

Son et mathématiques

1 Introduction

Dans le langage courant, on distingue souvent le son du bruit par l'intention de l'auteur. Cependant, le mécanisme physique en jeu est le même : une vibration de l'espace ambiant – généralement l'air – perçue par l'appareil auditif, et donnant naissance à des sensations auditives.

On donne souvent des analogies simples pour aider à la compréhension du phénomène de propagation des sons : un caillou jeté à la surface d'un lac, dont l'onde concentrique va se propager avec une oscillation propre, ou encore la vibration d'une corde quand on la pince. Ces analogies sont parfois trompeuses, car elle n'illustrent pas complètement le phénomène de compression/décompression qui s'opère dans le milieu ambiant.

Dans la suite de ce document, on donne quelques clés de compréhension des aspects physiques des sons, puis on s'intéresse à la manière dont les pratiques musicales occidentales ont formalisé la production artistique sonore. Enfin, on aborde la question de la synthèse additive, un moyen de s'approprier la notion de son et de note par la modélisation.

2 Éléments de physique du son

Considérons une source sonore ponctuelle, dans un volume d'air sans obstacle (situation théorique impossible à réaliser). Pour produire un son, la source sonore vibre, produisant localement une compression/décompression itérative de l'air. Cette vibration locale de l'air va se déplacer de proche en proche (ce ne sont pas les molécules qui se déplacent, mais le phénomène de compression/décompression), à une vitesse propre à l'espace ambiant.

Ainsi, la vitesse de propagation du son dans l'air sec sous une pression d'une atmosphère est de 343,4 mètres par seconde à une température de 20 degrés. Plus le matériau est rigide, plus cette propagation est rapide. Dans l'eau, le son parcourt environ 1500 mètres par seconde, et dans le béton plus de 3000 mètres par seconde.

Lorsqu'on mesure le son, on place à une distance fixe de la source un récepteur, ou capteur. Il s'agit généralement d'un microphone, constitué d'une membrane qui en vibrant restitue grâce à une bobine un signal électrique correspondant à l'oscillation du son produit.

La figure 1 illustre ce qui pourrait être capté depuis un point placé près d'une source produisant un son simple, composé d'une seule fréquence. Le tableau 1 donne quelques exemples de sons associés à leur fréquence.

objet sonore	diapason	violon	piano	roue et rail d'un train [FP14]
Fréquence (Hz)	440	200 à 2650 Hz	27,5 à 4186 Hz	autour de 1000 Hz

TABLE 1 – Exemple de quelques fréquences typiques.

L'intensité sonore est généralement exprimée en décibels (dB). Cette valeur exprime un rapport de puissance entre l'intensité mesurée et une valeur de référence.

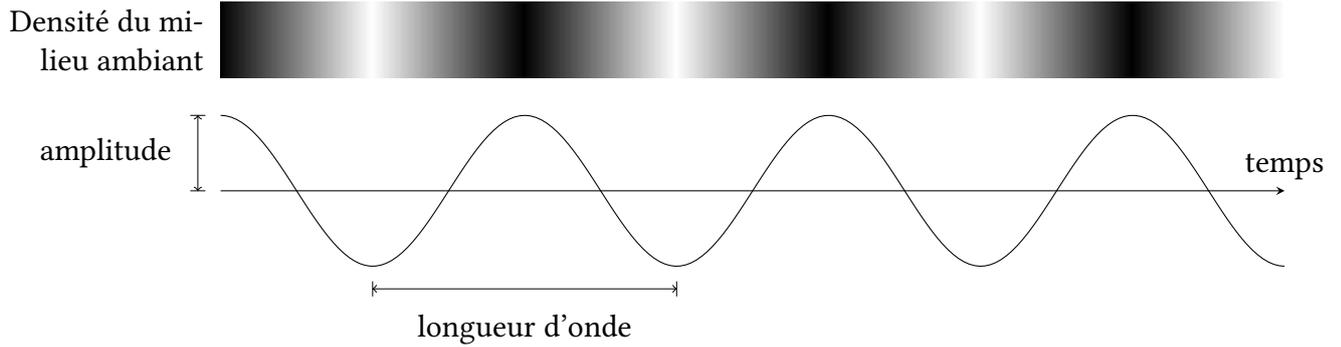


FIGURE 1 – Phénomène de compression/décompression du son (noir : en surpression ; gris : pression normale ; blanc : en sous-pression), et sa transcription en forme d'onde.

Lorsqu'il s'agit de mesures dans la vie réelle, on mesure la *pression acoustique*, et on choisira en général la valeur de référence comme celui d'un son imperceptible. Dans ce cas, on parle de *décibels Sound Pressure Level*, ou *dB SPL*. La pression acoustique suit une échelle approximativement logarithmique : une puissance double correspond environ à une augmentation de 3 dB.

3 Perception du son

L'*amplitude* du phénomène de compression/décompression est perçue par l'oreille humaine comme le volume sonore du son, quand la *longueur d'onde* et sa *fréquence* associée (nombre d'oscillations par seconde, exprimée en hertz) correspondent à la hauteur perçue.

La perception humaine des sons se situe entre 20 Hz (sons graves) et 20 000 Hz (sons aigus). En dessous de 20 Hz on parle des *infrasons*, qui sont perçus comme des phénomènes distincts. Au delà de 20 000 Hz, on parle d'*ultrasons*, phénomène utilisé dans l'industrie ou la médecine. Cette plage de perception varie d'une personne à l'autre, évolue notamment en fonction de l'âge, et est très différente suivant l'intensité du son considérée (figure 2).

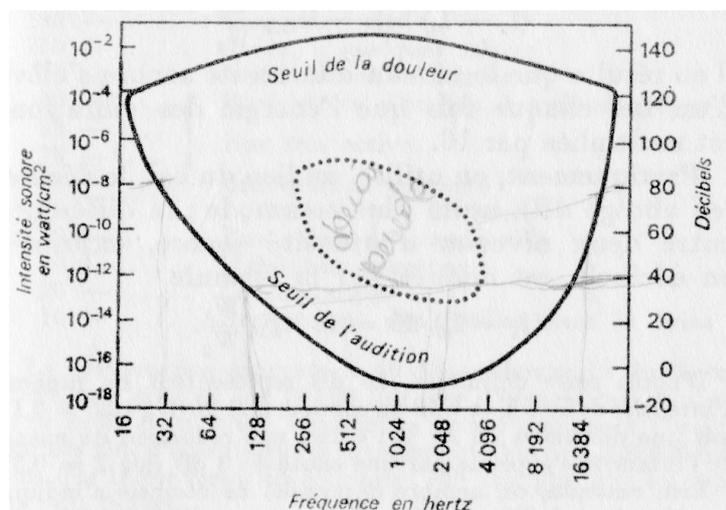


FIGURE 2 – Champ auditif d'un être humain (courbe en trait plein) comparé au domaine de la parole (courbe en pointillés). Source [Gri94].

Si la distance entre la source et le récepteur varie, on observe l'effet Doppler, phénomène

bien connu où les sons produits vont être déformés par le mouvement. On pense par exemple à la sirène d'une ambulance qui arrive au loin dans une rue, passe devant l'auditeur, puis s'éloigne dans la direction opposée.

3.1 Le son numérique

La figure 1 présente une courbe appelée forme d'onde (voir 3.3.1), qui décrit le son par l'évolution de la pression au fil du temps. Il s'agit donc d'une fonction $f : \mathbb{R}^+ \rightarrow \mathbb{R}$, considérée comme continue.

Tout comme l'encodage numérique d'une image, l'encodage numérique d'un son analogique passe par une étape d'échantillonnage (figure 3), puis par le stockage pour chaque échantillon de l'intensité sonore correspondante. La fonction f est alors décrite par un ensemble de valeurs discrètes f_0, f_1, \dots, f_n entières.

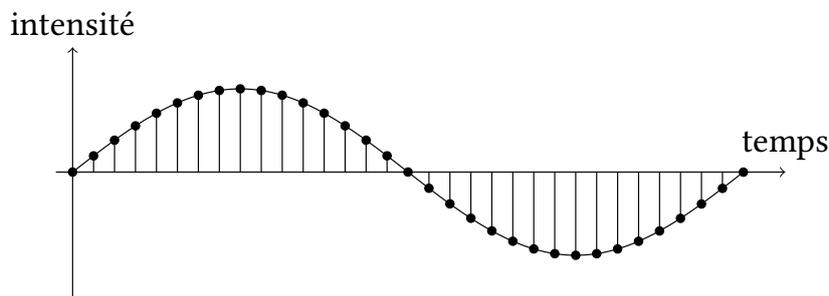


FIGURE 3 – Échantillonnage d'un signal analogique sinusoïdal.

Deux paramètres sont donc réglables, et il existe pour chacun d'eux des conventions issues de choix techniques historiques.

- La *fréquence d'échantillonnage*, correspondant au nombre d'échantillons par seconde. Puisque l'oreille humaine perçoit des fréquences jusqu'à 20 000 Hz, on a choisi d'échantillonner à au moins 40 000 Hz afin d'avoir une représentation assez fine du son pour que l'erreur ne soit pas perceptible par l'oreille humaine. Ainsi, le format de stockage Disque Compact a normalisé au début des années 80 l'usage à une fréquence de 44 100 Hz. C'est aujourd'hui encore la fréquence d'échantillonnage souvent utilisée dans les formats wav ou mp3. Dans l'audiovisuel, on a en parallèle opté pour un autre standard, en choisissant une fréquence multiple des 24 images par seconde, avec une fréquence d'échantillonnage sonore de 48 000 Hz. Avec l'augmentation de la performance des ordinateurs, on a aussi vu émerger l'utilisation de fréquences multiples de ces deux premières (88 200 Hz, 96 000 Hz ou 192 000 Hz).
- La *quantification* ou nombre de bits utilisés pour coder chaque échantillon. En pratique, on observe des encodage à 8, 16, 24, ou plus rarement 32 bits. Cette précision est notamment utile dans le cas où l'on souhaite augmenter le volume d'un son numériquement : si l'encodage s'est fait avec une quantification trop faible, l'approximation due à la discrétisation est amplifiée, et on produit artificiellement du bruit non désiré qui vient polluer le signal initial. Dans le cas des disques compact par exemple, la valeur numérique est codée sur 16 bits et prend donc une valeur entière entre -32 768 et +32 767, les bornes correspondant aux extrema de l'intensité sonore représentable.

Dans le cas d'un son numérique, l'unité d'intensité est toujours exprimée en décibels par rapport à une intensité de référence. Contrairement à la situation d'une mesure réelle, on choisira ici comme intensité de référence la valeur maximale représentable. Les dB seront

donc négatifs, et on parle de *décibels Full Scale*, ou dBFS. Un intérêt notable de cette échelle est que la précision de l'encodage n'influe pas la valeur d'une mesure.

Cette échelle reprend une progression logarithmique, diminuer la puissance l'intensité apparente par deux revenant environ à une diminution de 3 dB.

3.2 Décomposition en fréquences

Un son n'est quasiment jamais composé d'une unique fréquence, mais au contraire d'un assemblage de multiples fréquences. Cet assemblage constitue une propriété significative des sons. Une oreille humaine entraînée est capable de percevoir plusieurs notes composant un son entendu. Mathématiquement, on peut également procéder au même type de décomposition.

Étant donnée une fonction d'onde $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ telle que présentée sur la figure 1, on peut réaliser une *transformée de Fourier* afin d'obtenir une décomposition en fréquences, appelé le *spectre du signal*.

La transformée de Fourier est la fonction $\mathcal{F}(f) = \hat{f}$ donnée par la formule :

$$\mathcal{F}(f) : \xi \mapsto \hat{f}(\xi) = \int_{-\infty}^{+\infty} f(x) e^{-i\xi x} dx.$$

On réalise généralement cette transformation sur un intervalle de temps fixé, correspondant à l'échantillon sonore que l'on souhaite analyser. Voir [Coh10] pages 108 à 111 pour une introduction à cette transformation dans le cadre d'une application musicale.

Ce calcul n'est en pratique pas très simple dans le cas continu, et puisqu'on travaille généralement avec un signal numérisé, on a tendance à appliquer une transformée de Fourier discrète, pour laquelle des algorithmes rapides en $O(n \log(n))$ existent. Dans les faits, si un signal est donné par l'échantillonnage f_0, f_1, \dots, f_n dans l'intervalle souhaité, il s'agit de trouver les x_0, x_1, \dots, x_n tels que :

$$f_j = \sum_{k=0}^{n-1} x_k e^{-\frac{2\pi i}{n} jk} \quad j = 0, \dots, n-1.$$

Dans le cas d'une fonction sinusoïdale, les x_i sont tous nuls, sauf celui correspondant à la fréquence de la fonction (figure 4).

Cette décomposition en fréquences est couramment utilisée comme un outil pour la visualisation des propriétés d'un son (voir section 3.3.2).

Il arrive aussi que l'on fasse la transformation inverse, notamment en synthèse sonore, afin de produire un son depuis sa description en fréquences. On peut par exemple citer les *bruits de synthèse*, dont les utilisations sont très variées.

Le principe du *bruit blanc* consiste à générer, à l'aide d'une loi normale, un tirage aléatoire pour chaque fréquence, afin de produire un son dont la densité spectrale de puissance est la même pour toutes les fréquences. Ce type de bruit n'est pas parfaitement adapté à la perception humaine, qui a un comportement logarithmique. Lorsqu'on souhaite produire un signal sonore aléatoire pour l'oreille humaine, on privilégie donc le *bruit rose*, dont la densité spectrale est constante par bande d'octave.

Dans les faits, ces sons évoquent de manière très synthétique un bruit régulier et uniforme, comme celui d'une chute d'eau par exemple. La figure 9 propose une représentation graphique d'un bruit blanc et d'un bruit rose.

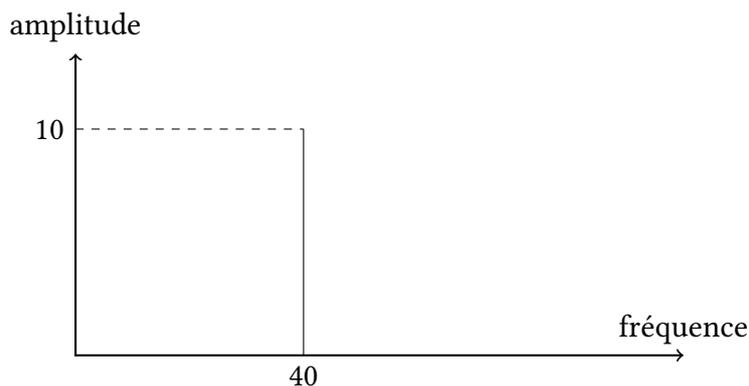


FIGURE 4 – Spectre de la fonction $f : x \mapsto 10 \sin(40 \times 2\pi \times x)$

3.3 Outils pour la visualisation

La représentation graphique des sons est aujourd'hui grandement facilitée par l'utilisation des ordinateurs, et de la chaîne numérique d'enregistrement des sons en général. Mais déjà en 1853, avant même l'invention du phonographe par Thomas Edison, Édouard-Léon de Scott de Martinville transformait les sons en tracés à l'aide d'une machine de son invention, le phonautographe. Le projet *First sounds*¹ a réussi en 2007 l'exploit de reconstituer ces premiers enregistrements sous forme sonore, rendant audible la plus vieille voix humaine jamais enregistrée.

Le phonautographe, tout comme le phonographe et les autres explorations techniques du XIX^e siècle, s'inspire du fonctionnement de l'oreille, en transformant les variations de pression que constitue le son en une variation mécanique, à l'aide d'une membrane ou d'un diaphragme. Ce mouvement mécanique est alors retransmis à un stylet qui vient marquer un cylindre.

Les représentations utilisées aujourd'hui pour visualiser le son ont dépassé les contraintes mécaniques, pour mettre en évidence la forme de l'onde sonore, ou son analyse spectrale.

Toutes les figures de cette section sont des captures d'écran du logiciel Audacity². Il s'agit d'un logiciel libre, gratuit et multiplateforme, qui permet de visualiser et de manipuler très simplement des fichiers son.

3.3.1 Forme d'onde

La représentation la plus couramment utilisée pour décrire un son consiste à représenter la forme de son onde. En abscisse, on place l'intensité sonore, et en ordonnée le temps. La figure 5 présente sur la première ligne deux tracés de forme d'onde, la première correspondant à une phrase parlée, la seconde à un bruit mécanique régulier. Le tracé du bas est un zoom sur un détail de la voix, qui illustre une certaine régularité du son quand on descend en dessous de la seconde, et que le son n'est pas trop explosif (ici, le son « nou » de « écran noir »).

Cette représentation graphique permet d'identifier facilement quelques aspects d'un son :

- le *profil d'intensité*, permettant d'identifier si le volume d'un son est faible ou fort, s'il évolue au fil du temps... Si on prête attention au début et à la fin du son mécanique présenté en haut à droite de la figure 5, on constate que le son commence par une montée en intensité, puis que le son est stable pendant une large partie du temps, et qu'il finit par une partie qui baisse en intensité. Cette propriété est notamment utilisée

1. <https://www.firstsounds.org/>

2. <https://www.audacityteam.org/>

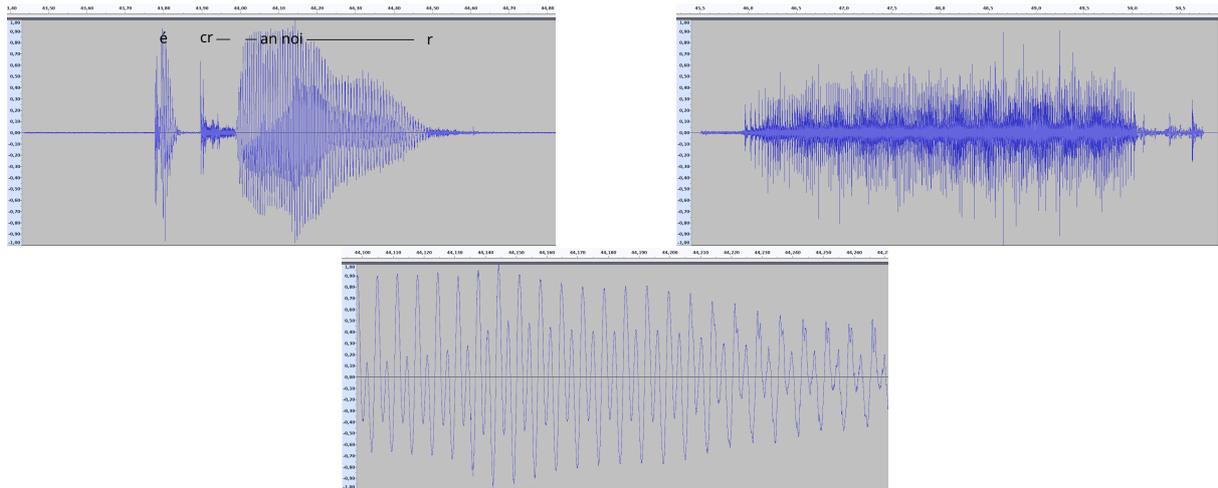


FIGURE 5 – Formes d’ondes : une voix en studio (« écran noir »), une roue libre de vélo qui crépite, un détail de la voix en studio (« [...] nou [...] »). Logiciel : audacity.

en synthèse sonore par la notion d’enveloppe (voir section 5.2).

- la *hauteur* (fréquence) d’un son. Par exemple, dans la phrase « écran noir », on observe que la forme d’onde des consonnes est décrite par des oscillations plus rapides (sons aigus) que celle des voyelles (sons graves).
- la *durée* d’un son. On peut voir sur la forme d’onde que le son « c » est un son très bref, comparé au son du « r » qui s’étale sur un temps plus long.
- le fait qu’un son soit *musical* ou au contraire plus proche d’un bruit, par sa régularité. Les voyelles de la phrase « écran noir » sont ainsi décrites par des oscillations plus régulières et douces que celles du bruit mécanique.

La forme d’onde est la représentation la plus brute d’un son, elle correspond au signal mono qui est capté par un microphone. Il est courant de manipuler un son stéréophonique, qui lui sera décrit par deux formes d’ondes, une par canal.

3.3.2 Spectre fréquentiel

En reprenant la décomposition en fréquences présentée à la section 3.2, on peut facilement produire pour un intervalle de temps choisi par l’utilisateur un graphique représentant la décomposition en fréquences d’une fonction. Dans les logiciels de traitement de son, ces diagrammes sont généralement accessibles via une fonction intitulée *analyse spectrale*. L’abscisse correspond aux fréquences, et l’ordonnée est exprimée en décibels négatifs, le zéro correspondant à l’intensité sonore la plus élevée que peut représenter l’encodage utilisé.

Les spectres ainsi produits permettent de comprendre les sons étudiés de manière plus synthétique que ne le propose la forme d’onde. Cependant, plus l’intervalle sélectionné est long, moins la précision du graphique est importante, d’autant plus si le son n’est pas musical.

La figure 6 donne deux spectres calculés par le logiciel Audacity (sélectionner un intervalle avec l’outil de sélection, puis menu *Analyse* ⇒ *Tracer le spectre* . . .). La figure de gauche est le spectre d’une voix, sur une phrase complète, où l’on constate un premier maximum local aux alentours de 145 Hz, et un second autour de 300 Hz. La figure de droite correspond à une note jouée au piano. On y observe très clairement la note fondamentale, située à 824 Hz, puis les harmoniques suivantes, qui s’estompent progressivement quand on monte en fréquence.

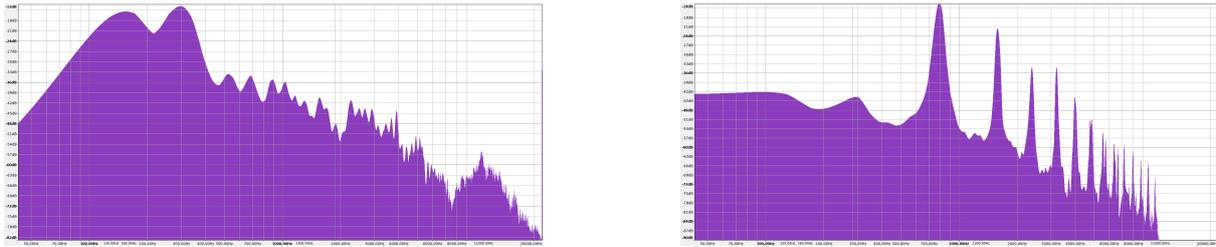


FIGURE 6 – Spectres de sons : une voix en studio (« écran noir »), un piano jouant une note (sol #4). Logiciel : audacity.

3.3.3 Spectrogramme

Si le spectre d'une fonction permet de l'analyser globalement, on perd la lecture temporelle du son. Un *spectrogramme* ou *sonogramme* permet de combiner les deux vues, en échantillonnant le temps pour t_0, t_1, \dots, t_n , et en calculant le spectre du signal sur une fenêtre glissante $[t_i - \frac{w}{2}, t_i + \frac{w}{2}]$ de largeur w . On obtient alors pour chaque échantillon de temps i et pour chaque fréquence j une amplitude f_{ij} , que l'on exprime en niveau de gris sur un diagramme où les abscisses correspondent au temps, et les ordonnées à la fréquence. On place habituellement les basses fréquences en bas du diagramme, et les hautes fréquences en haut. Dans ce type de diagramme, le noir représente souvent une amplitude nulle, quand le blanc représente l'amplitude maximale de la fréquence considérée. La figure 7 présente trois exemples de spectrogrammes.

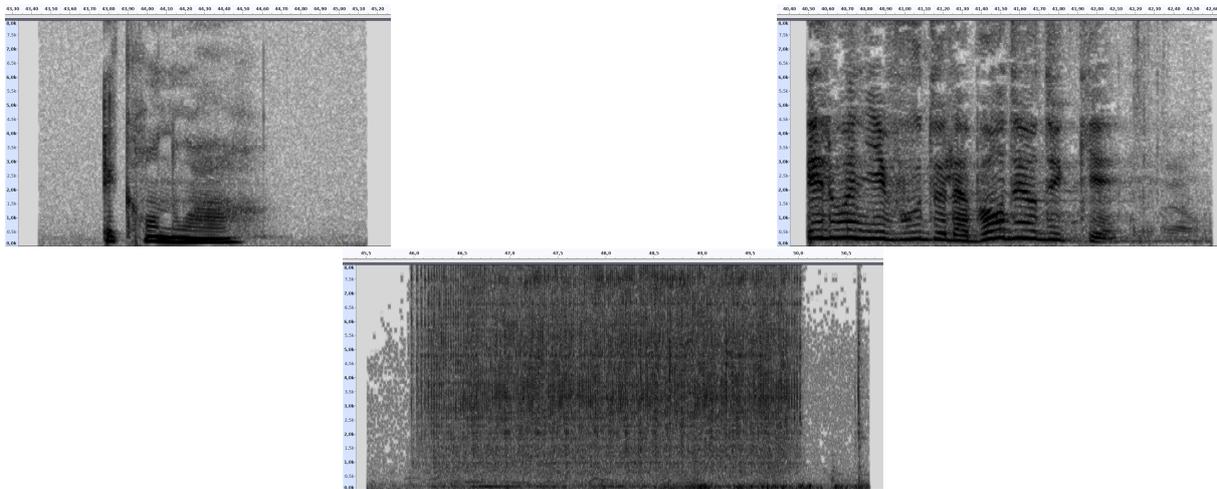


FIGURE 7 – Spectrogrammes de sons : une voix en studio (« écran noir »), la voix de la SNCF (« éloignez-vous de la bordure du quai »), une roue libre de vélo qui crépite. Logiciel : audacity.

Cette représentation permet de visualiser explicitement les partiels qui composent un son. Par exemple, dans la figure 7 droite, on observe des lignes noires qui marbrent le diagramme. Elles correspondent à des fréquences très présentes, dont la hauteur évolue de manière lente au fil du déroulé du son.

Le logiciel Audacity propose une visualisation par spectrogramme d'un son. Pour changer la représentation par défaut d'une piste, on utilise le menu déroulant qui apparaît quand on clique sur le nom de la piste, et on choisit « spectrogramme ». Par défaut, le logiciel utilise une représentation en fausses couleurs (figure 8), les couleurs froides représentant une faible amplitude, les couleurs chaudes représentant une grande amplitude de la fréquence.

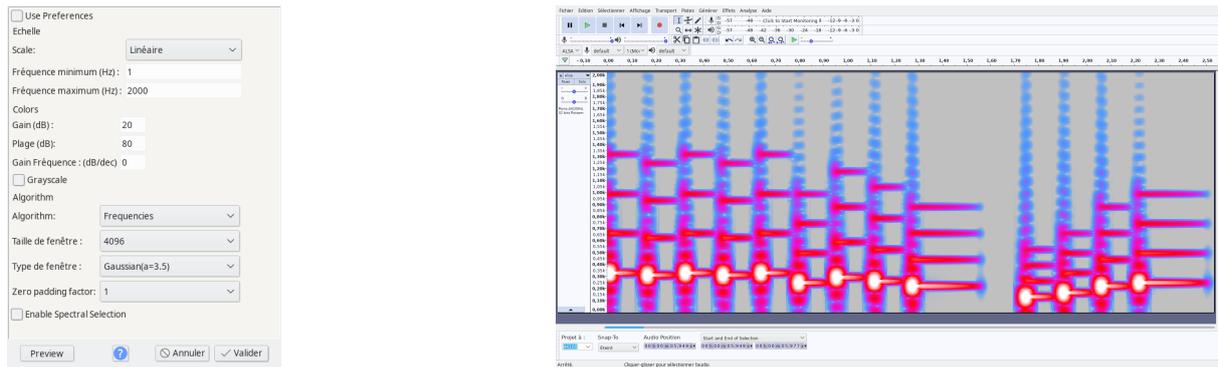


FIGURE 8 – Spectrogramme en fausses couleurs dans Audacity, et fenêtre de paramètres du spectrogramme du logiciel.

La représentation en fausses couleurs est souvent plus lisible que celle en niveaux de gris, pour distinguer les contrastes simples. Ainsi, un bruit blanc ou un bruit rose (voir section 3.2) seront très semblables dans le cas d'un affichage en niveaux de gris, mais la différence sur les basses fréquences sera beaucoup plus visible en couleur, comme illustré sur la figure 9.

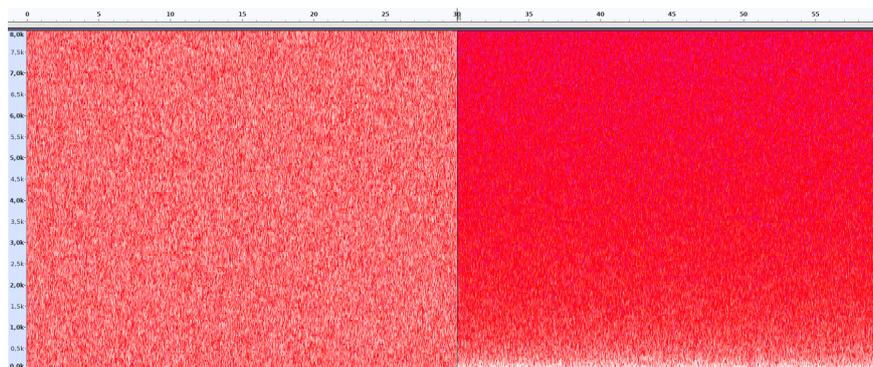


FIGURE 9 – Spectrogramme en couleur de 30 secondes de bruit blanc, et de 30 secondes de bruit rose.

4 Musique et mathématiques

L'histoire musicale est issue d'une longue tradition esthétique, propre à chaque culture, mais qui repose sur plusieurs éléments considérés comme universels, et liés à la perception que l'oreille humaine donne des sons.

L'étude de l'esthétique musicale a souvent été accompagnée d'une construction mathématique élaborée : Pythagore et la compréhension des fréquences, la construction d'une gamme, les intervalles harmoniques nécessaires à la musique polyphonique (quand plusieurs mélodies sont jouées simultanément). Dans la composition, les mathématiques ont une place importante, depuis la musique tonale jusqu'à la *musique sérielle* qui utilise comme contrainte les séries pour régir l'un ou l'autre des paramètres du son (hauteurs, rythmes, durées, timbres, etc.), ou la *musique stochastique*, qui s'appuie entre autres sur la théorie des jeux de John von Neumann.

On pourra par exemple écouter ou faire écouter [Mus14] pour une présentation vulgarisée des liens entre mathématiques et musique, et parcourir [Coh10] où l'on retrouve des articles

développant la notion d'harmoniques, de la construction d'une gamme, ainsi que nombre d'aspects mathématiques de la musique non évoqués ici.

4.1 Note

La musique occidentale s'est construite progressivement à partir de l'*échelle diatonique*, première convention de nommage des notes. Cette échelle est une gamme contenant 7 degrés, et composée de 5 tons et 2 demi-tons. Dans le système français, chaque degré ou *note naturelle* porte un nom, qui se répète de manière cyclique, du grave à l'aigu : *do, ré, mi, fa, sol, la, si*, et à nouveau *do*... Le système anglais adopte quant à lui les noms *a, b, c, d, e, f* et *g*. La figure 10 donne une représentation cyclique de l'échelle diatonique.

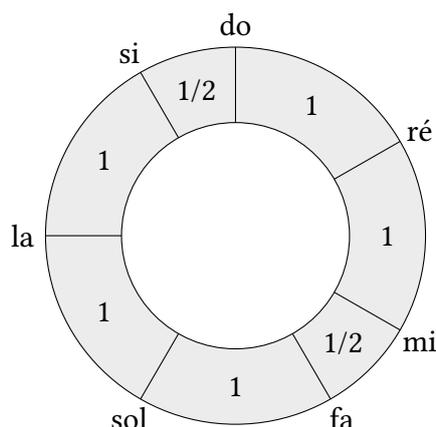


FIGURE 10 – Représentation cyclique de l'échelle diatonique, avec tons et demi-tons diatoniques.

Les fréquences de ces sept notes naturelles peuvent également être légèrement modifiées. On parle alors d'altération. Il existe trois type d'altérations simples :

- le dièse (#) est l'aténuation qui élève la note d'un demi-ton chromatique ;
- le bémol (♭) est l'aténuation qui abaisse la note d'un demi-ton chromatique ;
- le bécarre (♮) qui annule l'effet de toutes les altérations précédentes, et rend à la note sa hauteur naturelle.

La valeur d'un demi-ton chromatique est légèrement différente de celle d'un demi-ton diatonique, mais les deux sont liées par le fait qu'un ton est exactement composé d'un demi-ton chromatique et d'un demi-ton diatonique.

Le choix de la fréquence exacte des notes naturelles et des altérations correspondantes constitue l'élément central de la construction d'une gamme (voir section 4.3).

Sans entrer dans les détails, on peut également rappeler que la notation musicale contemporaine utilise une *portée*, composée de cinq lignes et de quatre interlignes, sur lesquelles on place principalement une *clé de note* et des *notes de musique*.

Il existe trois figures de clés (respectivement sol, ut et fa), qui peuvent être fixées sur n'importe quelle ligne de la portée, indiquant ainsi la position de la note correspondante (respectivement sol³, do³ et fa²).

La hauteur de chaque note est donnée par la position verticale sur la portée, chacune des lignes et interlignes correspondant à une note naturelle, les notes aigues vers le haut, les notes graves vers le bas. La hauteur des notes peut également être altérée par les aténuations #, ♭ et ♮ évoquées plus haut. Il existe également des symboles pour décrire les silences.

La figure 11 présente une portée où les notes de do³ à do⁴ sont placée après une clé de sol.

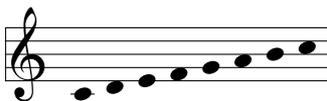


FIGURE 11 – Exemple de portée commençant par une clé de sol, avec une série de notes : do³, ré³, mi³, fa³, sol³, la³, si³, do⁴...

4.2 Octave

Nous avons tendance à percevoir comme identiques deux notes dont l'une aura une fréquence deux fois plus longue que l'autre. On dit alors que ces notes sont séparées par *une octave*. Cette propriété mène naturellement à la construction d'une échelle logarithmique, où les notes se répètent de manière cyclique, à chaque fois dans une nouvelle octave. La représentation de la figure 10 illustre la cyclicité, et la figure 12 représente cette échelle logarithmique avec la répartition des notes do par fréquence.

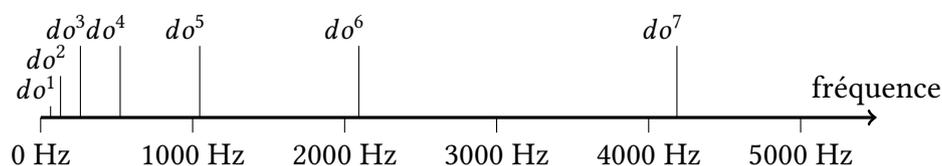


FIGURE 12 – Répartition logarithmique de la note do répartie sur chacune des octaves de 1 à 7. Fréquences empruntées à la gamme tempérée.

Pour différencier les octaves de deux notes de même nom, on indique généralement un numéro d'octave, issu d'une convention. Cette notation varie d'une culture à l'autre. En France, le changement d'octave se fait à partir du do (on passe du si² au do³), et on donne le numéro 3 au la du diapason à 440 Hz. Dans ce système, au-dessous de l'octave 1, on place l'octave -1 (il n'y a pas d'octave 0). La note la plus grave d'un piano normal (88 touches, accordées à A440) est un la⁻², correspondant dans la gamme tempérée à 27,5 Hz, ce qui atteint quasiment la limite de perception de l'oreille humaine. La note la plus aigüe de ce piano est le do⁷, correspondant à une fréquence de 4 186,01 Hz.

Certains logiciels utilisent également la notation américaine, qui fait varier le numéro des octaves de 0 à 8.

4.3 Gammes

Il existe une infinité de manières de découper une octave en intervalles plus petits. Dans ce document, nous avons commencé à évoquer le découpage en 7 notes, et les altérations associées (section 4.1). Il s'agit de considérations généralistes, quel l'on retrouve dans tous les systèmes de la musique occidentale.

Les contraintes utilisées lors de la conception d'un système de découpe de l'octave peuvent être listées ainsi³ :

- Le nombre d'intervalles doit être relativement faible, faute de quoi chacun d'entre eux est trop petit et les notes successives obtenues sont trop rapprochées pour être discernables par l'oreille.

3. Voir l'article Wikipédia consacré aux gammes musicales : https://fr.wikipedia.org/wiki/Gamme_musicale.

- La « panoplie » d'intervalles choisis doit correspondre à des notes qui peuvent être combinées (c'est-à-dire jouées en même temps en harmonie) entre elles sans irriter l'auditeur : elles doivent être le plus souvent possible *consonantes* (i.e. agréables à l'oreille).
- Les intervalles doivent, sinon être rigoureusement identiques, du moins diviser l'octave de façon suffisamment régulière pour permettre la *transposition* (pratique consistant à décaler une mélodie d'un intervalle fixe vers l'aigu ou le grave).

Dans la musique occidentale, trois types de gammes ont connu un succès important : l'accord pythagoricien, les gammes naturelles, et enfin la *gamme tempérée* à intervalles égaux. C'est aujourd'hui la gamme la plus souvent utilisée, et notamment celle qui est utilisée pour accorder les pianos.

Ces différentes constructions cherchent à optimiser les contraintes données ci-dessous, en intégrant le fait que deux rapports de fréquences sont souvent utilisés en musique occidentale :

- la *tierce*, dont la définition acoustique correspond à un rapport de $\frac{5}{4}$ entre deux notes ;
- la *quinte*, dont la définition acoustique correspond à un rapport de $\frac{3}{2}$ entre deux notes.

Une tierce ou une quinte juste ont une consonnance très forte, mais les respecter entraîne l'utilisation une division inégale de l'octave, et l'introduction d'intervalles très petits (les commas), qui complexifient énormément le système, et rendent difficile les transpositions.

La gamme tempérée la plus répandue consiste à découper l'octave en 12 intervalles égaux. À part les octaves, tous les intervalles sont acoustiquement parlant légèrement faux, mais ouvrent de nombreuses possibilités en terme de composition. Dans ce système, certaines altérations deviennent équivalentes (la #=si b, sol #=la b, fa #=sol b, ré #=mi b et do #=ré b). Le tableau 2 permet de comparer l'intonation juste comparée à celle de la gamme tempérée.

note	do	do #	ré	mi b	mi	fa	fa #	sol	sol #	la	si b	si	do
int. juste	264,00	275,00	297,00	316,80	330,00	352,00	371,25	396,00	412,50	440,00	475,20	495,00	528,00
g. tempérée	264,00	279,70	296,33	313,95	332,62	352,40	373,35	395,55	419,07	443,99	470,39	498,37	528,00

TABLE 2 – Fréquences des notes de l'octave do=264 Hz, en intonation juste, et dans la gamme tempérée.

4.4 Durée des notes

En musique, la durée des temps est le plus souvent constante, et exprimée en nombre de battements par minute. La durée des notes et des silences est quant à elle variable, et exprimée en fonction des temps.

Sur la portée, c'est la forme de chaque note qui va exprimer sa durée en fonction d'une durée de référence, donnée généralement au début de la partition. Ainsi, la notation « ♩ = 120 » indiqué au début de la partition indiquera que l'on doit jouer 120 croches (♩) en une minute. Les autres durées sont données de manière relative, en multipliant à chaque fois par deux la durée pour les noires (♩), les blanches (♪) et les rondes (♫), ou en divisant la durée par deux pour les double croches (♩♩) et les triples croches (♩♩♩). La figure 13 représente ces valeurs relatives de durée. En notation musicale, on utilise aussi le point de prolongation. Placé après la figure de la note, il permet de prolonger cette figure de la moitié de sa durée. Ainsi, ♩. = ♩♩♩.

Le découpage temporel décrit dans ce paragraphe s'utilise de deux manières différentes selon que l'on produit des *notes sans tenue* (par exemple avec un instrument à percussion), ou au contraire *avec tenue* (instruments à corde, à vent, etc.). Dans le premier cas, la position

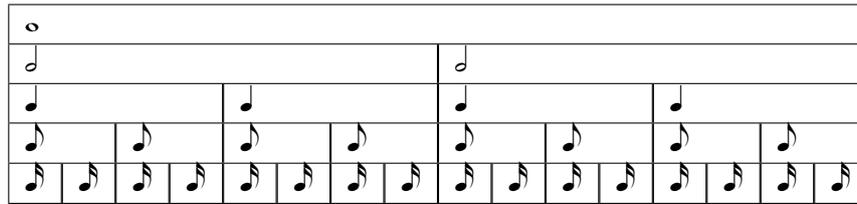


FIGURE 13 – Valeur relative des figures de notes (durées).

symbolise l'instant où la note est jouée, et la forme représente le temps écoulé avant le prochain symbole. Dans le second cas, le symbole représente l'instant où la note est jouée et le temps pendant lequel elle est tenue.

4.5 Le timbre des instruments de musique

Le son produit par deux instruments jouant une même partition peut être très différent, même si l'oreille entend qu'ils jouent la même série de sons. D'une famille instrumentale à l'autre, les caractéristiques des sons produits sont variables. On parle ainsi du *timbre* de l'instrument.

Tout d'abord, en étudiant le sonogramme d'un enregistrement musical, on constate que même lorsqu'un instrument joue une note unique, le son qu'il produit n'est pas composé d'une unique fréquence.

On observe bien sûr la fréquence de la note inscrite sur la partition, appelée *fréquence fondamentale*. Mais le son est également composé de nombreuses autres fréquences, les *partiels*. Lorsqu'un son semble musical, la plupart des partiels sont des *fréquences harmoniques*, c'est-à-dire des multiples entiers de la fréquence fondamentale. Une analyse spectrale comme celle présentée à droite sur la figure 6 permet d'observer que le piano joue une note à 824 Hz, laquelle est accompagnée de nombreuses harmoniques. À l'inverse, la voix humaine est composée de partiels qui ne se structurent pas en harmoniques (figure 6 gauche), et il est difficile d'identifier une fondamentale.

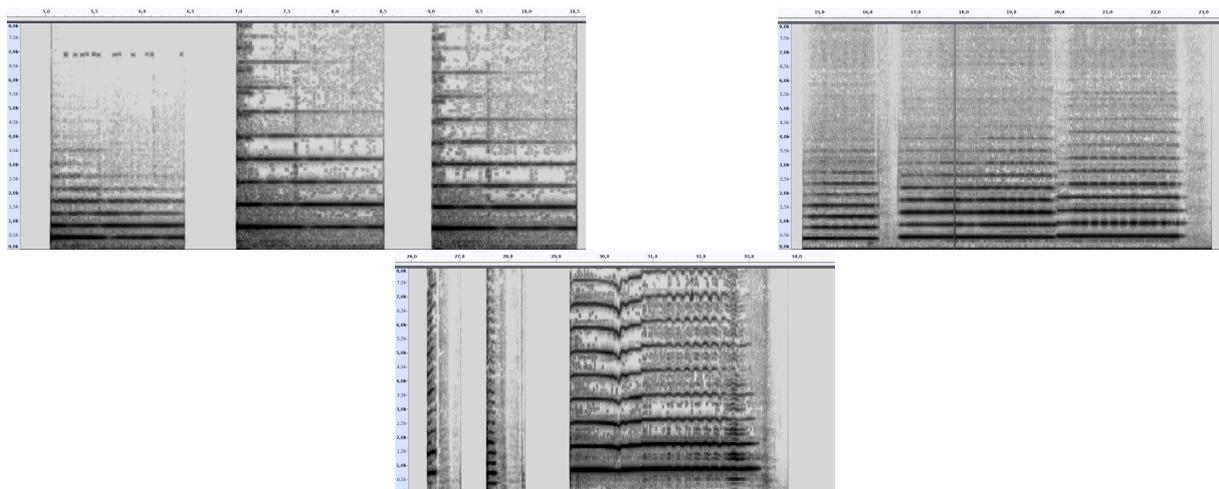


FIGURE 14 – Spectrogrammes de trois notes distinctes jouées au piano, puis de trois notes liées à la clarinette, puis de notes glissées jouées au saxophone. Logiciel : audacity.

Chaque instrument se caractérise ainsi par l'intensité des harmoniques, et par leur évolution temporelle. On observe par exemple sur la figure 14 la fondamentale (la ligne noire la

plus basse du diagramme), et les harmoniques (les lignes supérieures), dont l'intensité relative varie d'un instrument à l'autre. Dans la première illustration (piano), on observe que les harmoniques les plus aiguës s'estompent bien avant les harmoniques les plus graves. Dans la troisième illustration (saxophone), on observe également les variations de notes progressives, qui font infléchir le diagramme, les harmoniques quittant leurs lignes horizontales stables.

Cette première observation peut être complétée par un grand nombre de caractéristiques des sons produits, comme la durée de l'attaque, du corps ou de la chute du son, la variation. On pourra par exemple s'intéresser au travail de Pierre Schaeffer [Sch66], qui développe un vocabulaire très riche permettant de décrire les objets sonores : masse harmonique, grain, dynamique, allure, profil de masse, profil mélodique... Ainsi, sur la figure 14 on observe dans la dernière note de la clarinette, ou dans la dernière note du saxophone un grain, assimilable à un vibrato en musique.

Ces propriétés sont en partie des conséquences directes de techniques de jeu des musiciens, mais plus encore sont issues des propriétés physiques des instruments : famille, mécanisme de production du son, de résonance. L'attaque des sons sera par exemple différente suivant que l'on utilise une corde frottée ou pincée ; l'intensité des fondamentales sera très différente suivant la géométrie du tuyau d'un instrument à vent, qu'il soit symétrique ou conique ; le diamètre de l'étranglement de l'embouchure d'un cuivre fera varier le grain du son produit. Ainsi, sur la figure 14 on lit bien l'attaque percussive du piano, à opposer avec l'attaque lente des notes jouées à la clarinette.

En faisant abstraction du timbre des instruments, on peut obtenir une représentation équivalente à celle d'une partition d'un point de vue densité d'information, sous forme d'un *spectrogramme simplifié*. La figure 15 propose une illustration de cette représentation.

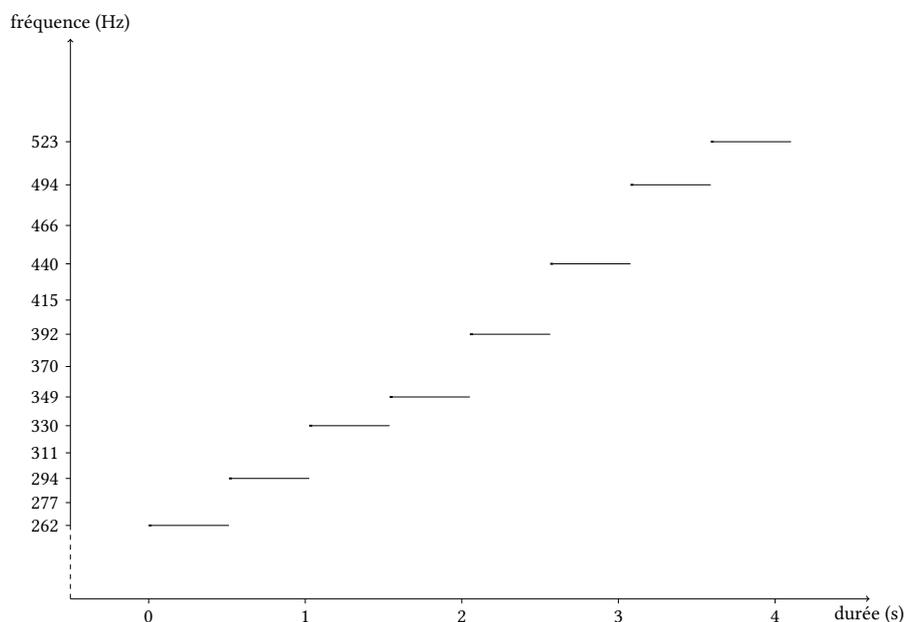


FIGURE 15 – Spectrogramme simplifié (sans timbre ni même harmoniques) équivalent à la portée de la figure 11.

5 Synthèse additive

Il existe différents moyens de produire du son de synthèse, pour imiter les instruments de musique existants, ou pour créer de toute pièce un instrument au timbre original.

Ces sons peuvent être créés de manière analogique, à partir de circuits électroniques, ou de manière numérique, en utilisant des logiciels. Parmi les grandes familles de synthèse, on peut évoquer la synthèse par modulation de fréquence, la modélisation physique, ou encore la synthèse additive.

- La *synthèse additive* est sans doute la plus simple à comprendre, mais a longtemps été difficile à utiliser, car elle nécessite de combiner de nombreux oscillateurs afin d'obtenir un son assez riche pour être satisfaisant à l'oreille. Le principe élémentaire consiste à générer autant de signaux sinus ou cosinus qu'il y a de partiels dans le son à produire.
- Le principe élémentaire de la *synthèse par modulation de fréquences* consiste à combiner deux oscillateurs, de sorte que le premier module le signal du second. On peut par exemple produire un signal $y(t) = \sin(2\pi f_p t + I_m \sin(2\pi f_m t))$ où f_p est appelée la *fréquence porteuse*, f_m la *fréquence modulante*, et I_m l'indice de modulation. L'intérêt de cette approche est qu'elle permet avec seulement deux oscillateurs de produire un son riche en partiels. Cependant, son utilisation n'est pas simplifiée par l'instabilité de l'ensemble, une petite variation d'un paramètre changeant grandement le rendu obtenu.
- La *synthèse par modélisation physique* consiste à modéliser la physique des instruments et de l'espace ambiant, puis de simuler informatiquement les vibrations et leur évolution au fil du temps, par exemple à l'aide de la méthode des éléments finis. Si cette représentation tente d'être la plus fidèle à la réalité, c'est aussi une approche qui nécessite beaucoup de calculs, et impose souvent de nombreux pré-calculs.

Dans la suite de cette section, nous proposons quelques éléments de compréhension de la synthèse additive. Une partie des idées développées dans ces lignes ont déjà été développées par son auteur dans un article de blog⁴ consacré à la synthèse additive, et s'appuie sur des expérimentations pédagogiques dont les sources en python sont disponible en ligne⁵. On peut également citer [WHC17] qui propose pages 242 à 299 un projet EPI intitulé « conception et programmation d'un synthétiseur », décomposé en huit séances allant de l'étude des sons à la fabrication d'un prototype matériel et logiciel basé sur Scratch.

5.1 Fondamentale et harmoniques

Le premier élément identifiable dans le son produit par un instrument, c'est la hauteur de la note fondamentale. Pour créer cette composante en synthèse additive, on utilise un premier oscillateur, par exemple une fonction sinus. Ainsi, pour une note dont la hauteur correspond à une fréquence f , on utilisera une fonction de la forme $t \mapsto a \sin(2\pi f t)$, où t est exprimé en seconde et a est l'intensité de la note.

Comme nous l'avons observé sur les spectrogrammes présentés dans la section 4.5, cette première note n'est jamais présente seule, et des partiels viennent enrichir le son, souvent sous forme d'harmoniques.

On se propose donc de rajouter à ce premier signal de la fondamentale des harmoniques, en adaptant leur intensité pour laisser plus de place à la fondamentale. Le son synthétisé pour une note de fréquence f et enrichi de n harmoniques s'écrit alors :

4. <http://jmtrivial.info/blog/2017/12/23/la-synthese-de-son-additive/>

5. <https://gitlab.isima.fr/jmafavre/basicsynth>

$$s(t) = a \sin(2\pi ft) + \sum_{k=2}^{n+1} a_k \sin(2\pi kft),$$

où les a_k correspondent aux intensités de chacune des harmoniques ajoutées au signal, généralement de la forme $a > a_2 > \dots > a_{n+1}$. La figure 16 présente un exemple d'un son synthétisé de manière additive avec une fondamentale et trois harmoniques dont les intensités sont données par la suite $a_k = 0.3^k$.

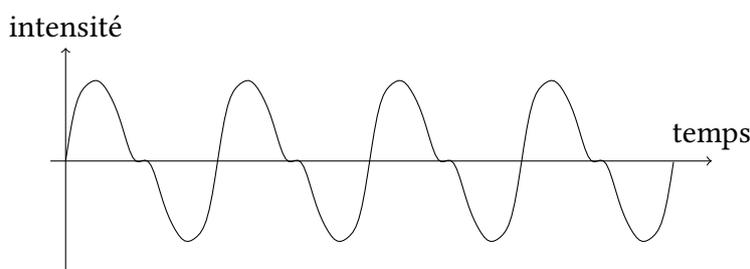


FIGURE 16 – Extrait d'un signal composé d'une fondamentale et de trois harmoniques d'intensité décroissante.

5.2 Enveloppe

Comme nous l'avons développé dans la section 4.5 consacrée au timbre des instruments, un aspect important dans l'identité sonore des instruments est l'évolution de l'intensité au fil du temps : est-ce que l'instrument a une attaque rapide ou au contraire lente, est-ce que l'instrument est percussif ou au contraire tient la note...

Pour reproduire de manière paramétrable ce comportement, les synthétiseurs les plus simples utilisent la notion d'*enveloppe* du volume. Le principe consiste à décrire l'évolution du volume de la note au fil du temps, à l'aide d'une fonction du temps, dont les quelques paramètres sont explicites d'un point de vue sonore.

L'une des enveloppes les plus simples, et très souvent utilisée utilise une fonction linéaire par morceaux, et décrivant quatre phases dans la progression d'un son (figure 17) :

- l'*attaque*, correspondant au début du son. On commence avec le volume à zéro, et on atteint plus ou moins rapidement le volume maximal.
- le *déclin*, correspondant à une baisse de volume qui suit l'attaque. Couplée à l'attaque pour produire un fort volume initial et bref, elle permet de décrire un son frappé par exemple.
- le *maintien*, correspondant au corps du son, plus ou moins long. Elle permet de décrire les instruments dont la note peut être tenue, un instrument à vent par exemple.
- la *relâche*, correspondant à la chute du son, avec un retour linéaire au volume à zéro.

Pour appliquer l'enveloppe $e(\cdot)$ à une note décrite par un signal $s(\cdot)$, il suffit de multiplier les deux fonctions : $\hat{s}(t) = s(t)e(t)$.

Plus un synthétiseur est complexe, plus il permet une description fine de l'enveloppe, et de son utilisation. Par exemple, on pourrait choisir d'appliquer une enveloppe différente à chacune des harmoniques qui composent un son.

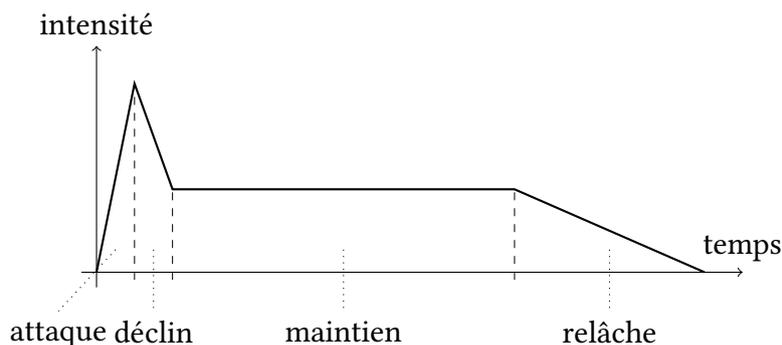


FIGURE 17 – Les quatre intervalles d'une enveloppe simple, utilisée en synthèse musicale pour simuler la dynamique d'un instrument réel.

5.3 Synthèse numérique

Dans la section 3.1 nous avons déjà évoqué le processus de numérisation d'un son. Dans le cas d'une synthèse numérique, on utilise donc un échantillonnage à N échantillons par seconde, et une représentation de l'intensité à l'aide d'un entier signé dans un intervalle $]i_{min}, i_{max}[$.

Par exemple, une sinusoïde oscillant à 440 Hz pendant une seconde sera calculée de la manière suivante :

$$s(k) = i_{max} \sin \left(2\pi \frac{440k}{N} \right), \text{ pour } k = 0, \dots, N.$$

Références

- [Coh10] Gilles Cohen. Titre : Tangente hors-série. num. 11. mathématiques & musique. editeur : Editions pôle paris, 2002 format : 17 cm x 24 cm, 80 p. bibliogr. pag. mult. issn : 1294-9949 type : périodique ou revue langue : Français support : papier. 2010.
- [FP14] Corinne Fillol and Franck Poisson. Bruit émis par les matériels roulants ferrés et pneumatiques. https://intranet.sfa.asso.fr/archives/J80-VISIBLE/presentations/S3_Poisson_Fillo1.pdf, 2014.
- [Gri94] A. Gribenski. *L'audition : « Que sais-je ? » n° 484*. Que sais-je? Presses Universitaires de France, 1994.
- [Mus14] France Musique. La musique et les mathématiques. <https://www.francemusique.fr/emissions/le-dossier-du-jour/la-musique-et-les-mathematiques-17588>, Novembre 2014.
- [Sch66] Pierre Schaeffer. *Traité des Objets Musicaux*. 1966.
- [WHC17] David Wilgenbus, Mathieu Hirtzig, and Claire Calmet. *1, 2, 3... codez! : six projets d'informatique pour le collège (cycle 4)*. 2017.