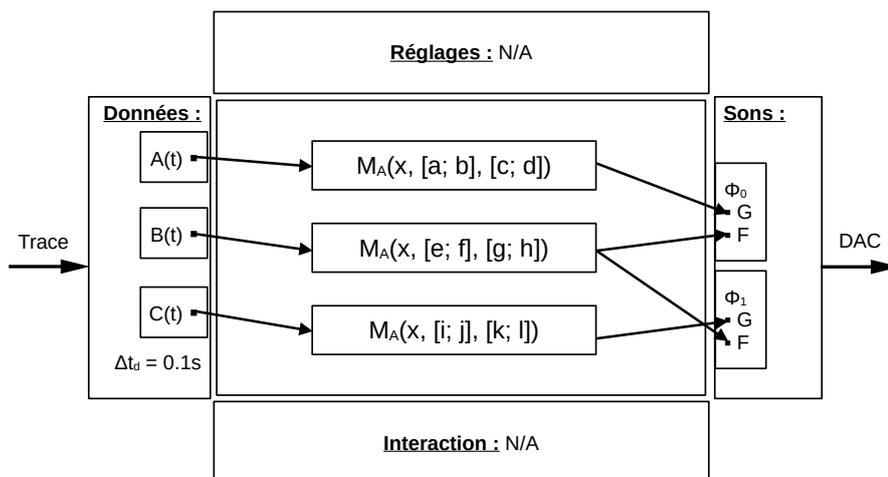


FIGURE 9.3 – Un exemple simple d'expression auxiliaire

On commence à développer l'expression en partant de la flèche « Vers Sons » pour Φ_0 . Le procédé est le même que pour l'exemple précédent, jusqu'à atteindre :

$$s(t, X) = M_A(A(t), [a; b], [c; d]) * \text{Sinus}(\mathbf{x})(t) + \Phi_1(t, X(t))$$

À cette étape, on détecte que le bloc donnant la prochaine variable x (qui est la fréquence F de Φ_0) sert d'entrée à une autre opération. On crée donc une variable f_0 pour cette expression auxiliaire, dans laquelle on développe cette branche du parcours :

$$\begin{aligned} f_0 &= \mathbf{x} \\ f_0 &= M_A(\mathbf{x}, [e; f], [g; h]) \\ f_0 &= M_A(B(t), [e; f], [g; h]) \end{aligned}$$

L'expression auxiliaire f_0 remplace alors la variable de l'expression principale à partir de laquelle elle a été détectée :

$$s(t, X) = M_A(A(t), [a; b], [c; d]) * \text{Sinus}(f_0)(t) + \Phi_1(t, X(t))$$

On continue ensuite le développement de l'expression en parcourant le schéma pour Φ_1 . Après quelques étapes, on rencontre à nouveau le bloc d'opération pour lequel on a créé la variable f_0 . Cette fois, c'est pour obtenir le paramètre F de Φ_1 :

$$\begin{aligned} s(t, X) &= M_A(A(t), [a; b], [c; d]) * \text{Sinus}(f_0)(t) \\ &+ M_A(C(t), [i; j], [k; l]) * \text{Carrée}(\mathbf{x})(t) \end{aligned}$$

On remplace alors la variable x par f_0 pour terminer le développement de l'expression :

$$s(t, X) = M_A(A(t), [a; b], [c; d]) * \text{Sinus}(f_0)(t) \\ + M_A(C(t), [i; j], [k; l]) * \text{Carrée}(f_0)(t)$$

Dans le programme ainsi généré, l'ordre des opérations à chaque pas de temps de la synthèse sonore est tel que l'expression auxiliaire est calculée avant son utilisation dans l'expression principale :

1. Pour chaque instant t :
2. $f_0 = M_A(B(t), [e; f], [g; h])$
3. $s(t, X) = M_A(A(t), [a; b], [c; d]) * \text{Sinus}(f_0)(t) \\ + M_A(C(t), [i; j], [k; l]) * \text{Carrée}(f_0)(t)$

Opérations initiales

Les blocs d'analyses préalables sont dénotés dans le schéma par le fait qu'ils se trouvent dans un des blocs de sources des données (valeurs, interactions ou réglages), qu'ils sont reliés à des dimensions des données par des flèches courbes, et que leurs intitulés décrivent les analyses à effectuer pour les calculer (par exemple dans la figure 8.7 : $\min(L_i(t)), \forall i, \forall t$). Ces sources de valeurs, quand elles sont rencontrées dans le parcours du schéma, sont ajoutées à une liste des opérations initiales à effectuer au début du programme de sonification généré. Par ailleurs, les blocs isolés donnant les premières valeurs des variables mémoires sont aussi à compter parmi les opérations initiales de la sonification.

Exemple La figure 9.4 donne le schéma pour un exemple de sonification impliquant des opérations à l'initialisation. Les valeurs de l'unique dimension A au cours du temps sont associées à une fréquence fondamentale pour le complexe Φ_0 , dont le détail est donné par la figure 9.5.

On suppose l'ensemble des valeurs de A au cours du temps entièrement connu a priori, ce qui permet d'en extraire les valeurs min et max pour définir précisément les bornes de l'intervalle d'entrée dans le mapping à la fréquence.

Le parcours du schéma donne l'expression :

$$s(t, X) = \text{Sinus}(M_A(A(t), [\min_A; \max_A], [c; d]))(t)$$

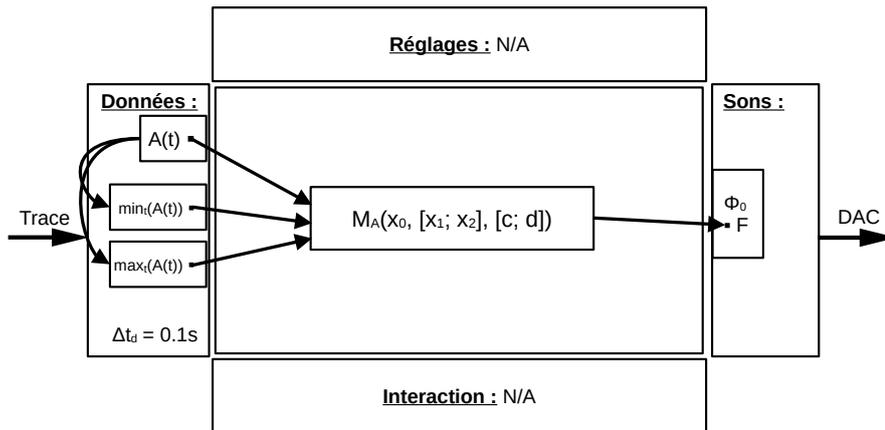


FIGURE 9.4 – Un exemple simple d'opération initiale

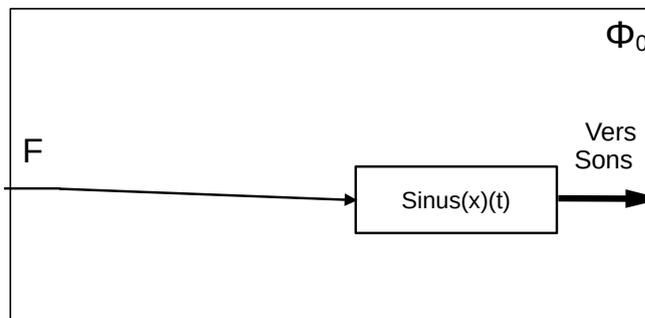


FIGURE 9.5 – Le complexe sonore employé dans les exemples des figures 9.4 et 9.7

Les valeurs $min_A = min_t(A(t))$ et $max_A = max_t(A(t))$, positionnées dans le bloc des données d'entrée et reliées à la dimension A par des flèches courbes, sont détectées comme des opérations initiales. Dans le programme généré, l'ordre des opérations est donc tel qu'elles sont calculées à l'initialisation du bloc des données, avant le début de la boucle temporelle du rendu sonore :

1. Calculer $min_A = min_t(A(t))$
2. Calculer $max_A = max_t(A(t))$
3. Pour chaque instant t :
4. $s(t, X) = Sinus(M_A(A(t), [min_A; max_A], [c; d]))(t)$

Affectations de variables mémoires

Dans le modèle, certaines variables dites « mémoires » servent à enregistrer des valeurs pour des usages ultérieurs. Leurs affectations sont indiquées dans des blocs

ne donnant sur aucune autre opération ultérieure, à la suite de blocs d'opérations conditionnelles.

On considère que la valeur d'une variable mémoire change avant tous les usages qui en sont faits pour le traitement de l'instant correspondant. Ainsi dans une traduction algorithmique d'un schéma de sonification, les affectations de variables détectées sont placées avant toutes les opérations qui nécessitent ces variables. Les affectations de variables sont elles-mêmes ordonnées en fonction de quelles variables dépendent de quelles autres variables (sachant que les boucles ne sont pas permises).

Exemple La figure 9.6 donne le schéma pour un exemple de représentation utilisant des variables mémoires. Il s'agit d'une audification pour laquelle, au fur et à mesure de la représentation, on décide de mettre à jour les bornes minimale et maximale de l'intervalle d'entrée du mapping qui donne les valeurs de pression d'air entre -1 et 1. Une vérification que $min_A \neq max_A$ évite que le mapping effectue une division par zéro lors du premier pas de temps.

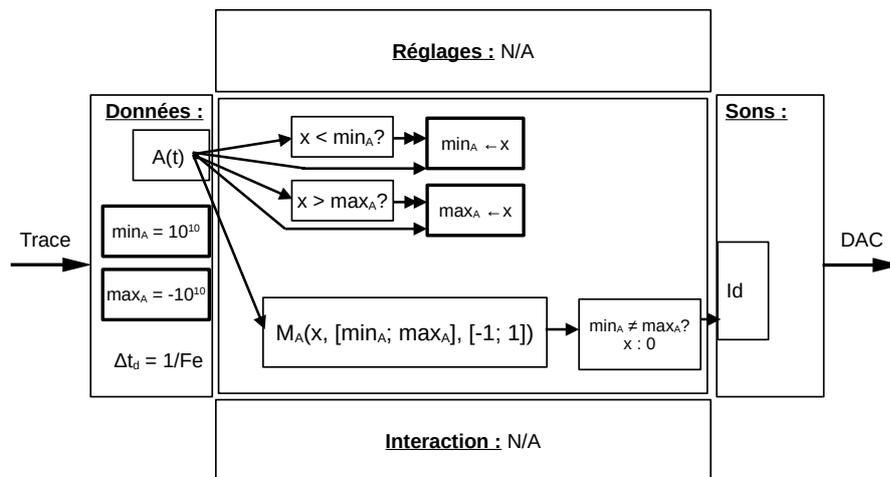


FIGURE 9.6 – Un exemple simple d'affectation de variables mémoires

Le parcours du schéma donne l'expression principale suivante :

$$s(t, X) = (min_A \neq max_A) ? M_A(A(t), [min_A; max_A], [-1; 1]) : 0$$

On détecte par ailleurs l'initialisation des deux variables min_A et max_A respectivement à 10^{10} et -10^{10} , ainsi que leurs deux affectations :

Si $A(t) < min_A$, alors $min_A \leftarrow A(t)$
 Si $A(t) > max_A$, alors $max_A \leftarrow A(t)$

L'ordre des opérations dans le programme ainsi généré est alors :

1. $min_A = 10^{10}$
2. $max_A = -10^{10}$
3. Pour chaque instant t :
4. Si $A(t) < min_A$, alors $min_A \leftarrow A(t)$
5. Si $A(t) > max_A$, alors $max_A \leftarrow A(t)$
6. $s(t, X) = (min_A \neq max_A)?M_A(A(t), [min_A; max_A], [-1; 1]) : 0$

Complexes sonores factorisés

Il est possible d'indiquer dans le modèle que plusieurs complexes sonores sont tous définis d'une même façon pour une collection d'éléments des données. Une telle factorisation se traduit graphiquement par un complexe sonore dont le nom est indexé par un i , et encadré par un bloc indiquant les bornes de la factorisation (par exemple dans le schéma 8.11 : $Modèle_i, \forall i \in [0; M - 1]$).

Dans ces cas-là, la traduction sous une forme algorithmique consiste à définir une expression auxiliaire, obtenue en additionnant un à un les éléments calculés pour chaque indice selon une boucle de type « Pour », puis ajoutée à l'expression principale.

Exemple La figure 9.7 donne le schéma pour un exemple de définition factorisée de complexes sonores. Les données d'entrée comportent un nombre $NCols$ de colonnes, qu'on choisit de représenter selon un même type de complexe sonore Φ_i (le signal sinusoïdal basique déjà détaillé sur la figure 9.5), de sorte que chacune soit associée, selon la même fonction de mapping, à une fréquence fondamentale dans le rendu sonore.

Le parcours du schéma pour les complexes factorisés Φ_i donne, pour $M = NCols$ et pour tout $i \in [0, M - 1]$:

$$\Phi_i(t, X(t)) = \text{Sinus}(M_A(A_i(t), [a; b], [c; d]))(t)$$

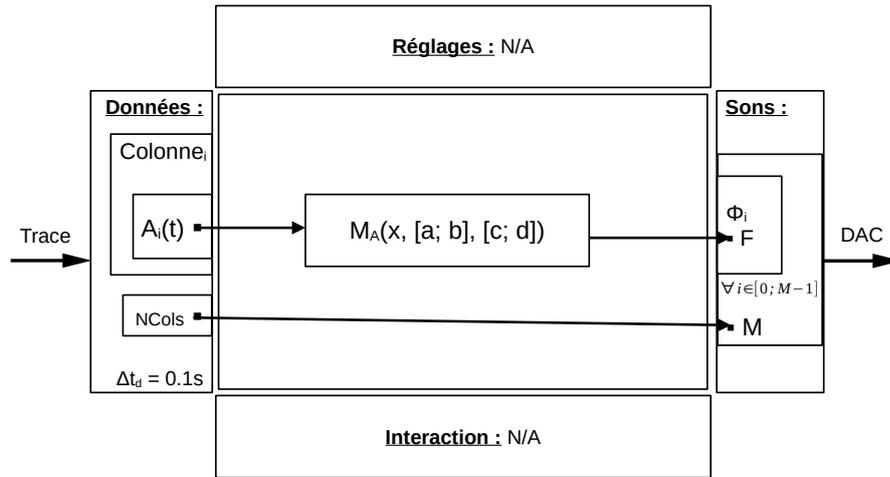


FIGURE 9.7 – Un exemple simple de factorisation des complexes sonores

Un complexe global Φ pour cette factorisation est alors construit à chaque instant t en additionnant les complexes Φ_i :

$$\Phi = 0$$

Pour i de 0 à $M - 1$:

$$\Phi \leftarrow \Phi + \text{Sinus}(M_A(A_i(t), [a; b], [c; d]))(t)$$

L'ordre des opérations dans le programme ainsi généré est alors :

1. $M = \text{NCols}$
2. Pour chaque instant t :
3. $\Phi = 0$
4. Pour i de 0 à $M - 1$:
5. $\Phi \leftarrow \Phi + \text{Sinus}(M_A(A_i(t), [a; b], [c; d]))(t)$
6. Fin Pour (i)
7. $s(t, X) = \Phi$

9.1.3 Prototypage sous ossia score

L'un des avantages attendus d'un modèle générique pour la sonification est qu'il permette de représenter et communiquer une sonification d'une façon qui contribue à sa reproductibilité. Nous nous sommes ainsi interrogés sur la possibilité de reconstruire sur un support commun les sonifications *Mixte1* et *FluoSim-sonif*, respectivement programmées initialement sous *SuperCollider* et en C++, d'après

leurs traductions dans le format du modèle. Notre préférence de support pour cette expérience s'est portée sur le séquenceur *ossia score*.

En effet, l'interface de patching proposée par *ossia score* se rapporte assez bien à la manipulation graphique des mappings envisagée dans un outil dédié à notre modèle, tout en offrant une courbe d'apprentissage qui paraît, à notre sens, plus rapide que celles de *Pure Data* ou *Max/MSP*.

Le séquenceur offre par ailleurs la possibilité de programmer l'activation d'éléments sonores en fonction de conditions ou de déclenchements, ce qui nous épargne dans beaucoup de cas de programmer explicitement la gestion du temps, comme notamment pour les complexes *Méلودie* et *Friture* dans la représentation *Mixte1*.

La souplesse de l'outil a aussi constitué un critère important pour sa sélection. Il est en effet possible, pour l'ajout de fonctionnalités, de choisir parmi un certain nombre d'options permettant de programmer, directement depuis l'interface, des modules compilés *Just-in-time* (JIT) en C++ ou Javascript, des blocs d'opérations basiques en ExprTK ou des **DSP** en *FAUST*.

L'interface d'*ossia score* encadre assez intuitivement la gestion de *devices* : des conteneurs structurés de valeurs numériques qui peuvent être connectés, entre autres, à des modules externes à l'aide de protocoles comme **OSC** ou **MIDI**. Ces conteneurs peuvent aussi servir pour gérer l'interaction de l'utilisateur avec une fenêtre de rendu visuel, ou être utilisés comme récipients pour mémoriser des états de paramètres, à l'instar des variables mémoires décrites dans notre modèle.

Quelques outils de base sont par ailleurs présents pour programmer un retour visuel dans des fenêtres dédiées, ce qui a été utile pour reprendre la représentation *FluoSim-sonif* qui inclut un affichage du positionnement de l'auditeur parmi les points en mouvement.

Ajout d'un module de lecture de **CSV**

Comme les cas étudiés au cours de la thèse partent tous d'une acquisition de données à partir de fichiers traces, nous avons commencé par utiliser les fonctionnalités de programmation JIT offertes par *ossia score* pour créer un module permettant la lecture et la restitution de données depuis des fichiers de type **CSV**.



FIGURE 9.8 – Le module de lecture de CSV (à gauche) suivi de la distribution des dimensions des données. Détail de la reconstruction de *Mixte1* sous *ossia score*

Ce module de lecture a été conçu comme un prototype du bloc « Données » de notre modèle. Deux phases sont définies dans son code : l’initialisation, au cours de laquelle les données sont acquises, organisées en structures et éventuellement analysées, puis le pas de temps pendant lequel les données ainsi définies sont communiquées au reste du système au cours du temps.

Pourtant, il s’agit encore d’une version assez spécifique du bloc « Données ». En effet on y considère que les données sont entièrement connues a priori (puisque lues depuis un fichier) et sont obligatoirement déroulées au cours du temps (plutôt que statiques).

Il n’a par ailleurs pas été possible de programmer un module dont les ports de sorties sont alloués et affichés dynamiquement en fonction des champs de données détectés pendant la lecture du fichier. À la place, la sortie du module a donc pris la forme d’un unique vecteur dont la taille est déterminée d’après la lecture. La récupération des différents champs des données se passe alors après la sortie du module, dans les blocs $xv[indice]$ (à droite de la figure 9.8).

Reconstruction de *Mixte1*

À partir des fonctionnalités offertes par *ossia score*, une grande partie des choix de complexes sonores de *Mixte1* ont pu être reconstitués :

- *Mélodie* est géré par une série de trois intervalles, constituant respectivement la note principale, la tierce et la quinte. L’enchaînement de ces intervalles est conditionné par le déroulement du temps de la

note : quand le temps écoulé dans l'intervalle dépasse la durée passée en paramètre, le démarrage de l'intervalle suivant est déclenché.

- *Bourdon* est géré par un **DSP FAUST** reconstituant les opérations de synthèse sonore décrites dans l'expression 8.3.
- *Ébullition* et *Craquèlement* sont gérés par deux pistes audio qui se déroulent en boucle et en continu. Chacun possède un paramètre de gain sonore, relié aux calculs de mappings et qui détermine si la piste doit être audible et à quel niveau.
- *Friture* est géré par un intervalle contenant une piste audio, déclenché en cas de dépassement du seuil de **TP** dans les calculs de mappings.

La figure 9.9 montre le projet score dans sa globalité, mettant en avant la reconstitution sur différentes pistes des différents complexes sonores. La figure 9.10 montre le détail des embranchements de mappings qui ont été reconstitués à partir des représentations schématiques de *Mixte1*. Une démonstration de cette reconstitution est disponible en ligne à l'adresse en pied de page¹.

Remarque L'expérimentation pour le prototypage sous *ossia score* a eu lieu en parallèle d'une première phase de réflexions sur le modèle, mais n'a pas suivi ces réflexions jusqu'au bout, faute de temps. Ainsi certaines idées apparues tardivement, comme la gestion explicite du paramètre de temps, n'ont pas pu être entièrement testées dans les prototypes. Pour cette raison, le redémarrage de la séquence de *Mélodie* en cas de changement des paramètres, ainsi que le compteur de temps pour l'alerte *Friture* ne sont pas encore pris en compte ici.

1. <https://youtu.be/vix0EFgGs1o> – visité le 4/1/23



FIGURE 9.9 – Capture d'écran du projet *ossia score* entier pour la reconstruction de *Mistel*

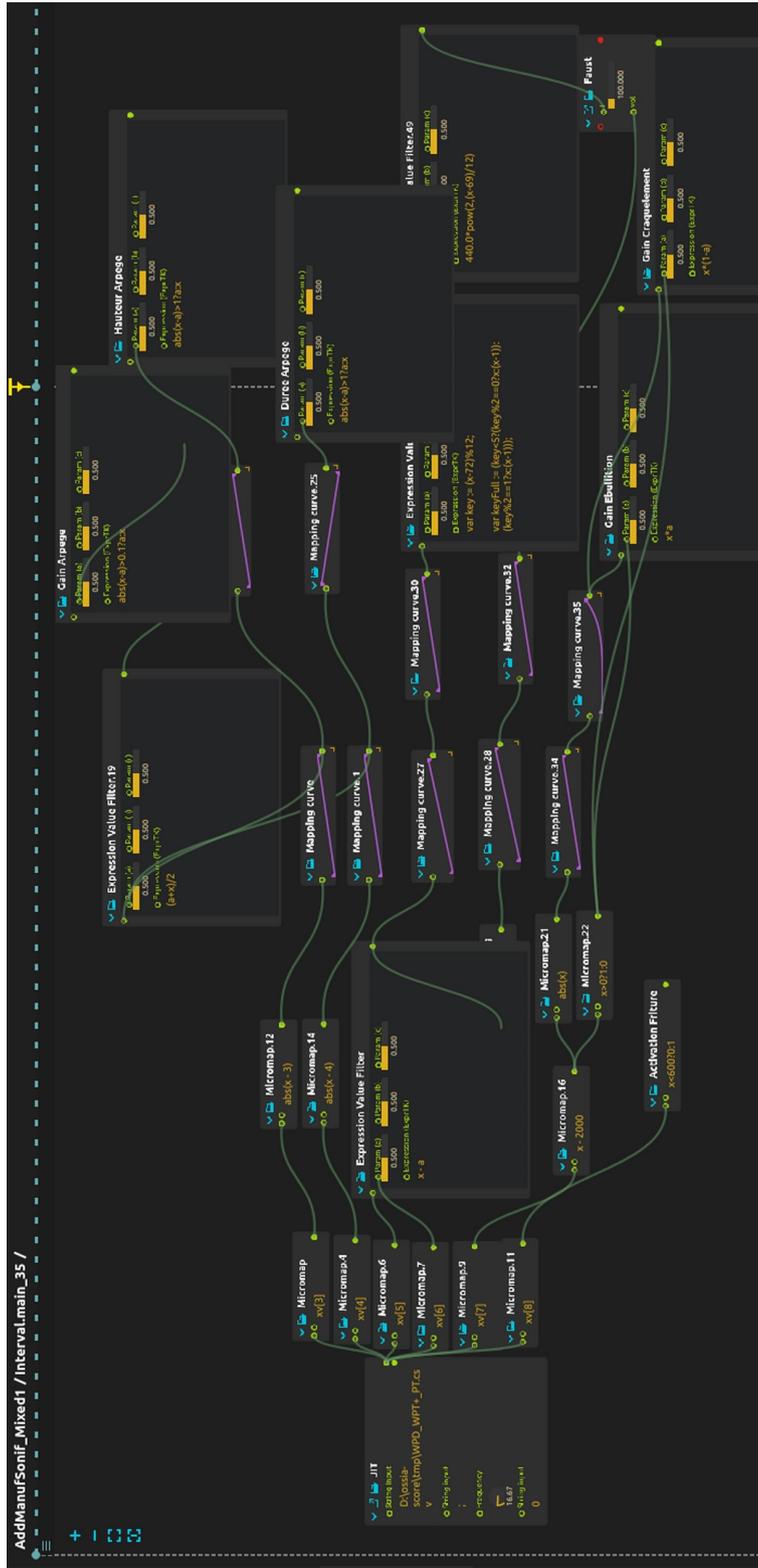


FIGURE 9.10 – Capture d'écran du détail des mappings pour la reconstruction de *Marte1* sous *ossia score*

Reconstruction de *FluoSim-sonif* (simplifiée)

Comme son nom l'indique, la représentation *FluoSim-sonif* part normalement du principe que les données sont préalablement acquises et analysées par l'outil d'analyse *FluoSim*, de sorte à pouvoir traiter directement des structures donnant les positions successives de l'ensemble des trajectoires au cours du temps. Le fait de la reconstituer à partir d'une lecture de fichiers traces, qui ne contiennent que les positions successives des points pour chaque trajectoire, oblige à faire certaines concessions quant à son fonctionnement, notamment pour l'utilisation des taux de diffusivité. Ainsi, dans cette version, les taux de diffusivité instantanée sont seulement calculés comme les distances entre les positions du point à un moment donné par rapport au précédent. Les taux de diffusivité globaux sont quant à eux omis dans un premier temps, faute d'implémentation des techniques d'analyse qui les produisent. La sélection du timbre liée à ces taux de diffusivité globaux a donc aussi été omise.

La reconstitution de l'interface d'exploration spatialisée est aussi simplifiée : l'auditeur virtuel est un point qui suit la position de la souris dans la fenêtre de rendu et il n'y a pas de moyen de l'orienter autrement que vers le haut de la fenêtre.

La version d'*ossia score* utilisée pour les expériences de prototypage (3.0.6) ne permet pas de factoriser la définition de complexes sonores. Notre reconstitution de *FluoSim-sonif* se contente donc de représenter chaque trajectoire de la trace l'une après l'autre, sans les faire coexister. Le premier état de chaque trajectoire est marqué par une hauteur caractéristique choisie arbitrairement à 1100 Hz, de sorte qu'un « bip » plus aigu que le reste soit émis à chaque changement de trajectoire.

La représentation *FluoSim-sonif* simplifiée permet ainsi à son utilisateur de se positionner dans une fenêtre de rendu visuel montrant l'évolution de chaque trajectoire l'une après l'autre, sous la forme d'un point changeant de couleur pour chaque trajectoire. Le rendu sonore consiste en un seul complexe sonore : une tonalité sinusoïdale dont la hauteur tonale suit le mapping initialement décrit pour *FluoSim-sonif*, et dont la répartition dans les canaux stéréo simule son positionnement autour de l'auditeur virtuel.

La figure 9.11 est une capture d'écran du projet score de la reconstitution de *FluoSim-sonif*. Il s'agit d'un seul sous-programme dans lequel les sources de données sont mappées à un générateur de signal et un module *FAUST* de spatialisation. On peut voir en bas à gauche les fonctionnalités qui gèrent le rendu visuel et son interaction avec le rendu sonore. La figure 9.12 offre une vue légèrement rapprochée afin de donner le détail de la chaîne de traitement spécifiquement pour le rendu sonore. Une démonstration de cette reconstitution est disponible en ligne².

2. <https://youtu.be/170fs1DXQ40> – visité le 4/1/23

9.2 Conclusion

Notre objectif pour la présente thèse était de concevoir une modélisation générique du principe de sonification qui soit aussi inclusive que possible des différents choix envisageables pour ce type de représentation.

En partant d'une vision relativement bas niveau de la sonification comme une fonction de synthèse d'un signal sonore dont les paramètres sont donnés par les données numériques d'entrée, nous avons décrit un modèle permettant de détailler les opérations menant à la synthèse depuis les données. Nous proposons notamment de compartimenter les entrées de cette fonction selon qu'elles traitent du système de données à représenter ou du rôle que peut jouer l'utilisateur dans le procédé de représentation.

De par sa modularité, ce modèle se prête assez bien à une représentation visuelle schématique, à la façon des langages de type « patching » comme *Pure Data* ou *ossia score*. Dans la présente thèse, nous avons commencé à formaliser la syntaxe de ce langage graphique, qui est envisagé comme la forme que prendront le plus souvent les descriptions de sonification d'après le modèle.

De plus, des réflexions ont été entamées sur la perspective d'une interface permettant de créer des sonifications d'après les spécificités du modèle. L'implémentation d'un tel outil constitue la prochaine étape de notre recherche sur un modèle générique de la sonification des données.

En guise de perspectives, nous notons pour finir un certain nombre d'éléments qui peuvent être impliqués dans la sonification, mais que le modèle ne semble pas encore permettre de clarifier tout à fait

- Les **sons pré-enregistrés** ont un statut particulier que le modèle ne permet pas encore d'explicitier entièrement. S'il est bien possible d'indiquer l'enchaînement précis d'opérations donnant la synthèse d'un signal, il n'a pas de réel moyen de donner des informations sur l'utilisation d'un pré-enregistré. Dans l'état actuel du modèle, on suppose juste que le chargement d'un fichier sonore est géré à un autre niveau par d'autres modules, que son contenu est placé dans un tableau dont la taille est connue, et que le son est lisible échantillon par échantillon en itérant sur les éléments de ce tableau.
- Bien que l'éventualité soit évoquée dans les réflexions préliminaires, l'application du modèle générique à une **sortie MIDI** n'est pas étudiée dans cette thèse. Intuitivement, il ne semble pas qu'une extension à ce format impose de contradiction ou de complication particulière puisque, au lieu d'exprimer l'échantillon du signal sonore à un instant t , on devrait en fait exprimer le message (ou absence de message) à émettre à chaque instant qui passe. Dans le cas d'une écriture de fichier **MIDI**, la temporalité de la

sortie sonore se retrouverait dans les intervalles de temps qui ponctuent les évènements successifs.

- Le **son sur plusieurs canaux** est utilisé à quelques reprises dans les cas d'études : le complexe *Bourdon* est construit différemment sur les deux canaux stéréo, et les éléments sonores de *FluoSim-sonif* ou du sonogramme des données sont spatialisés dans un rendu binaural. Cependant, le format du modèle ne prévoit pas encore de notation générale permettant de clarifier la répartition de la (ou des) sortie(s) de la fonction de synthèse dans un ou plusieurs canaux audio. On se contente actuellement d'une fonction *Stéréo*(x_0, x_1), utilisée par exemple pour *Bourdon*, qui permet d'expliciter que la valeur x_0 doit être communiquée au canal gauche et x_1 au canal droit. Les fonctions *SpatialBinaural* de *FluoSim-sonif* et *Spatial* du sonogramme produisent un rendu réparti et filtré sur deux canaux gauche et droite. Dans tous les autres cas, sauf si le contraire est explicitement indiqué, on suppose que la sortie d'une fonction de synthèse est communiquée à l'identique à tous les canaux de rendu sonore.

Conclusion générale

La représentation auditive des données n'a commencé à être reconnue comme une possibilité sérieuse qu'à une époque relativement récente, ce qui fait qu'elle n'a pas encore tout à fait atteint la maturité qui lui permettrait d'être mise en œuvre aussi naturellement que la visualisation face à des problématiques nécessitant la représentation de données.

Notre objectif au cours de cette thèse a été de contribuer à cette maturation de la sonification en plaçant les bases d'un modèle générique pour encadrer les réflexions sur ce type de représentation. Ce modèle doit offrir une façon de se représenter le concept en général, ainsi que ses applications spécifiques, en prenant toujours la même forme à quelques variables près.

À partir de notre étude de l'existant et de notre travail sur trois cas d'études répondant à divers besoins de représentation, nous avons décrit de façon informelle une première version de notre modélisation générique de la sonification, dont le but est de permettre de définir assez librement des techniques de sonifications employant différentes modalités de rendu sonore, d'acquisition de données, d'interaction ou de réglages personnalisés.

Dans ce chapitre, nous résumons les problématiques soulevées et les travaux commencés pour les trois cas d'études abordés, puis la progression qui a mené au modèle générique proposé. Pour chacun de ces points, nous rapportons la façon dont le travail a été entamé à différents degrés allant de l'étude de l'existant à l'évaluation de prototypes, puis nous listons les divers axes de poursuite des travaux dont beaucoup pourraient, à notre sens, faire l'objet de nouvelles thèses plus spécifiques à l'avenir.

9.3 Cas d'étude - Supervision de la fabrication additive

En collaboration avec **ESTIA** Recherche, Addimadour et Sigma-Clermont, ce cas d'étude s'inscrit dans une initiative plus large de recherche visant à

optimiser l'implication humaine dans les processus industriels en renforçant les modalités d'interaction et de supervision offertes aux opérateurs. Pour ce travail en particulier, il s'agit d'offrir une modalité auditive à la supervision d'un procédé de fabrication additive de type wire-arc. La représentation auditive développée doit permettre d'écouter à distance des notifications périphériques sur le déroulement du procédé, émises en fonction de captures (encore en grande partie hypothétiques) de grandeurs physiques sur la machine. Le chapitre 2 de la présente thèse donne un compte-rendu détaillé du travail sur ce sujet.

9.3.1 Contribution

L'essentiel du travail sur ce cas d'étude a été axé sur un processus de conception itérative d'une représentation auditive optimale pour cette problématique. Cette conception a consisté à s'interroger sur les critères de qualité pour l'évaluation d'une représentation auditive pour la supervision périphérique de processus, puis à en déduire les données à mesurer au cours d'expériences utilisateurs pour obtenir une quantification de cette qualité. Un protocole expérimental a alors été développé afin de comparer, dans un contexte de laboratoire, les représentations auditives successivement proposées pour la tâche de supervision périphérique étudiée.

L'échantillon de dimensions physiques spécifiques à cette tâche de supervision consiste en deux catégories assez distinctes de grandeurs : deux températures d'une part et deux géométries d'autre part. Cette particularité nous a encouragés à formuler une stratégie pour faciliter la détection des notifications, consistant à les répartir dans deux groupes de sons distincts. Dans l'écologie sonore volontairement incohérente ainsi construite, les températures ont été représentées par des sons naturels se rapportant aux états de l'eau et les géométries ont été représentées par des sons tonals synthétiques.

Afin de vérifier que cette stratégie était judicieuse, nous avons aussi produit deux autres écologies dans lesquelles les choix de sons naturels ou tonals étaient cohérents, puis nous avons effectué l'évaluation comparative de ces différents choix selon le protocole susmentionné. Les résultats préliminaires suggèrent que les sons naturels seuls offriraient l'expérience utilisateur la moins stressante, que les sons synthétiques seuls permettraient la meilleure réactivité aux notifications, et que le mélange des deux permettrait la meilleure distinction des notifications.

9.3.2 Perspectives

Améliorer la sonification Afin de répondre à la problématique originale de ce cas d'étude, un premier axe de recherche pourra consister à continuer d'itérer sur différents designs de sonifications pour la tâche de supervision. En attendant

la possibilité d'expérimenter en situation réelle, on pourra continuer de mettre à profit le protocole d'évaluation comparative en laboratoire pour mesurer les progrès d'une représentation à la suivante.

Améliorer le protocole expérimental Ce protocole d'évaluation lui-même gagnera à être perfectionné à l'avenir. On pourra ainsi chercher à rendre l'exercice de distraction plus difficile, dans une certaine mesure, afin de rendre l'estimation de la capture attentionnelle du son plus précise. Pour cela, entre autres, une attention plus grande pourra être apportée à l'enregistrement des erreurs commises lors de l'exécution de la tâche principale. Des phases supplémentaires de mesure des capacités individuelles pourraient aussi permettre de débiaiser et uniformiser les résultats des testeurs, notamment vis-à-vis de l'utilisation d'une souris. La simulation du contexte de travail prévu pourrait être renforcée par l'ajout des sons de la fabrication dans l'environnement sonore de test, voire par l'implication de ce procédé d'évaluation dans des conditions hybrides rappelant celles d'une fabrication wire-arc.

Expérimenter sur la cohérence écologique La problématique de la cohérence écologique soulevée dans le cadre de ce cas d'étude pourra être poursuivie à l'avenir. Le travail pourra consister à continuer de faire varier les quantités mélangées de sons naturels ou synthétiques en s'interrogeant sur lesquels sont, d'après l'expérience, les plus judicieux pour l'un ou l'autre type de grandeur physique. En plus de cette tâche de supervision déjà étudiée, d'autres cas d'études pourront être abordés sous la même optique, afin de s'assurer de l'homogénéité des résultats observés pour diverses collections de grandeurs physiques, voire pour des représentations vouées à d'autres usages que la supervision de processus (sous réserve d'utiliser un autre protocole expérimental approprié pour ces autres usages).

Expériences en situation réelle Il est déjà prévu qu'une thèse démarrée en 2022 à **ESTIA** Recherche reprenne la problématique de la supervision des processus de fabrication par le biais de diverses techniques de réalité augmentée, y compris la sonification partiellement développée dans la présente thèse. Comme ce travail aura lieu à proximité du site sur lequel les fabrications sont effectuées, il sera plus aisé de rapprocher les expériences de supervision du contexte réel de travail.

Alors que les travaux sur les capteurs des critères de fabrication finiront par permettre d'obtenir de réels flux de données en temps réel, il deviendra nécessaire de produire un véritable programme remplaçant le prototype *SuperCollider* pour la synthèse du son de supervision.

Sur le long terme, l'espoir est que les progrès pour la capture des critères de fabrication et les techniques de représentation de ces critères convergeront vers une situation où les deux seront prêts au même moment et compatibles l'un avec l'autre, de sorte à permettre naturellement une implémentation réelle de la supervision sonifiée du processus.

9.4 Cas d'étude - Suivi des trajectoires individuelles de protéines

Ce cas d'étude effectué en collaboration avec l'IINS porte sur l'analyse d'imagerie à super résolution de l'activité des cellules, en particulier des trajectoires de protéines capturées par fluorescence. La représentation et l'observation visuelles jouant déjà un rôle assez important dans la détection de comportements lors de l'analyse de trajectoires individuelles, on s'interroge ici sur le potentiel de la modalité auditive pour renforcer (voire remplacer dans certains cas) cette observation. Le chapitre 3 de la présente thèse donne un compte-rendu détaillé du travail sur ce sujet.

9.4.1 Contribution

En nous basant sur une étude des solutions de sonifications souvent associées à l'exploration de données évoluant dans l'espace et dans le temps, nous avons décrit une solution consistant à offrir une interaction immersive au sein d'un espace auditif de sons basiques associés aux données. Un prototype de cette représentation a été intégré à l'outil de simulation et d'analyse *FluoSim*.

9.4.2 Perspectives

Évaluations de l'expérience utilisateur Dans l'état actuel du projet, la prochaine étape du travail consistera à réaliser des évaluations de l'expérience utilisateur pour le prototype conçu. Un manuel utilisateur a été rédigé pour expliquer les fonctionnalités de sonification ajoutées à *FluoSim*. Sur sa couverture, ce manuel encourage les utilisateurs à répondre à un sondage en ligne sur leurs impressions. Idéalement, une expérience plus formelle pourra même être organisée, impliquant la présence physique et l'interaction du testeur et de l'expérimentateur ainsi qu'une tâche normalisée à effectuer à l'aide de l'interface auditive, à l'issue de laquelle une série de questions pourront être posées au testeur.

Évaluation de la représentation Une fois que des retours utilisateurs préliminaires auront confirmé que cette piste de représentation mérite d'être

explorée, des expériences visant à quantifier plus précisément l'apport de l'outil pourront être réalisées. Une version modifiée de l'application sera programmée afin de permettre la collecte de données expérimentales (la trace des interactions avec l'interface et les temps de résolution de problèmes), puis cette application expérimentale pourra être employée pour tester l'efficacité de la représentation. L'expérience pourra prendre la forme d'exercices de détection de zones d'intérêt selon les différentes modalités d'observation prévues : la sonification interactive, la visualisation ou une combinaison des deux.

Améliorations générales Plus généralement, les travaux futurs pourront tenir compte des retours des utilisateurs pour améliorer progressivement les choix de représentation, l'interface de contrôle, l'optimisation des temps de calcul et le débogage de l'outil. En particulier, il a déjà été suggéré de permettre de nouvelles modalités de déplacement du point d'audition selon des trajectoires pré-programmées, ainsi que de permettre la sélection libre de régions d'intérêt au delà d'un unique point d'audition.

9.5 Cas d'étude - Sons environnementaux vs. mesures physiques

Ce cas d'étude rejoint les sujets de recherche de l'**UMR ISPA**, à l'**INRAE** de Bordeaux, pour l'étude et la modélisation des interactions au sein d'écosystèmes agricoles et forestiers. La piste privilégiée ici est une étude de l'apport de l'acoustique d'un environnement. Le chapitre 4 de la présente thèse donne un compte-rendu détaillé du travail sur ce sujet.

9.5.1 Contribution

La contribution que nous avons proposée pour ce cas d'étude consiste en la conception d'un outil informatique capable d'apprendre, à partir de données annotées, les relations sous-jacentes entre des phénomènes physiques mesurés (ici, la vitesse du vent) et les sons qui en résultent dans l'environnement étudié (ici, une forêt de pins dans le Val de l'Eyre). Il est alors prévu qu'une telle modélisation puisse, par réciprocité, entrer en jeu dans un processus de sonification en permettant de reconstituer le son d'un environnement à partir de paramètres donnés par les mesures physiques.

La principale difficulté de ce sujet réside dans le fait que le rapport temporel entre le son ambiant et les mesures physiques a peu de chances d'être direct : une mesure instantanée du vent peut provoquer (ou avoir provoqué auparavant) divers

phénomènes lors de son trajet spatial, en dehors de la fenêtre de son enregistrée au même instant.

Notre étude de l'existant porte sur les applications d'apprentissage machine pour la synthèse sonore informée (synthèse vocale à partir de texte, musique à partir de partitions, effets sonores à partir de vidéos), et suggère que les mécanismes d'attention actuellement appliqués pour des besoins de traduction et de synthèse vocale pourraient aider à la modélisation des rapports temporels non séquentiels susmentionnés.

Des captures synchronisées de mesures d'anémomètre et de sons environnementaux ont été obtenues au début de l'année 2021. Nos analyses préalables de la possibilité de corrélérer ces données se sont avérées peu concluantes jusqu'à présent. Toutefois, l'observation croisée de leurs spectres de nombres d'ondes suggèrent certaines interactions spatiales entre la structure de l'environnement (notamment la géométrie des arbres) et le son produit par le vent dans cet environnement.

9.5.2 Perspectives

Construction d'une base de données Afin de construire une base de données de bonne qualité pour cette étude, il sera tout d'abord utile d'effectuer de nouvelles captures synchronisées des mesures et du son dans des conditions mieux contrôlées.

Pistes pour la corrélation Les données collectées pourront ensuite être mises à profit pour poursuivre la recherche de corrélations entre les mesures et le son. Deux pistes qui semblent intéressantes à privilégier d'après nos études préliminaires sont les suivantes :

- Poursuivre la recherche d'un apprentissage des corrélations par un réseau de neurones profond, en s'inspirant des mécanismes d'attention déjà utilisés pour la synthèse vocale et la traduction.
- Explorer plus avant la corrélation, suggérée par l'étude des spectres de nombres d'ondes, entre la structure du milieu et la constitution du son ambiant.

9.6 Modèle générique de la sonification

9.6.1 Contribution

Réflexions sur l'exhaustivité Un critère particulièrement important pour un modèle générique de la sonification est la possibilité de réellement décrire n'importe quels choix de représentation dans un même cadre théorique. Nous avons donc

commencé par lister un ensemble de huit variables de la sonification : la source des données à représenter, l'extraction de propriétés sous-jacentes des données pour la représentation, la synthèse du son, la paramétrisation de cette synthèse, les fonctions de mappings employées, l'interaction offerte à l'utilisateur, les paramètres laissés réglables et le rapport temporel entre le déroulement du son et l'organisation des données.

Étude de l'existant Notre étude de l'existant a porté sur les tentatives de modéliser le principe de la sonification du point de vue de différentes disciplines, puis sur les outils génériques de sonification proposés au fil des années. Cette étude des différentes approches disciplinaires nous a permis de détecter un certain nombre de notions particulièrement prometteuses qui ont pu être incluses dans notre nouvelle modélisation synthétique :

- considérer la sonification comme une fonction de synthèse du son, d'après Hermann et al. [101], puis Rohrhuber [185]
- combler le manque de précision des approches liées au design qui pourrait, d'après nous, expliquer le faible taux d'adoption des techniques proposées par Barrass [15] puis Frauenberger [75]
- expliciter d'un point de vue perceptif et cognitif la progression des données d'entrées vers des paramètres perceptifs, puis les paramètres de synthèse sonore correspondants, d'après Arfib et al. [7].

Description d'un modèle En repartant des notions listées ci-dessus, nous avons décrit notre propre modélisation de la sonification en tant que fonction de synthèse sonore. Dans cette modélisation, un concepteur indique comment les composantes d'un son de sortie sont calculées en fonction d'un paramètre de temps, des données vouées à la représentation, des paramètres réglables et des données d'interaction.

Des bases d'un langage graphique ont ensuite été posées, mettant à profit la modularité suggérée par nos réflexions sur l'exhaustivité et pouvant servir de support visuel pour une description structurée selon le modèle. Un cadre théorique supplémentaire est suggéré dans cette représentation afin d'encourager les concepteurs à séparer explicitement les étapes successives de la transformation des données, d'abord en paramètres perceptifs, puis en paramètres de synthèse sonore.

Validation d'après cas d'études Nous avons ensuite effectué une vérification partielle de l'exhaustivité du modèle et en avons tiré nos premières constatations sur la souplesse du modèle en l'appliquant à quatre cas d'études. Deux sont tirés

de notre travail au cours de la thèse et deux sont des cas particuliers tirés de l'existant. A priori, la combinaison d'une description textuelle structurée et d'une représentation graphique nous ont assez bien permis de rendre compte des choix de représentations pour les quelques types d'outils étudiés.

Vers un outil dédié Enfin, nous avons rapporté un certain nombre de réflexions préliminaires quant à la possibilité de construire un outil, basé sur le modèle, pour la création de sonifications. Notamment, cela a consisté à théoriser la compilation d'une représentation graphique de la sonification en un programme qui puisse alors être exporté et intégré à d'autres logiciels. Les mêmes principes de compilation pourront, on l'espère, être mis en œuvre en temps réel pour permettre un rendu sonore directement depuis l'interface afin de faciliter l'expérimentation. Le séquenceur *ossia score*, qui comporte déjà une bonne partie des fonctionnalités souhaitées pour l'outil envisagé, nous a permis d'expérimenter sur la réécriture de solutions de sonification à partir de leurs descriptions d'après le modèle.

9.6.2 Perspectives

Expérimentation pédagogique Tout le long de la thèse, les réflexions et expérimentations sur le modèle ont été le fait d'une seule personne : l'auteur. Une technique de validation plus robuste à l'avenir pourra consister à enseigner le format de description à des personnes plus ou moins habituées à la conception de sonifications, puis à étudier leur capacité à l'apprendre, l'utiliser, relire des instances produites par d'autres personnes et reproduire les outils de sonification ainsi décrits.

Expérimenter sur l'exhaustivité La vérification de l'exhaustivité du modèle ne sera finie que quand il aura été prouvé qu'il n'est pas exhaustif. D'une façon organique au fil des prochains travaux de sonification, la démarche consistera à continuer de tester l'applicabilité du modèle, jusqu'à rencontrer des cas pour lesquels il ne fonctionne pas. Dans ces cas-là, on s'interrogera sur la possibilité de l'adapter pour palier à ces problèmes.

Pour les deux perspectives énoncées ci-dessus, le séquenceur *ossia score* pourrait continuer de servir de support pour la pratique, comme commencé dans la section 9.1.3. Le développement de cet outil étant encore largement dans une phase active, beaucoup des fonctionnalités notées comme manquantes pour l'équivalence avec le modèle (notamment la factorisation de complexes sonores) devraient y être ajoutées dans un avenir relativement proche.

Implémentation en tant qu'outil La partie la plus importante du travail à venir consistera à reprendre et approfondir les réflexions entamées dans cette thèse sur la présentation du modèle sous la forme d'une interface utilisateur, ainsi que sur la compilation du langage graphique pour la création de sonifications.

L'objectif sera ensuite de développer cet outil informatique en reprenant les principes de modularité, lisibilité et souplesse énoncés pour le modèle, tout en gardant à l'esprit les critères énoncés par la littérature pour ce type d'outil [128, 51] :

- assurer la portabilité sur différentes plateformes. On envisagera par exemple une application web, java ou python.
- permettre un contrôle bas niveau pour l'extension des fonctionnalités, soit directement par le biais des outils de personnalisation fournis, soit par la distribution libre du code source.
- permettre l'intégration d'une sonification conçue avec cet outil dans d'autres programmes pouvant par exemple gérer l'acquisition des données ou un retour visuel.
- mettre à la disposition des utilisateurs un certain nombre de fonctionnalités classiques, librement extensibles selon les besoins (par exemple des courbes auditives, un exemple de **MBSon**, un outil basique d'exploration spatiale immersive, ou une audification).

9.6.3 Conclusion

Une fois que cette modélisation de la sonification aura été confirmée comme judicieuse et concrètement utilisable, sa contribution pourra s'étendre aux différentes facettes de ce domaine très interdisciplinaire. Comme préparé en perspective à la fin de la partie III, la modélisation pourrait ainsi servir de cadre pour la conception d'un outil générique de création de sonifications. Cet outil pourrait à son tour servir de support pour faciliter l'expérimentation et le prototypage de représentations auditives, ce qui permettrait d'accélérer la recherche en perception auditive en facilitant la découverte de règles empiriques ou la confirmation expérimentale de tendances, donc la création de modèles psychoacoustiques de plus en plus complets menant à la conception d'outils de qualité pour divers usages.

À terme, l'adoption plus large de la sonification comme une technique de représentation pourrait offrir une nouvelle perspective sur l'étude de données scientifiques, qui habituerait les chercheurs à mettre en évidence d'autres types de propriétés des données lors de l'observation. La représentation auditive pourrait

aussi être plus souvent considérée comme une alternative pour faciliter l'accès des malvoyants à la recherche scientifique.

On pourra par ailleurs espérer mieux mettre à profit le potentiel humain dans les tâches d'analyse et de supervision afin de retarder un peu plus son remplacement complet par l'intelligence artificielle, dont les besoins actuels en énergie en font une technologie peu souhaitable à grande échelle dans le contexte de crise climatique dont nous sommes témoins, chanceuse génération que nous sommes.

Annexe A

Manuel utilisateur pour FluoSim-sonif

Manuel FluoSim-sonif

Maxime Poret, Octobre 2021

Adapté du travail de Johan David
et Matthieu Lagardère

Votre avis nous intéresse : rendez-vous sur [cette page](#)*
pour répondre à un rapide sondage sur votre expérience
avec cet outil.

Chargement d'une trace SPT pour la sonification

The screenshot displays the 'SPT ANALYSIS' tab of a software interface. It is divided into several sections:

- Trajectories:** Includes a file browser field with a 'Browse...' button and a red 'X' icon. Below it are 'Time Step [s] 0.020' and 'Pixel Size [µm] 0.160'.
- Movie:** Includes a file browser field with a 'Browse...' button and a red 'X' icon.
- Rendering:** Contains several controls:
 - Event Rendering Mode:** Set to 'CURRENT EVENTS'.
 - Spot Size:** Slider set to 1.000.
 - Spot Intensity:** Slider set to 2.000.
 - Trace Rendering Mode:** Set to 'CURRENT TRACES'.
 - Line Size:** Slider set to 1.000.
 - Line Intensity:** Slider set to 1.000.
 - Color Code:** Set to 'RANDOM'.
 - Look Up Table:** Set to 'Jet' with a color gradient bar.
 - LUT Min Value:** -5.0
 - LUT Max Value:** 1.0
- Filtering:** Includes sliders for 'Min Length', 'Max Length', 'Min Log(D)', and 'Max Log(D)' (set to 10.000).
- Advanced Parameters:** Includes input fields for 'D: Nb Points MSD Fit' (4), 'DInst: Nb Points Before Fit' (10), and 'DInst: Nb Points After Fit' (10).
- Export:** Includes 'Export Type' set to 'MSD Per Tracks' and an 'Export...' button.

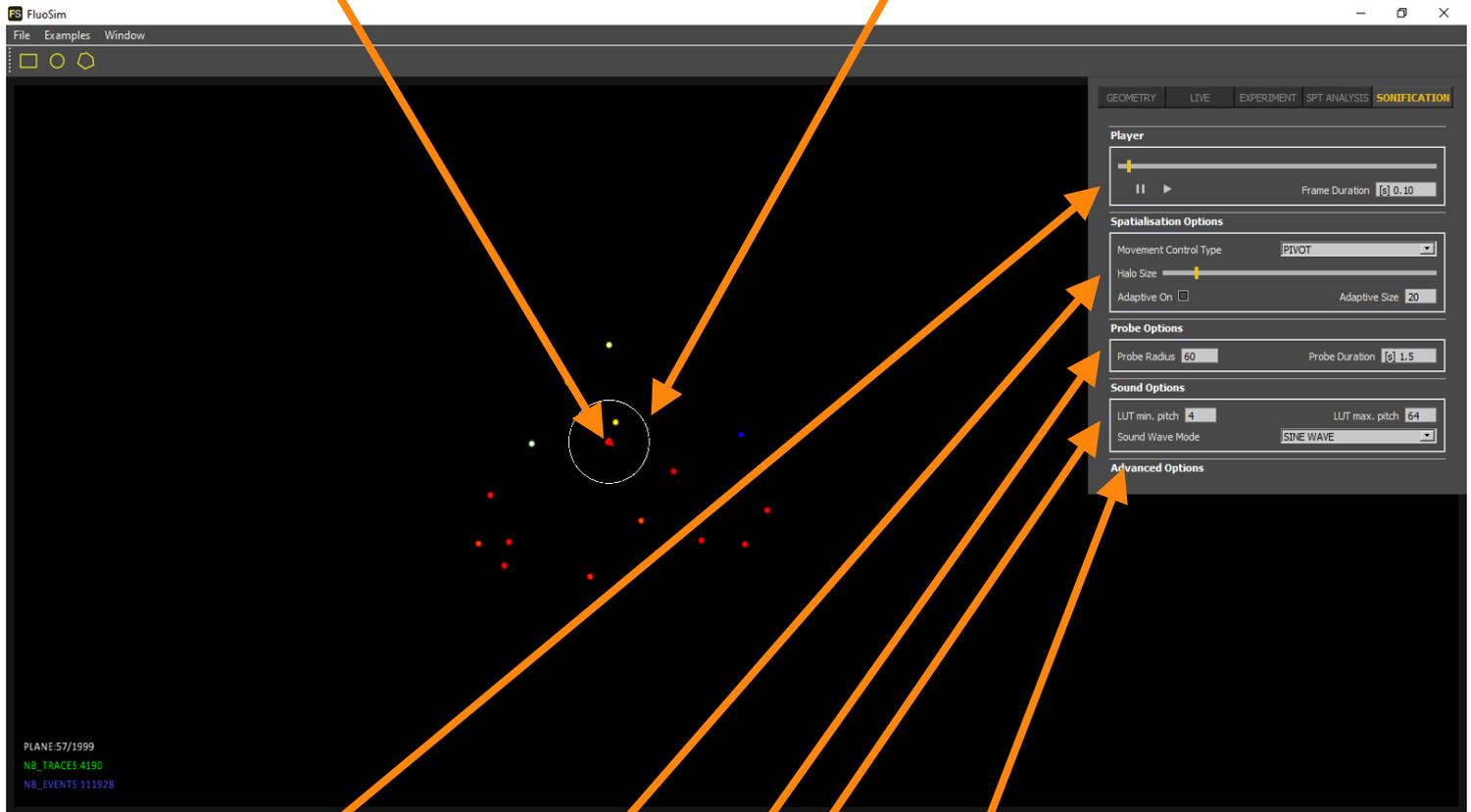
De la même façon que pour une analyse de trajectoires normale, le fichier de trajectoires est chargé depuis l'onglet "SPT Analysis"

Il est recommandé de jouer sur les paramètres d'affichage pour expliciter les positions des protéines au cours du temps (spots), plutôt que leurs trajectoires (lines).

Onglet Sonification: vue d'ensemble

Avatar de l'utilisateur. Le triangle, ou "nez", donne l'orientation de l'auditeur virtuel.

Halo d'audition centré sur l'avatar. Donne la distance à laquelle les protéines sont audibles au cours de l'exploration.



Lecteur de la trace (p.4)

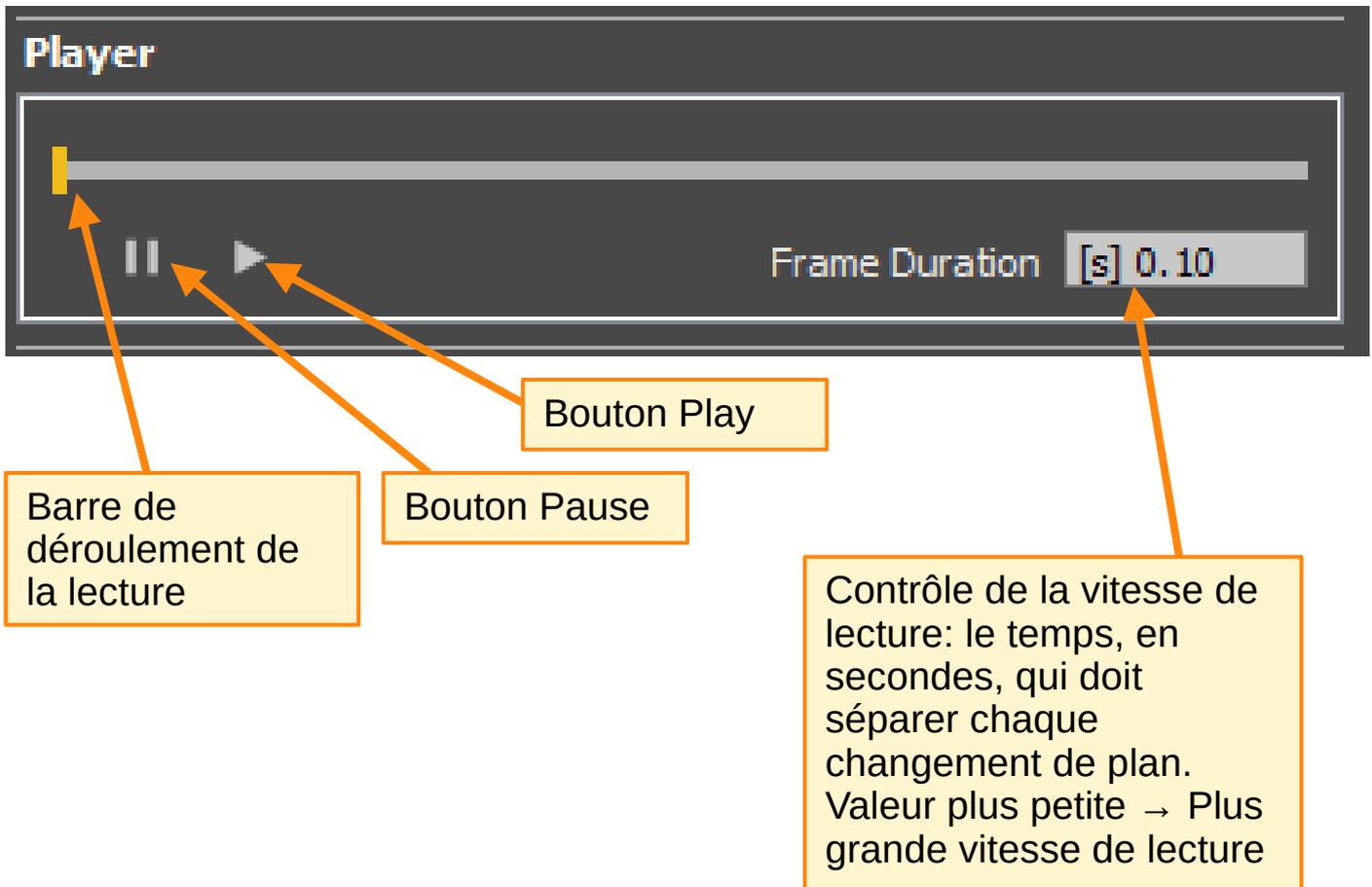
Options de spatialisation pour le mode Exploration (p.5)

Options pour le mode Sonde (p.6)

Options pour la synthèse sonore (p.7)

Options avancées (p.8)
Cliquer pour dérouler

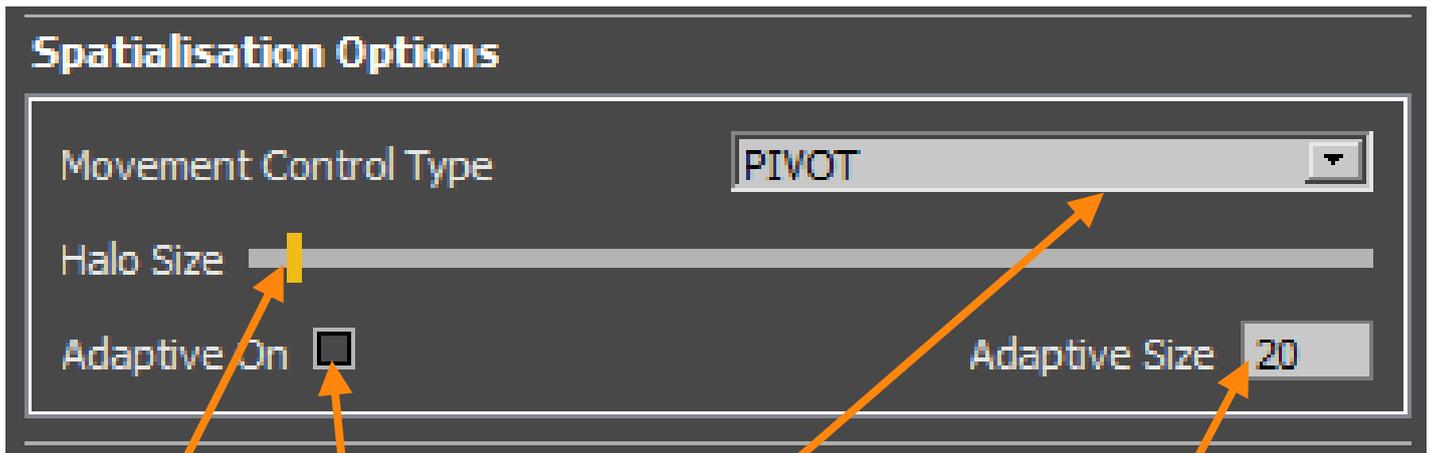
Player: Lecteur de la trace



En mode Play, la trace se déroule à la vitesse choisie par l'utilisateur et les points présents dans le halo d'audition (voir p.5) sont représentés par des sons virtuellement positionnés autour de l'avatar. C'est le mode dit "exploration".

En mode Pause, la trace reste immobile, le son est coupé, mais l'utilisateur peut lancer une sonde (voir p.6) en cliquant, avec la touche shift maintenue, à un endroit souhaité dans le tableau.

Spatialisation Options



Contrôle de la taille du halo d'audition

Sélection du type de contrôle pour le mouvement de l'avatar. Voir ci-dessous

Activation du mode de halo adaptatif, dans lequel la taille du halo est mise à jour en temps réel pour toujours inclure un nombre donné de points.

Nombre de points à inclure dans le halo adaptatif

Movement Control Type



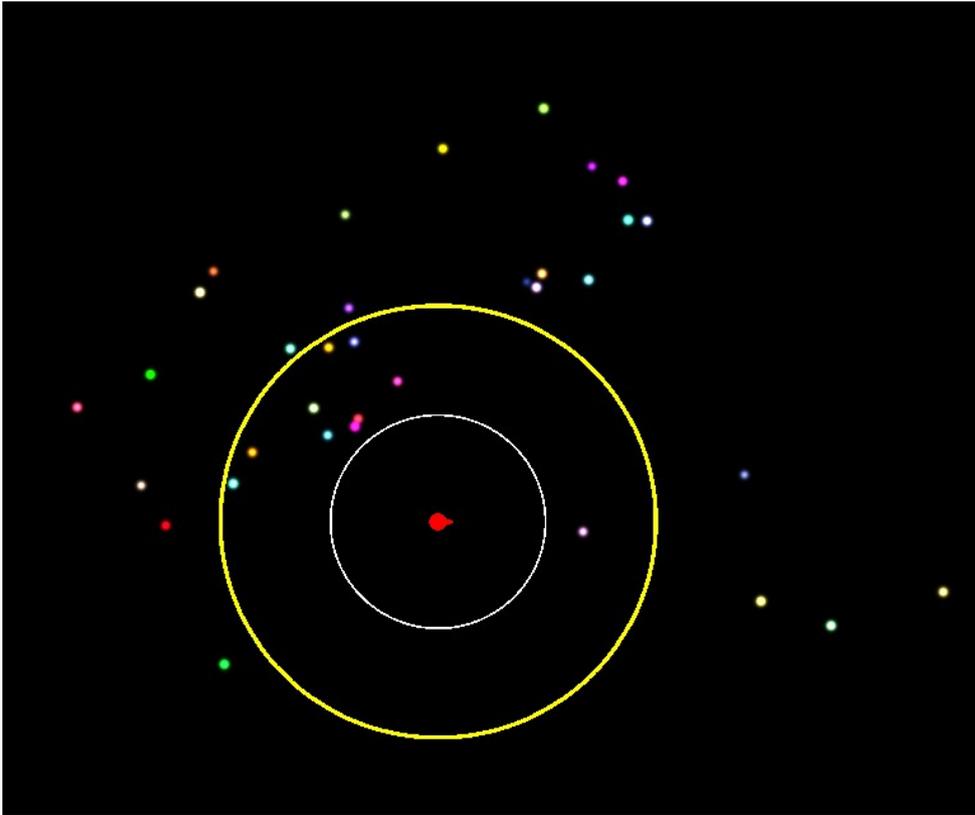
PIVOT:
Déplacement de l'avatar:
← ↑ → ↓

Orientation:
Ctrl+clic sur la position que doit viser l'avatar

TANK:
Déplacement avant/arrière de l'avatar:
↑ ↓

Orientation:
← →

Mode Sonde (probe in english)



Quand la lecture est en pause, l'utilisateur peut lancer une sonde en cliquant, avec la touche shift maintenue, à un endroit souhaité dans le tableau.

Ce point devient la nouvelle position de l'avatar, et une onde circulaire se propage de cette position jusqu'à un rayon déterminé par

l'utilisateur (voir ci-dessous) en activant les sons caractéristiques (et spatialisés) des points rencontrés au fur et à mesure de son déploiement.

Probe Options: Réglages du mode Sonde



Contrôle de la taille finale que doit atteindre la sonde

Contrôle du temps total que doit durer la sonde (par l'inverse, contrôle de la vitesse de déploiement de la sonde)

Sound Options: paramètres de la synthèse sonore

LUT min. pitch LUT max. pitch

Numéro de la note minimale pour le mapping à la diffusion instantanée d'une protéine (par défaut C₁)

Numéro de la note maximale pour le mapping (par défaut C₆)

Sound Wave Mode

SINE WAVE
SINE WAVE
SAWTOOTH WAVE
SQUARE WAVE
TRIANGLE WAVE
SYNTH FM
RANDOM
DIFFUSION SYNTH
DIFFUSION VOX

Formes d'onde basiques

Synthèse FM (voir ci-dessous)

Sélection aléatoire parmi les formes d'onde basiques

(En travaux)
Sélection du timbre en fonction de la diffusion

SYNTH FM

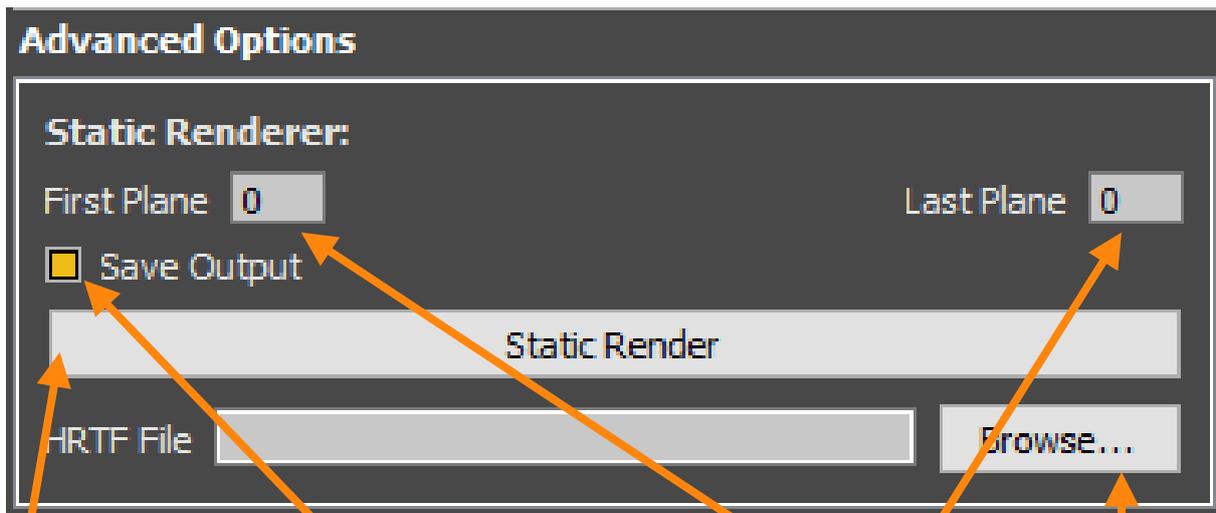
Sound Wave Mode

Modulation Frequency Modulation Index

Fréquence de modulation de la synthèse FM

Indice de modulation de la synthèse FM

Advanced Options: rendu statique et sélection de la source HRTF



Lance le rendu statique de la sonification: les sons produits par la lecture de la trace (entre un plan de début et un plan de fin) pour la position et l'orientation actuelles de l'avatar

Active/désactive la sauvegarde du rendu statique dans un fichier "outputFromStaticRender.raw"

Plans de début et de fin du rendu statique

Sélection du fichier source de Head-Related Transfer Function (HRTF) donnant le filtrage du signal, en fonction de la forme de la tête et des oreilles, qui permet de simuler la spatialisation des sources sonores.

Deux fichiers de HRTF sont actuellement disponibles dans le répertoire "HRTF/" de la release, pour des oreilles de forme "normale" ou "large".

Glossaire

audification L'audification est une technique de sonification qui consiste à transposer un signal acoustique (ou tout signal suffisamment long sur une seule dimension) dans le domaine de l'audible. 29

auditory icons « icônes auditives » en français. Les auditory icons sont une technique de sonification consistant à utiliser des sons inspirés de phénomènes naturels comme métaphores pour des événements dans un système de données. 30

complexe sonore Terme introduit par Roddy et Bridges [184]. Un complexe sonore est défini comme n'importe quel groupement perceptif constitué de plusieurs dimensions perceptives et pouvant être identifié comme formant une unité perceptive intelligible. 113

courbe auditive Les courbes auditives (*auditory graphs* en anglais), sont une application assez primitive et très répandue de la PMSon qui consistent à représenter des séries de valeurs d'une seule dimension chacune selon un axe des abscisses, le plus souvent en modulant des paramètres psychophysiques basiques comme la hauteur tonale ou la sonie. 30

earcons Les earcons sont une technique de sonification consistant à utiliser de brèves séquences musicales paramétrées par les données pour exprimer des événements dans un système. 31

hauteur tonale La hauteur tonale est la propriété d'un son harmonique (c'est-à-dire un son dont les composantes spectrales de plus forte énergie ont un diviseur commun) qui correspond à sa perception comme plus ou moins aigu, ou grave. 25

mapping (en mathématiques) Un mapping est l'association d'une valeur appartenant à un domaine vers une valeur appartenant à un autre domaine. 4

model-based sonification La model-based sonification, ou « MBSon », est une technique de sonification consistant à construire un modèle physique

sonore paramétré par les données afin de permettre leur exploration selon l'interaction de l'utilisateur. 32

parameter mapping sonification La parameter mapping sonification, ou « PMSon », est une technique de sonification qui consiste à construire un son paramétré par l'évolution de données souvent multidimensionnelles, pour une lecture du rendu selon un axe temporel. 30

représentation auditive La représentation auditive englobe tous les moyens pouvant être mis en œuvre pour communiquer des informations par le biais d'un signal sonore. 11

sonie La sonie est un paramètre perceptif lié à l'amplitude d'un signal sonore. Elle correspond à l'intensité perçue pour ce son. 25

sonification La sonification est la génération de son en fonction de données, sous réserve que la transformation soit systématique, objective et reproductible, de sorte à pouvoir être utilisée comme une méthode scientifique (traduit d'après [97]). 12

sonogramme des données De l'anglais « data sonogram ». Le sonogramme des données est une technique de model-based sonification qui consiste à associer chaque point d'un espace de données à un ressort dont les propriétés de masse, de raideur, etc, sont associées aux propriétés des données. L'interaction utilisateur sert alors à mettre les ressorts en oscillation de proche en proche selon l'extension d'un front d'onde circulaire. 168

spatialisation La spatialisation (sonore) consiste à simuler le positionnement de sources sonores autour d'un auditeur. 26

spectre Le spectre (fréquentiel) est la représentation d'un signal dans le domaine fréquentiel, c'est-à-dire la représentation selon un axe des fréquences de l'amplitude et de la phase des différentes composantes de ce signal. 20

spectrogramme (ou sonogramme) Une adaptation temporelle du principe du spectre. Au lieu d'effectuer l'analyse du signal entier, on le découpe en fenêtres d'un nombre relativement petit d'échantillons qui sont analysés individuellement. La représentation est à nouveau donnée selon un axe temporel. 21

timbre Le timbre est la propriété perceptive qui permet de différencier à l'écoute deux sons dont la hauteur et la sonie sont identiques. 26

échantillonnage L'échantillonnage consiste à mesurer un plus ou moins grand nombre de points du signal sonore au cours du temps. Cette quantité est

donnée par une grandeur appelée la fréquence d'échantillonnage, exprimée en Hertz (Hz), qui est le nombre de point d'échantillons constituant une seconde de signal. [16](#)

Acronyms

- HC* Hauteur du Cordon. 40, 41, 43–48, 61, 63, 150, 151
- HP* Hauteur de la Pièce. 40, 41, 44–49, 61–63, 67, 150, 151
- HP₀* Hauteur théorique de la Pièce. 40, 150, 151
- LC* Largeur du Cordon. 40, 41, 43–48, 61, 63, 150, 151
- TC* Température du Cordon. 40, 41, 45–48, 61–63, 67, 150–152, 159
- TP* Température de la Pièce. 40, 41, 43, 45–48, 61–63, 67, 150, 151, 159, 191
- API** Application Programming Interface. 22, 81
- CD** Compact Disc. 16, 17
- CIAC** Centre Interactif d'Art Cognitif. 72, 76
- CNRS** Centre National de la Recherche Scientifique. 72
- CSS** Cascading Style Sheets. 50
- CSV** Comma-Separated Values. 41, 67, 151, 167, 179, 189
- DSP** Digital Signal Processor. 22, 23, 85, 189, 191
- EcoFun** Relations hydriques et fonctionnement des écosystèmes. 91, 92
- ENSC** École Nationale Supérieure de Cognitique. 102
- ENSEIRB-MATMECA** École Nationale Supérieure d'Électronique, Informatique, Télécommunications, Mathématique et Mécanique. 75, 102
- ESTIA** École Supérieure des Technologies Industrielles Avancées. 37, 58, 199, 201
- GAN** Generative Adversarial Network. 95
- HRTF** Head-Related Transfer Function. 26, 85, 118, 140, 160, 166
- HTML** HyperText Markup Language. 50, 55, 111
- IA** Intelligence Artificielle. 95, 97, 106

- ICAD** International Conference for Auditory Display. 5, 12, 23
- ICMC** International Computer Music Conference. 5
- IDE** Integrated Development Environment. 179
- IINS** Institut Interdisciplinaire de NeuroSciences. 71, 72, 86, 202
- INRAE** Institut National de Recherche pour l’Agriculture, l’alimentation et l’Environnement. 91, 203
- ISPA** Interaction Sol Plante Atmosphère. 91, 92, 94, 203
- MBSon** Model-Based Sonification. 32, 78, 109, 207
- ME** Mécanique Environnementale. 91, 92
- MFCC** Mel Frequency Cepstral Coefficients. 104
- MIDI** Musical Instrument Digital Interface. 22–24, 26, 76, 96, 113, 114, 129, 143, 145, 146, 154, 162, 176, 189, 197
- MSD** Mean Squared Displacement. 73
- OSC** Open Sound Control. 22–24, 129, 138, 176, 189
- PALM** PhotoActivated Localization Microscopy. 72
- PHP** PHP Hypertext Preprocessor. 50, 55
- PMSon** Parameter Mapping Sonification. 23, 30, 109, 132, 149
- SCRIME** Studio de Création et de Recherche en Informatique pour les Musiques Expérimentales. 42, 72, 81, 97, 100
- SMC** Sound and Music Computing conference. 5
- SPT** Single-Particle Tracking. 73–75, 77, 83
- TMBASS** Turning Molecular Behavior into Architecture Shape Sound. 72, 76
- UEQ-S** User Experience Questionnaire Short. 86
- UMR** Unité Mixte de Recherche. 91, 203

Références

- [1] M. Adcock and S. Barrass, “Cultivating Design Patterns for Auditory Display,” in *Proceedings of the 10th International Conference on Auditory Display (ICAD2004)*, Sydney, Australie, Jul. 2004.
- [2] A. L. Aldana Blanco, S. Grautoff, and T. Hermann, “ECG Sonification to Support the Diagnosis and Monitoring of Myocardial Infarction,” *Journal on Multimodal User Interfaces*, vol. 14, no. 2, pp. 207–218, 2020.
- [3] E. Amsen. (2019) What Does Climate Change Sound Like As Music? [Online]. Available : <https://www.forbes.com/sites/evaamsen/2019/04/18/what-does-climate-change-sound-like-as-music/>
- [4] A. Andreopoulou and N. Kotsani, “Polyfōnía : an Interactive Vocal Processing System using Finger Tracking,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6798269>
- [5] I. Angeletti, G. Colliou–Clourenec, J. Arnault-Macabiau, A. Riveiro, A. Patrone, J. Marzin, and M. El Moumni, “Musique Moléculaire,” *Rapport de Projet au Fil de l’Année, 2A Informatique, ENSEIRB-MATMECA*, 2015.
- [6] J. Ao, R. Wang, L. Zhou, S. Liu, S. Ren, Y. Wu, T. Ko, Q. Li, Y. Zhang, Z. Wei *et al.*, “SpeechT5 : Unified-modal encoder-decoder pre-training for spoken language processing,” *arXiv preprint arXiv :2110.07205*, 2021.
- [7] D. Arfib, J.-M. Couturier, L. Kessous, and V. Verfaillie, “Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces,” *Organised Sound*, vol. 7, pp. 127–144, Aug. 2003.
- [8] S. Ö. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, A. Ng, J. Raiman *et al.*, “Deep voice : Real-time neural text-to-speech,” in *International Conference on Machine Learning*. PMLR, 2017, pp. 195–204.
- [9] M. Barra, T. Cillo, A. De Santis, U. F. Petrillo, A. Negro, and V. Scarano, “Multimodal Monitoring of Web Servers,” *IEEE MultiMedia*, vol. 9, no. 3, pp. 32–41, 2002.

-
- [10] M. Barra, T. Cillo, A. de Santis, U. F. Petrillo, A. Negro, V. Scarano, T. Matlock, and P. P. Maglio, “Personal webmelody : Customized sonification of web servers,” in *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, Espoo, Finland, Jul. 2001.
- [11] S. Barrass, “Personify : A Toolkit for Perceptually Meaningful Sonification,” *ACMA*, vol. 95, 1995.
- [12] —, “EarBenders : Using stories about listening to design auditory interfaces,” in *Proceedings of the First Asia-Pacific Conference on Human Computer Interaction APCHI*, vol. 96, 1996.
- [13] —, “Sculpting a sound space with information properties,” *Organised Sound*, vol. 1, no. 2, pp. 125–136, 1996.
- [14] —, “TaDa ! Demonstrations of auditory information design,” in *Proceedings of the 3rd International Conference on Auditory Display (ICAD1996)*, Palo Alto, California, Nov. 1996.
- [15] —, “Sonification design patterns,” in *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, Boston, MA, Jul. 2003, pp. 170–175.
- [16] O. Ben-Tal, J. Berger, B. Cook, G. Scavone, M. Daniels, and P. Cook, “SONART : the sonification application research toolbox,” in *Proceedings of the eighth international conference on auditory display*, Kyoto, Japan, Jul. 2002, pp. 2–5.
- [17] J. Bernstein, “Telephonische wahrnehmung der schwankungen des muskelstromes bei der kontraktion : Sitzung am 28. mai,” *Bericht über die Sitzungen der Naturforschenden Gesellschaft zu Halle*, 1881.
- [18] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, “Imaging Intracellular Fluorescent Proteins at Nanometer Resolution,” *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006. [Online]. Available : <https://www.science.org/doi/abs/10.1126/science.1127344>
- [19] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, “Earcons and icons : Their structure and common design principles,” *Human-Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.
- [20] G. Blum, M. d. Blois, N. Tadjine *et al.*, *L’impression 3D : de l’émerveillement technique aux enjeux organisationnels, économiques et sociétaux*, 2017.
- [21] S. Bly, “Presenting information in sound,” in *Proceedings of the 1982 conference on Human factors in computing systems*, 1982, pp. 371–375.
- [22] —, *Sound and computer information presentation*. University of California, Davis, 1982.

- [23] T. L. Bonebright and J. H. Flowers, “Evaluation of auditory display,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Berlin, 2011, pp. 111–144.
- [24] T. Bovermann, T. Hermann, and H. Ritter, “Tangible Data Scanning Sonification Model,” pp. 77–82, Jun. 2006.
- [25] A. S. Bregman, *Auditory scene analysis : The perceptual organization of sound*. MIT press, 1994.
- [26] R. Bresin, A. Hunt, and H. Thomas, “Human Interaction with Auditory Displays,” in *Explorations in Sonic Interaction Design*, D. Rocchesso, Ed. Berlin, Germany : Logos Publishing House, 2011, ch. 5, pp. 20–21.
- [27] S. A. Brewster, P. C. Wright, and A. D. Edwards, “A detailed investigation into the effectiveness of earcons,” in *Santa Fe Institute studies in the sciences of complexity—proceedings*, vol. 18. Addison-Wesley Publishing CO, 1994.
- [28] J. Brooke, “SUS : A quick and dirty usability scale,” *Usability Eval. Ind.*, vol. 189, Nov. 1995.
- [29] D. S. Brungart and B. D. Simpson, “Design, validation, and in-flight evaluation of an auditory attitude indicator based on pilot-selected music.” Paris, France : International Community for Auditory Display, Jun. 2008.
- [30] M. Bujacz, P. Skulimowski, and P. Strumillo, “Sonification of 3D Scenes Using Personalized Spatial Audio to Aid Visually Impaired Persons,” in *Proceedings of the 17th International Conference on Auditory Display (ICAD2011)*, Budapest, Hungary, Jun. 2011.
- [31] J. Calandra, J.-M. Chouvel, and M. Desainte-Catherine, “Multi-Scale Oracle and Automated Representation of Formal Diagrams Based on the Cognitive Algorithm,” in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR’20/21*, R. Gottfried, G. Hajdu, J. Sello, A. Anatrini, and J. MacCallum, Eds. Hamburg, Germany : Hamburg University for Music and Theater, 2021, pp. 240–250.
- [32] R. Candey, A. Schertenleib, and W. Diaz Merced, “xSonify sonification tool for space physics,” in *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, London, UK, Jun. 2006, pp. 289–290.
- [33] S. Carlile, “Psychoacoustics,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany : Logos Publishing House, 2011, ch. 3, pp. 41—61.
- [34] A. Chabot, “Méthodologie de monitoring multiphysique des procédés DED : développement par une démarche expérimentale,” Theses, École centrale de Nantes, Oct. 2020. [Online]. Available : <https://tel.archives-ouvertes.fr/tel-03164132>

-
- [35] A. Chabot, M. Rauch, and J.-Y. Hascoët, “Novel control model of Contact-Tip-to-Work Distance (CTWD) for sound monitoring of arc-based DED processes based on spectral analysis,” *The International Journal of Advanced Manufacturing Technology*, vol. 116, no. 11, pp. 3463–3472, 2021.
- [36] C. Chafe, “Browser-Based Sonification,” in *Proceedings of the 17th Linux Audio Conference (LAC-19)*, 2019.
- [37] A. Chazeau, M. Garcia, K. Czöndör, D. Perrais, B. Tessier, G. Giannone, and O. Thoumine, “Mechanical coupling between transsynaptic n-cadherin adhesions and actin flow stabilizes dendritic spines,” *Molecular biology of the cell*, vol. 26, no. 5, pp. 859–873, 2015.
- [38] A. Chazeau, A. Mehidi, D. Nair, J. J. Gautier, C. Leduc, I. Chamma, F. Kage, A. Kechkar, O. Thoumine, K. Rottner *et al.*, “Nanoscale segregation of actin nucleation and elongation factors determines dendritic spine protrusion,” *The EMBO journal*, vol. 33, no. 23, pp. 2745–2764, 2014.
- [39] J. Chen, X. Tan, Y. Leng, J. Xu, G. Wen, T. Qin, and T.-Y. Liu, “Speech-T : Transducer for Text to Speech and Beyond,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 6621–6633, 2021.
- [40] E. C. Cherry, “Some experiments on the recognition of speech, with one and with two ears,” *The Journal of the acoustical society of America*, vol. 25, no. 5, pp. 975–979, 1953.
- [41] J. Cherston and J. A. Paradiso, “Rotator : Flexible Distribution of Data Across Sensory Channels,” in *Proceedings of the International Conference on Auditory Display (ICAD) 2017*, Pennsylvania State University, USA, Jun. 2017, pp. 86–93.
- [42] E. Childs, “Achorripsis : A sonification of probability distributions,” in *Proceedings of the 8th International Conference on Auditory Display (ICAD 2002)*, Kyoto, Japan, Jul. 2002.
- [43] J. M. Chowning, “The synthesis of complex audio spectra by means of frequency modulation,” *Journal of the audio engineering society*, vol. 21, no. 7, pp. 526–534, 1973.
- [44] J. Cohen, “"Kirk Here" : Using Genre Sounds to Monitor Background Activity,” in *INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems*, ser. CHI '93. New York, NY, USA : Association for Computing Machinery, 1993, p. 63–64. [Online]. Available : <https://doi.org/10.1145/259964.260073>
- [45] P. R. Cook. (2012) Sonification in ChuckK, MiniAudicle and STK. [Online]. Available : <https://www.cs.princeton.edu/~prc/ChuckKU/SonifyingInChuckKICAD2012.pdf>

- [46] M. Daniel, “Shape-changing Cylindrical Displays : Application to Data Physicalization and Peripheral Interaction for Energy Demand-Side Management,” Theses, Univ. Bordeaux, Nov. 2018. [Online]. Available : <https://hal.archives-ouvertes.fr/tel-01972386>
- [47] J. David, “Exploration Sonore Interactive de l’Activité Cellulaire,” *Rapport de Stage pour Master Informatique, Université de Bordeaux*, 2021.
- [48] B. K. Davison and B. N. Walker, “AudioPlusWidgets : Bringing Sound to Software Widgets and Interface Components,” in *Proceedings of the 14th International Conference on Auditory Display (ICAD2008)*, Paris, France, Jun. 2008.
- [49] A. de Campo, “A data sonification design space map,” in *Proc. of the 2nd International Workshop on Interactive Sonification, York, UK*, 2007.
- [50] A. de Campo, C. Frauenberger, and R. Hoeldrich, “Sonenvir – a Progress Report,” in *ICMC*. Citeseer, 2005.
- [51] A. de Campo, C. Frauenberger, and R. Höldrich, “Designing a generalized sonification environment,” in *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia, Jul. 2004.
- [52] A. de Campo, J. Rohrhuber, T. Bovermann, and C. Frauenberger, *Sonification and Auditory Display in SuperCollider*, S. Wilson, D. Cottle, and N. Collins, Eds. The MIT Press, 2011.
- [53] O. M. de Météorologie. (2010) New world record wind gust. [Online]. Available : <https://public.wmo.int/en/media/news/new-world-record-wind-gust>
- [54] L. Demany and C. Semal, “Article “Son”,” in *Dictionnaire de Psychologie*, R. Doron and F. Parot, Eds. Paris : Presses Universitaires de France, pp. 640–641.
- [55] E. Dimogerontakis, “Spatialized polyphonic granular,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6800822>
- [56] F. Dombois, “Using audification in planetary seismology,” in *International Conference on Auditory Display (ICAD 2001)*, Espoo, Finland, Jul. 2001.
- [57] —, “Auditory seismology : On free oscillations, focal mechanisms, explosions and synthetic seismograms,” in *International Conference on Auditory Display (ICAD 2002)*, Kyoto, Japan, Jul. 2002.
- [58] F. Dombois, O. Brodewolf, O. Friedli, I. Rennert, and T. Koenig, “SonifYer – A Concept, a Software, a Platform,” in *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, Jun. 2008.

-
- [59] S. Drake, “The role of music in Galileo’s experiments,” *Scientific American*, vol. 232, no. 6, pp. 98–105, 1975.
- [60] G. Dubus and R. Bresin, “A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities,” *PLOS ONE*, vol. 8, no. 12, pp. 1–28, Dec. 2013. [Online]. Available : <https://doi.org/10.1371/journal.pone.0082491>
- [61] S. Dupont and E. G. Patton, “Momentum and scalar transport within a vegetation canopy following atmospheric stability and seasonal canopy changes : the CHATS experiment,” *Atmospheric Chemistry and Physics*, vol. 12, no. 13, pp. 5913–5935, 2012. [Online]. Available : <https://acp.copernicus.org/articles/12/5913/2012/>
- [62] S. Dupont, P. Défossez, J.-M. Bonnefond, M. R. Irvine, and D. Garrigou, “How stand tree motion impacts wind dynamics during windstorms,” *Agricultural and Forest Meteorology*, vol. 262, pp. 42–58, 2018. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0168192318302132>
- [63] M. Eddy, “Diagrams,” in *A Companion to the History of Information*, A. B. Anthony Grafton and A. S. Goeing, Eds. Princeton : Princeton University Press, 05 2021, pp. 397–401.
- [64] A. D. N. Edwards, “Auditory Display in Assistive Technology,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany : Logos Publishing House, 2011, ch. 17, pp. 431—453.
- [65] J. Edworthy, S. Loxley, and I. Dennis, “Improving auditory warning design : Relationship between warning sound parameters and perceived urgency,” *Human factors*, vol. 33, no. 2, pp. 205–231, 1991.
- [66] S. K. Everton, M. Hirsch, P. Stravroulakis, R. K. Leach, and A. T. Clare, “Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing,” *Materials & Design*, vol. 95, pp. 431–445, 2016. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0264127516300995>
- [67] J. Ferguson and S. Brewster, “Evaluating the magnitude estimation approach for designing sonification mapping topologies,” in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Northumbria University, Newcastle upon Tyne, UK, Jun. 2019.
- [68] M. Fernstrom and C. McNamara, “After direct manipulation - direct sonification,” in *Proceedings of the 5th International Conference on Auditory Display (ICAD1998)*, Glasgow, UK, Nov. 1998.

- [69] H. Fletcher and W. Munson, “Loudness, Its Definition, Measurement and Calculation,” *The Journal of the Acoustical Society of America*, vol. 5, no. 2, pp. 82–108, 1933.
- [70] J. H. Flowers, “Thirteen years of reflection on auditory graphing : Promises, pitfalls, and potential new directions.” Georgia Institute of Technology, 2005.
- [71] J. Fox, J. Carlile, and J. Berger, “Sonimime : Sonification of fine motor skills,” in *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, Limerick, Ireland, Jul. 2005.
- [72] G. Frantti and L. Levereault, “Auditory discrimination of seismic signals from earthquakes and explosions,” *Bulletin of the seismological society of America*, vol. 55, no. 1, pp. 1–25, 1965.
- [73] C. Frauenberger, “Auditory Display Design : An Investigation of a Design Pattern Approach,” Ph.D. dissertation, Queen Mary University of London, 2009.
- [74] C. Frauenberger, T. Stockman, and M.-L. Bourguet, “Pattern Design in the Context Space : a Methodological Framework for Auditory Display Design,” in *Proceedings of the 13th International Conference on Auditory Display*, Montréal, Canada, Jun. 2007, pp. 513–518.
- [75] —, “Pattern design in the context space : paco – a methodological framework for designing auditory display with patterns,” in *Proceedings of the 14th Conference on Pattern Languages of Programs*, 2007.
- [76] F. Ganis, S. Serafin, and M. Vatti, “Tickle Tuner - Haptic Smartphone Cover for Cochlear Implant Users’ Musical Training,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6800792>
- [77] M. Garcia, C. Leduc, M. Lagardère, A. Argento, J.-B. Sibarita, and O. Thoumine, “Two-tiered coupling between flowing actin and immobilized n-cadherin/catenin complexes in neuronal growth cones,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 22, pp. 6997–7002, 2015. [Online]. Available : <https://www.pnas.org/doi/abs/10.1073/pnas.1423455112>
- [78] L. S. Gardiner, C. Deser, M. Quinn, B. Hatheway, S. Clark, T. Scheitlin, M. Rehme, and A. Phillips, “Sounding Climate : An Exhibit Showcasing Data Sonification and Visualization from the Community Earth System Model,” in *AGU Fall Meeting Abstracts*, vol. 2018, Dec. 2018, pp. PA41B–03.
- [79] W. W. Gaver, “The sonicfinder : An interface that uses auditory icons,” *Human-Computer Interaction*, vol. 4, no. 1, pp. 67–94, 1989.

-
- [80] —, “Auditory interfaces,” in *Handbook of human-computer interaction*. Elsevier, 1997, pp. 1003–1041.
- [81] W. W. Gaver and D. A. Norman, “Everyday listening and auditory icons,” Ph.D. dissertation, University of California, San Diego, Department of Cognitive Science and Psychology, 1988.
- [82] W. W. Gaver, R. B. Smith, and T. O’Shea, “Effective sounds in complex systems : The ARKola simulation,” in *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, 1991, pp. 85–90.
- [83] G. Giannone, B. J. Dubin-Thaler, H.-G. Döbereiner, N. Kieffer, A. R. Bresnick, and M. P. Sheetz, “Periodic lamellipodial contractions correlate with rearward actin waves,” *Cell*, vol. 116, no. 3, pp. 431–443, 2004.
- [84] G. Giannone, B. J. Dubin-Thaler, O. Rossier, Y. Cai, O. Chaga, G. Jiang, W. Beaver, H.-G. Döbereiner, Y. Freund, G. Borisy *et al.*, “Lamellipodial actin mechanically links myosin activity with adhesion-site formation,” *Cell*, vol. 128, no. 3, pp. 561–575, 2007.
- [85] G. Giannone, G. Jiang, D. H. Sutton, D. R. Critchley, and M. P. Sheetz, “Talin1 is critical for force-dependent reinforcement of initial integrin–cytoskeleton bonds but not tyrosine kinase activation,” *Journal of Cell Biology*, vol. 163, no. 2, pp. 409–419, Oct. 2003. [Online]. Available : <https://doi.org/10.1083/jcb.200302001>
- [86] M. Gilfix and A. L. Couch, “Peep (The Network Auralizer) : Monitoring Your Network with Sound,” in *Proceedings of the 14th USENIX Conference on System Administration (LISA’00)*. Berkeley, CA, USA : USENIX Association, 2000, pp. 109–118.
- [87] G. Grinstein, R. Pickett, and M. G. Williams, “EXVIS : An exploratory visualization environment,” in *Graphics Interface*, vol. 89, 1989, pp. 254–261.
- [88] F. Halbig. (2009) ANTARKTIKA – a climatic time-travel. [Online]. Available : <http://www.antarktika.at/>
- [89] R. Hamilton, “Sonifying Game-Space Choreographies With UDKOSC,” in *NIME*, 2013, pp. 446–449.
- [90] Y. C. Han and B.-j. Han, “Skin Pattern Sonification Using NMF-based Visual Feature Extraction and Learning-based PMSon.” Georgia Institute of Technology, 2014.
- [91] J. D. Harris, “Pitch discrimination,” *The Journal of the Acoustical Society of America*, vol. 24, no. 6, pp. 750–755, 1952. [Online]. Available : <https://doi.org/10.1121/1.1906970>

- [92] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research,” in *Human Mental Workload*, ser. Advances in Psychology, P. A. Hancock and N. Meshkati, Eds. North-Holland, 1988, vol. 52, pp. 139–183. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0166411508623869>
- [93] D. Hausen, A. Tabard, A. Thermann, K. Holzner, and A. Butz, “Evaluating Peripheral Interaction,” Munich Germany, Feb. 2014, pp. 21–28.
- [94] C. Hawthorne, A. Stasyuk, A. Roberts, I. Simon, C.-Z. A. Huang, S. Dieleman, E. Elsen, J. Engel, and D. Eck, “Enabling factorized piano music modeling and generation with the MAESTRO dataset,” *arXiv preprint arXiv :1810.12247*, 2018.
- [95] C. Hayward, “Listening to the earth sing,” in *Auditory Display. Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed. Reading : Addison-Wesley, 1994, p. 369–404.
- [96] T. Hermann, “Sonification for Exploratory Data Analysis,” Ph.D. dissertation, 2002. [Online]. Available : <https://nbn-resolving.org/urn:nbn:de:0070-pub-20172633>, <https://pub.uni-bielefeld.de/record/2017263>
- [97] —, “Taxonomy and definitions for sonification and auditory display.” Paris, France : International Community for Auditory Display, Jun. 2008.
- [98] —, “Model-Based Sonification,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany : Logos Publishing House, 2011, ch. 16, pp. 399–427.
- [99] —. (2011) Sonification – A Definition. [Online]. Available : <https://sonification.de/son/definition/>
- [100] T. Hermann, G. Baier, U. Stephani, and H. Ritter, “Vocal Sonification of Pathological EEG Features,” in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, Jun. 2006, pp. 158–163.
- [101] T. Hermann, K. Bunte, and H. Ritter, “Relevance-based interactive optimization of sonification,” in *Proceedings of the 13th International Conference on Auditory Display*, Montréal, Canada, Jun. 2007, pp. 461–467.
- [102] T. Hermann, T. Hildebrandt, P. Langeslag, and S. Rinderle-Ma, “Optimizing Aesthetics and Precision in Sonification for Peripheral Process Monitoring,” in *Proceedings of the 21st International Conference for Auditory Display (ICAD 2015)*, Graz, Austria, Jul. 2015, pp. 317–318.
- [103] T. Hermann and A. Hunt, “The importance of interaction in sonification,” in *Proceedings of ICAD 04. Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia, Jul. 2004.

-
- [104] T. Hermann, A. Hunt, and J. G. Neuhoff, “Introduction,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany : Logos Publishing House, 2011, ch. 1, pp. 1—6.
- [105] T. Hermann, J. Krause, and H. Ritter, “Real-time Control of Sonification Models with a Haptic Interface,” in *Proceedings of the 8th international conference on auditory display*, Kyoto, Japan, Jul. 2002.
- [106] T. Hermann, P. Meinicke, and H. Ritter, “Principal Curve Sonification,” in *Proceedings of the International Conference on Auditory Display*, Atlanta, GA, Apr. 2000.
- [107] T. Hermann and H. Ritter, “Listen to your Data : Model-Based Sonification for Data Analysis,” in *Advances in Intelligent Computing and Multimedia Systems*. Windsor, Ontario : Int. Inst. for Advanced Studies in System research and cybernetics : G.E. Lasker and M.R. Syed, eds., 1999, pp. 189–194.
- [108] M. Heymann and M. Hansen, “A new set of sound commands for R; sonification of the HMC algorithm,” *ASA Proceedings, Statistical Computing Section*, pp. 1439–1443, 2002.
- [109] T. Hildebrandt, T. Hermann, and S. Rinderle-Ma, “A Sonification System for Process Monitoring as Secondary Task,” in *5th IEEE Conference on Cognitive Infocommunications (CogInfoCom 2014)*. Vietri sul Mare, Italy : IEEE, 2014, pp. 191–196.
- [110] —, “Continuous Sonification Enhances Adequacy of Interactions in Peripheral Process Monitoring,” *International Journal of Human-Computer Studies*, vol. 95, pp. 54–65, 2016. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S107158191630074X>
- [111] B. Hilgren, “The music of science : Environmentalist data sonifications, interdisciplinary art, and the narrative of climate change,” Ph.D. dissertation, The Florida State University, 2019.
- [112] C. Ho and C. Spence, “Using Peripersonal Warning Signals to Orient a Driver’s Gaze,” *Human Factors*, vol. 51, no. 4, pp. 539–556, 2009, pMID : 19899363. [Online]. Available : <https://doi.org/10.1177/0018720809341735>
- [113] S. E. Hudson and I. Smith, “Electronic Mail Previews Using Non-Speech Audio,” in *Conference Companion on Human Factors in Computing Systems (CHI96)*, Vancouver, British Columbia, Canada, 1996, pp. 237–238.
- [114] O. Höner, “Aiding Movement with Sonification in “Exercise, Play and Sport”,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany : Logos Publishing House, 2011, ch. 21, pp. 525—554.

- [115] S. Ibarboure, “Perception d’information par des retours vibrotactiles pour une mise en œuvre plus flexible des procédés robotisés de fabrication additive,” Theses, Université de Bordeaux, Jun. 2021. [Online]. Available : <https://tel.archives-ouvertes.fr/tel-03361862>
- [116] V. Isnard, “L’efficacité du système auditif humain pour la reconnaissance de sons naturels,” Theses, Université Pierre et Marie Curie - Paris VI, Nov. 2016. [Online]. Available : <https://hal.archives-ouvertes.fr/tel-01444576>
- [117] V. Isnard, M. Taffou, I. Viaud-Delmon, and C. Suied, “Auditory Sketches : Very Sparse Representations of Sounds Are Still Recognizable,” *PLOS ONE*, vol. 11, no. 3, pp. 1–15, Mar. 2016. [Online]. Available : <https://doi.org/10.1371/journal.pone.0150313>
- [118] G. Jiang, G. Giannone, D. R. Critchley, E. Fukumoto, and M. P. Sheetz, “Two-piconewton slip bond between fibronectin and the cytoskeleton depends on talin,” *Nature*, vol. 424, no. 6946, pp. 334–337, 2003.
- [119] M. R. Jones, “Time, our lost dimension : toward a new theory of perception, attention, and memory,” *Psychological review*, vol. 83, no. 5, pp. 323–355, 1976.
- [120] M. R. Jones and M. Boltz, “Dynamic attending and responses to time,” *Psychological review*, vol. 96, no. 3, pp. 459–491, 1989.
- [121] R. Jung, “Ambience for Auditory Displays : Embedded Musical Instruments as Peripheral Audio Cues,” in *Proceedings of the International Conference on Auditory Display (ICAD 2008)*. Paris, France : International Community for Auditory Display, Jun. 2008.
- [122] M. Kaltenbrunner, “Y-windows : Proposal for a standard AUI environment,” in *Proceedings of the 8th International Conference on Auditory Display (ICAD)*, Kyoto, Japan, Jul. 2002.
- [123] A. Kapur, G. Tzanetakis, N. Virji-Babul, G. Wang, and P. R. Cook, “A framework for sonification of vicon motion capture data,” in *Conference on Digital Audio Effects*, 2005, pp. 47–52.
- [124] F. Kilander and P. Lönnqvist, “A Whisper in the Woods - An Ambient Soundscape for Peripheral Awareness of Remote Processes,” in *Proceedings of the 8th International Conference on Auditory Display (ICAD 2002)*, Kyoto, Japan, Jul. 2002.
- [125] J. W. Kim, R. Bittner, A. Kumar, and J. P. Bello, “Neural music synthesis for flexible timbre control,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 176–180.

-
- [126] J. Kirakowski and M. Corbett, “SUMI : the Software Usability Measurement Inventory,” *British Journal of Educational Technology*, vol. 24, no. 3, pp. 210–212, Sep. 1993. [Online]. Available : <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.1993.tb00076.x>
- [127] J. Kong, J. Kim, and J. Bae, “Hifi-gan : Generative adversarial networks for efficient and high fidelity speech synthesis,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 17 022–17 033, 2020.
- [128] G. Kramer, B. N. Walker, T. Bonebright, P. Cook, J. H. Flowers, and N. Miner, “The Sonification Report : Status of the Field and Research Agenda. Report prepared for the National Science Foundation by Members of the International Community for Auditory Display,” in *International Community for Auditory Display (ICAD)*, Santa Fe, NM, 1999.
- [129] M. Lagardère, I. Chamma, E. Bouilhol, M. Nikolski, and O. Thoumine, “FluoSim : Simulator of single molecule dynamics for fluorescence live-cell and super-resolution imaging of membrane proteins,” *Scientific Reports*, vol. 10, no. 1, p. 19954, Nov. 2020.
- [130] S. Langlois, C. Suied, T. Laguear, and A. Charbonneau, “Cross cultural study of auditory warnings.” Paris, France : International Community for Auditory Display, Jun. 2008.
- [131] B. Laugwitz, T. Held, and M. Schrepp, “Construction and Evaluation of a User Experience Questionnaire,” vol. 5298, Nov. 2008, pp. 63–76.
- [132] A. M. Leaver and J. P. Rauschecker, “Cortical Representation of Natural Complex Sounds : Effects of Acoustic Features and Auditory Object Category,” *Journal of Neuroscience*, vol. 30, no. 22, pp. 7604–7612, 2010.
- [133] S. Lenzi, T. Riccardo, T. Ginevra, S. Galelli, P. Ciuccarelli *et al.*, “Disclosing Cyber-Attacks on Water Distribution Systems. An Experimental Approach to the Sonification of Threats and Anomalous Data,” in *25th International Conference on Auditory Display*. Northumbria University, Newcastle upon Tyne, UK : International Community on Auditory Display, Jun. 2019, pp. 125–132.
- [134] N. Li, S. Liu, Y. Liu, S. Zhao, and M. Liu, “Neural speech synthesis with transformer network,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 6706–6713.
- [135] S. K. Lodha, J. Beahan, T. Heppe, A. Joseph, and B. Zane-Ulman, “MUSE : A musical data sonification toolkit,” in *Proceedings of the 4rd International Conference on Auditory Display (ICAD1997)*, Palo Alto, California, Nov. 1997.

- [136] S. G. Lomber and S. Malhotra, “Double Dissociation of ‘What’ and ‘Where’ Processing in Auditory Cortex,” *Nature neuroscience*, vol. 11, no. 5, pp. 609–616, 2008.
- [137] L. A. Ludovico and G. Presti, “The sonification space : A reference system for sonification tasks,” *International Journal of Human-Computer Studies*, vol. 85, pp. 72–77, 2016, data Sonification and Sound Design in Interactive Systems. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S1071581915001329>
- [138] N. A. Macmillan and C. D. Creelman, *Detection Theory : A User’s Guide*. Psychology press, 2004.
- [139] T. M. Madhyastha and D. A. Reed, “Data sonification : Do you see what i hear?” *IEEE Software*, vol. 12, no. 2, pp. 45–56, 1995.
- [140] S. Matinfar, T. Hermann, M. Seibold, P. Fürnstahl, M. Farshad, and N. Navab, “Sonification for Process Monitoring in Highly Sensitive Surgical Tasks,” in *Proceedings of the Nordic Sound and Music Computing Conference 2019 (Nordic SMC 2019)*, 2019.
- [141] B. S. Mauney and B. N. Walker, “Creating Functional and Livable Soundscapes for Peripheral Monitoring of Dynamic Data,” in *Proceedings of the 10th International Conference on Auditory Display (ICAD 2004)*, Sydney, Australia, Jul. 2004.
- [142] D. S. McCrickard, C. M. Chewar, J. P. Somervell, and A. Ndiwalana, “A Model for Notification Systems Evaluation—Assessing User Goals for Multitasking Activity,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 10, no. 4, pp. 312–338, 2003.
- [143] D. K. McGookin and S. A. Brewster, “Understanding concurrent earcons : Applying auditory scene analysis principles to concurrent earcon recognition,” *ACM Transactions on Applied Perception (TAP)*, vol. 1, no. 2, pp. 130–155, 2004.
- [144] D. McGrory. (2022) Hearing the Sound of Climate Change. [Online]. Available : <https://medium.com/sonify/listening-to-the-sound-of-climate-change-48b25940673e>
- [145] M. Meier and A. Saranti, “Sonic explorations with earthquake data,” in *International Conference on Auditory Display (ICAD 2008)*, Paris, France, Jun. 2008.
- [146] J. J. Mezrich, S. Frysinger, and R. Slivjanovski, “Dynamic Representation of Multivariate Time Series Data,” *Journal of the American Statistical Association*, vol. 79, no. 385, pp. 34–40, 1984. [Online]. Available : <https://www.tandfonline.com/doi/abs/10.1080/01621459.1984.10477059>

- [147] P. Michaud, Communication personnelle, juillet 2022.
- [148] J. C. Middlebrooks, “Narrow-band Sound Localization Related to External Ear Acoustics,” *The Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2607–2624, Nov. 1992.
- [149] J. Miele, “Smith-Kettlewell display tools : A sonification toolkit for MATLAB,” in *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, Boston, MA, Jul. 2003, pp. 288–291.
- [150] A. Mihalic, “Modèles et données extra-musicales : étude de leur incidence dans le processus compositionnel,” Ph.D. dissertation, 2000, thèse de doctorat dirigée par Vaggione, Horacio Musicologie Paris 8 2000. [Online]. Available : <http://www.theses.fr/2000PA081899>
- [151] M. Milazzo and M. J. Buehler, “Designing and fabricating materials from fire using sonification and deep learning,” *iScience*, vol. 24, no. 8, p. 102873, 2021. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S2589004221008415>
- [152] B. R. Molesworth and M. Burgess, “Improving intelligibility at a safety critical point : In flight cabin safety,” *Safety Science*, vol. 51, no. 1, pp. 11–16, 2013. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0925753512001506>
- [153] B. R. Molesworth, M. Burgess, and D. Kwon, “The use of noise cancelling headphones to improve concurrent task performance in a noisy environment,” *Applied Acoustics*, vol. 74, no. 1, pp. 110–115, 2013. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0003682X12001752>
- [154] J. J. Mulder, “Designing the (Re)Context : Stolen Goods (Stocketus),” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6798291>
- [155] M. M. Murray, C. Camen, S. L. G. Andino, P. Bovet, and S. Clarke, “Rapid Brain Discrimination of Sounds of Objects,” *Journal of Neuroscience*, vol. 26, no. 4, pp. 1293–1302, 2006.
- [156] E. D. Mynatt, M. Back, R. Want, M. Baer, and J. B. Ellis, “Designing Audio Aura,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1998, pp. 566–573.
- [157] J. G. Neuhoff, R. Knight, and J. Wayand, “Pitch change, sonification, and musical expertise : Which way is up ?” in *Proceedings of the 8th International Conference on Auditory Display (ICAD)*, Kyoto, Japan, Jul. 2002.
- [158] J. G. Neuhoff, “Perception, Cognition and Action in Auditory Displays,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany : Logos Publishing House, 2011, ch. 4, pp. 63–85.

- [159] I. Neuman, “Tartyp Generative Grammars and The Analysis of Hip Hop Beats,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6798298>
- [160] M. Olivo, “Les instruments de musique numériques pour les personnes en situation de handicap cognitif et mental : exemples d’applications via le web,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6800845>
- [161] A. Oord, Y. Li, I. Babuschkin, K. Simonyan, O. Vinyals, K. Kavukcuoglu, G. Driessche, E. Lockhart, L. Cobo, F. Stimberg *et al.*, “Parallel wavenet : Fast high-fidelity speech synthesis,” in *International conference on machine learning*. PMLR, 2018, pp. 3918–3926.
- [162] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “Wavenet : A generative model for raw audio,” *arXiv preprint arXiv :1609.03499*, 2016.
- [163] A. Owens, P. Isola, J. McDermott, A. Torralba, E. H. Adelson, and W. T. Freeman, “Visually indicated sounds,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2405–2413.
- [164] T. H. Park, J. Turner, C. Jacoby, A. Marse, M. Musick, A. Kapur, and J. He, “Locative sonification : Playing the world through citygram,” in *ICMC*, 2013.
- [165] M. J. Paszek, C. C. DuFort, O. Rossier, R. Bainer, J. K. Mouw, K. Godula, J. E. Hudak, J. N. Lakins, A. C. Wijekoon, L. Cassereau *et al.*, “The cancer glycolyx mechanically primes integrin-mediated growth and survival,” *Nature*, vol. 511, no. 7509, pp. 319–325, 2014.
- [166] R. D. Patterson, “Auditory warning sounds in the work environment,” *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, vol. 327, no. 1241, pp. 485–492, 1990.
- [167] S. Pauletto and A. Hunt, “A toolkit for interactive sonification,” in *Proceedings of ICAD 04. Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia, Jul. 2004.
- [168] E. R. Pedersen and T. Sokoler, “AROMA : Abstract Representation of Presence Supporting Mutual Awareness,” in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI97)*, ser. CHI ’97. Atlanta, GA, USA : Association for Computing Machinery, 1997, pp. 51–58.
- [169] J. Pereira. (2020) Climate Data Sonification. [Online]. Available : <https://www.jamieperera.com/climate-data-sonification>

- [170] S. C. Peres and D. M. Lane, “Auditory Graphs : The effects of redundant dimensions and divided attention,” in *Proceedings of the International Conference on Auditory Display (ICAD2005)*. Limerick, Ireland : International Community for Auditory Display, Jul. 2005, pp. 169–174.
- [171] L. S. Petrucci, E. Harth, and P. Roth, “Websound : A generic web sonification tool, and its application to an auditory web browser for blind and visually impaired users,” in *Proceedings of the 6th International Conference on Auditory Display (ICAD2000)*, Atlanta, GA, Apr. 2000.
- [172] S. Phillips and A. Cabrera, “Sonification workstation,” in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Northumbria University, Newcastle upon Tyne, UK, Jun. 2019, pp. 184–190.
- [173] W. Ping, K. Peng, and J. Chen, “Clarinet : Parallel wave generation in end-to-end text-to-speech,” *arXiv preprint arXiv :1807.07281*, 2018.
- [174] W. Ping, K. Peng, A. Gibiansky, S. O. Arik, A. Kannan, S. Narang, J. Raiman, and J. Miller, “Deep voice 3 : Scaling text-to-speech with convolutional sequence learning,” *arXiv preprint arXiv :1710.07654*, 2017.
- [175] I. Pollack and L. Ficks, “Information of elementary multidimensional auditory displays,” *The Journal of the Acoustical Society of America*, vol. 26, no. 2, pp. 155–158, 1954.
- [176] M. Poret, “Sonification,” *Rapport de Stage pour Master Informatique, Université de Bordeaux*, 2019.
- [177] N. Privato, O. Rampado, and A. Novello, “Scramble Live : Combining LSTM and Markov Chains for Real-time Musical Interaction,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6800786>
- [178] W. S. Project and B. Truax, *The World Soundscape Project’s Handbook for Acoustic Ecology*, ser. Music of the environment series. A.R.C. Publications, 1978. [Online]. Available : <https://books.google.fr/books?id=hJ5tQgAACAAJ>
- [179] M. Puckette and K. L. Hagan, “PD on an Iot-Class Processor,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6800795>
- [180] M. Rauterberg and E. Styger, “Positive Effects of Sound Feedback During the Operation of a Plant Simulator,” in *International Conference on Human-Computer Interaction*. Springer, 1994, pp. 35–44.

- [181] Y. Ren, J. Liu, and Z. Zhao, “Portaspeech : Portable and high-quality generative text-to-speech,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 13 963–13 974, 2021.
- [182] Y. Ren, Y. Ruan, X. Tan, T. Qin, S. Zhao, Z. Zhao, and T.-Y. Liu, “Fastspeech : Fast, robust and controllable text to speech,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [183] F. Riess, P. Heering, and D. Nawrath, “Reconstructing Galileos Inclined Plane Experiments for Teaching Purposes,” in *Proc. of the International History, Philosophy, Sociology and Science Teaching Conference, Leeds*, 2005.
- [184] S. Roddy and B. Bridges, “Mapping for meaning : the embodied sonification listening model and its implications for the mapping problem in sonic information design,” *Journal on Multimodal User Interfaces*, vol. 14, no. 2, pp. 143–151, 2020.
- [185] J. Rohrhuber, “S – Introducing sonification variables,” in *Proceedings of the Supercollider Symposium*, 2010.
- [186] N. Rönnerberg, J. Lundberg, and J. Löwgren, “Sonifying the Periphery : Supporting the Formation of Gestalt in Air Traffic Control,” in *Proceedings of The 5th Interactive Sonification Workshop (ISon-2016)*, CITEC, Bielefeld University, Germany, 2016, pp. 23–27.
- [187] O. Rossier, V. Octeau, J.-B. Sibarita, C. Leduc, B. Tessier, D. Nair, V. Gatterdam, O. Destaing, C. Albiges-Rizo, R. Tampé *et al.*, “Integrins $\beta 1$ and $\beta 3$ exhibit distinct dynamic nanoscale organizations inside focal adhesions,” *Nature cell biology*, vol. 14, no. 10, pp. 1057–1067, 2012.
- [188] A. Savery, “Distracted - A piece for Violin and an Audio-Visual Interface,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6800823>
- [189] N. Sawhney and C. Schmandt, “Nomadic Radio : Speech and Audio Interaction for Contextual Messaging in Nomadic Environments,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 7, no. 3, pp. 353–383, 2000.
- [190] A. Schertenleib and S. Barrass, “A social platform for information sonification : many-ears.com,” in *16th International Conference on Auditory Display*, Washington, DC, Jun. 2010, pp. 295–299.
- [191] C. E. Schmidt, T. Chen, and D. A. Lauffenburger, “Simulation of integrin-cytoskeletal interactions in migrating fibroblasts,” *Biophysical journal*, vol. 67, no. 1, pp. 461–474, 1994.

-
- [192] A. Schoon and F. Dombois, “Sonification in Music,” in *Proceedings of the 15th International Conference on Auditory Display (ICAD2009)*, Copenhagen, Denmark, May 2009.
- [193] M. Schrepp, A. Hinderks, and J. Thomaschewski, “Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S),” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, p. 103, Jan. 2017.
- [194] R. Selfridge, C. Barone, and S. Pauletto, “Sonifying an Office Gadget to Indicate Air Quality,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6572709>
- [195] Y. Seznec and S. Pauletto, “The Singing Shower : A Melody-Sensitive Interface For Physical Interaction and Efficient Energy Consumption,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6798253>
- [196] R. Shelton, N. Aeronautics, and S. Administration. (2005) MathTrax. Accessible Math and Science for 6th-12th Grade Students. [Online]. Available : <https://prime.jsc.nasa.gov/mathtrax/>
- [197] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. Skerrv-Ryan *et al.*, “Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions,” in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 4779–4783.
- [198] A. Simon-Lewis. (2017) Climate change data is being transformed into beautiful, haunting symphonies. [Online]. Available : <https://www.wired.co.uk/article/climate-symphony-data-sonification>
- [199] D. R. Smith and B. N. Walker, “Tick-marks, axes, and labels : The effects of adding context to auditory graphs,” in *Proceedings of the 8th International Conference on Auditory Display (ICAD)*, Kyoto, Japan, Jul. 2002.
- [200] L. Sousa and A. Pinto, “MuSec : Sonification of alarms generated by a SIEM,” in *International Symposium on Ambient Intelligence*. Springer, 2017, pp. 32–39.
- [201] S. D. Speeth, “Seismometer sounds,” *The Journal of the Acoustical Society of America*, vol. 33, no. 7, pp. 909–916, 1961.
- [202] C. Spence, “Audiovisual multisensory integration,” *Acoustical science and technology*, vol. 28, no. 2, pp. 61–70, 2007.

- [203] R. Stevens, S. Brewster, P. Wright, and A. Edwards, “Design and evaluation of an auditory glance at algebra for blind readers,” in *Proceedings of the 2nd International Conference on Auditory Display (ICAD1994)*, Santa Fe, New Mexico, Nov. 1994, pp. 21–30.
- [204] S. S. Stevens, J. Volkman, and E. B. Newman, “A scale for the measurement of the psychological magnitude pitch,” *The journal of the acoustical society of america*, vol. 8, no. 3, pp. 185–190, 1937.
- [205] T. Stockman, G. Hind, and C. Frauenberger, “Interactive Sonification of Spreadsheets,” in *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, Limerick, Ireland, Jul. 2005.
- [206] X. Tan, J. Chen, H. Liu, J. Cong, C. Zhang, Y. Liu, X. Wang, Y. Leng, Y. Yi, L. He *et al.*, “NaturalSpeech : End-to-End Text to Speech Synthesis with Human-Level Quality,” *arXiv preprint arXiv :2205.04421*, 2022.
- [207] M. Tilemachos, N. Kotsani, and A. Georgaki, “Kinesthesia : An Interactive Voice Software for Intertextual Improvisation in Theatre Performance,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6797798>
- [208] Q. T. Tran and E. D. Mynatt, “Music Monitor : Ambient Musical Data for the Home,” in *Proceedings of the IFIP WG 9.3 International Conference on Home Oriented Informatic and Telematics (HOIT 2000)*, A. Sloan and F. van Rijn, Eds., vol. 173. Kluwer, 2000, pp. 85–92.
- [209] J. Tricot, “Aristote : De l’âme, traduction nouvelle et notes,” *Tricot, Paris*, 1934.
- [210] G. M. Vallejo Rosas, “ListenIn : Ambient Auditory Awareness at Remote Places,” Ph.D. dissertation, Massachusetts Institute of Technology, 2003.
- [211] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [212] D. Venton. (2018) LISTEN : 1,200 Years of Earth’s Climate, Transformed into Sound. [Online]. Available : <https://www.kqed.org/science/1918660/listen-1200-years-of-earths-climate-transformed-into-sound>
- [213] A. Verrier, V. Goudard, E. Hong, and H. Genevois, “Interactive Software for the Sonification of Neuronal Activity,” in *Proceedings of the 17th Sound and Music Computing Conference SMC 2020*, Jun. 2020.
- [214] P. Vickers, “Sonification for Process Monitoring,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. Neuhoff, Eds. Logos Verlag, 2011, ch. 18, pp. 455–491.

-
- [215] B. N. Walker, “Magnitude estimation of conceptual data dimensions for use in sonification,” *Journal of experimental psychology : Applied*, vol. 8, no. 4, p. 211, 2002.
- [216] —, “Consistency of magnitude estimations with conceptual data dimensions used for sonification,” *Applied Cognitive Psychology*, vol. 21, no. 5, pp. 579–599, 2007. [Online]. Available : <https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.1291>
- [217] B. N. Walker and J. T. Cothran, “Sonification Sandbox : a graphical toolkit for auditory graphs,” in *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, Boston, MA, Jul. 2003, pp. 161–163.
- [218] B. N. Walker and G. Kramer, “Mappings and metaphors in auditory displays : An experimental assessment,” *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 4, pp. 407–412, 2005.
- [219] B. N. Walker, G. Kramer, and D. M. Lane, “Psychophysical scaling of sonification mappings,” in *Proceedings of the 6th International Conference on Auditory Display (ICAD2000)*, Atlanta, GA, Apr. 2000.
- [220] B. N. Walker and D. M. Lane, “Psychophysical scaling of sonification mappings : A comparison of visually impaired and sighted listeners,” in *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, Espoo, Finland, Jul. 2001.
- [221] B. N. Walker and L. M. Mauney, “Individual differences, cognitive abilities, and the interpretation of auditory graphs,” in *Proceedings of ICAD 04. Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia, Jul. 2004.
- [222] B. N. Walker and M. A. Nees, “Theory of Sonification,” in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany : Logos Publishing House, 2011, ch. 2, pp. 9—39.
- [223] M. M. Wanderley, N. Schnell, and J. Rován, “ESCHER – modeling and performing composed instruments in real-time,” in *SMC’98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218)*, vol. 2. IEEE, 1998, pp. 1080–1084.
- [224] Y. Wang, R. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio *et al.*, “Tacotron : Towards end-to-end speech synthesis,” *arXiv preprint arXiv :1703.10135*, 2017.
- [225] C. Wegener, S. Stang, and M. Neupert, “FPGA-accelerated Real-Time Audio in Pure Data,” in *Proceedings of the 19th Sound and Music Computing Conference (SMC22)*. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6797805>

- [226] S. W. Williams, F. Martina, A. C. Addison, J. Ding, G. Pardal, and P. Colegrove, “Wire + Arc Additive Manufacturing,” *Materials Science and Technology*, vol. 32, no. 7, pp. 641–647, 2016. [Online]. Available : <https://doi.org/10.1179/1743284715Y.0000000073>
- [227] C. M. Wilson and S. K. Lodha, “Listen : A data sonification toolkit,” in *Proceedings of the 3rd International Conference on Auditory Display (ICAD1996)*, Palo Alto, California, Nov. 1996.
- [228] R. Wilson, J. Berger, and W. Yeo, “A flexible framework for real-time sonification with SONART,” in *Proceedings of ICAD 04. Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia, Jul. 2004.
- [229] K. E. Wolf and R. Fiebrink, “Sonnet : A code interface for sonifying computer network data,” in *NIME*, 2013, pp. 503–506.
- [230] —, “Toward supporting end-user design of soundscape sonifications,” in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Northumbria University, Newcastle upon Tyne, UK, Jun. 2019, pp. 268–276.
- [231] D. Worrall, “Sonification : A Prehistory,” in *Proceedings of the 24th International Conference on Auditory Display (ICAD2018)*, Michigan Technological University, Houghton, MI, USA, Jun. 2018, pp. 177–182.
- [232] D. Worrall, M. Bylstra, and R. Dean, “SoniPy : the Design of an Extendable Software Framework for Sonification Research and Auditory Display,” in *Proceedings of the 13th International Conference on Auditory Display*, Montréal, Canada, Jun. 2007, pp. 445–452.
- [233] C. Xia, Z. Pan, J. Polden, H. Li, Y. Xu, S. Chen, and Y. Zhang, “A Review on Wire Arc Additive Manufacturing : Monitoring, Control and a Framework of Automated System,” *Journal of Manufacturing Systems*, vol. 57, pp. 31–45, 2020. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0278612520301412>
- [234] R. Yamamoto, E. Song, and J.-M. Kim, “Parallel WaveGAN : A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6199–6203.
- [235] E. S. Yeung, “Pattern recognition by audio representation of multivariate analytical data,” *Analytical Chemistry*, vol. 52, no. 7, pp. 1120–1123, 1980.
- [236] J. Zhang, C. Zong *et al.*, “Deep Neural Networks in Machine Translation : An Overview,” *IEEE Intell. Syst.*, vol. 30, no. 5, pp. 16–25, 2015.

- [237] H. Zhao, C. Gan, W.-C. Ma, and A. Torralba, “The sound of motions,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1735–1744.
- [238] H. Zhao, C. Gan, A. Rouditchenko, C. Vondrick, J. McDermott, and A. Torralba, “The sound of pixels,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 570–586.
- [239] Y. Zhou, Z. Wang, C. Fang, T. Bui, and T. L. Berg, “Visual to Sound : Generating Natural Sound for Videos in the Wild,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3550–3558.
- [240] D. N. Zotkin, R. Duraiswami, and L. S. Davis, “Customizable auditory displays,” in *Proceedings of the 8th International Conference on Auditory Display (ICAD)*, Kyoto, Japan, Jul. 2002.

Publications

Articles in Peer-reviewed Conferences

- [241] M. Poret, M. Desainte-Catherine, and C. Semal, “A Perceptual Study of Sound Ecology in Peripheral Sonification,” ser. Proceedings of the 19th Sound and Music Computing Conference. Saint-Étienne, France : Zenodo, Jun. 2022. [Online]. Available : <https://doi.org/10.5281/zenodo.6572730>
- [242] M. Poret, S. Ibarboure, M. Robine, E. Duc, C. Semal, M. Desainte-Catherine, and N. Couture, “Sonification for 3D Printing Process Monitoring,” in *International Computer Music Conference*, Santiago, Chile, 2020. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-03118034>
- [243] M. Poret, S. Ibarboure, C. Semal, M. Desainte-Catherine, and N. Couture, “Peripheral Auditory Display for 3D-Printing Process Monitoring,” in *18th Sound and Music Computing Conference SMC 2021*, D. A. Mauro, S. Spagnol, and A. Valle, Eds., Turin, Italy, Jun. 2021, pp. 268–275, iISBN 978-88-945415-4-0. [Online]. Available : <https://hal.archives-ouvertes.fr/hal-03617168>

Posters

- [244] J. David, M. Poret, M. Desainte-Catherine, and C. Semal, “Exploration sonore interactive de l’activité cellulaire,” in *Séminaire des doctorants et doctorantes en informatique*, juin 2021. [Online]. Available : <https://sifdoctorants21.sciencesconf.org/357323>