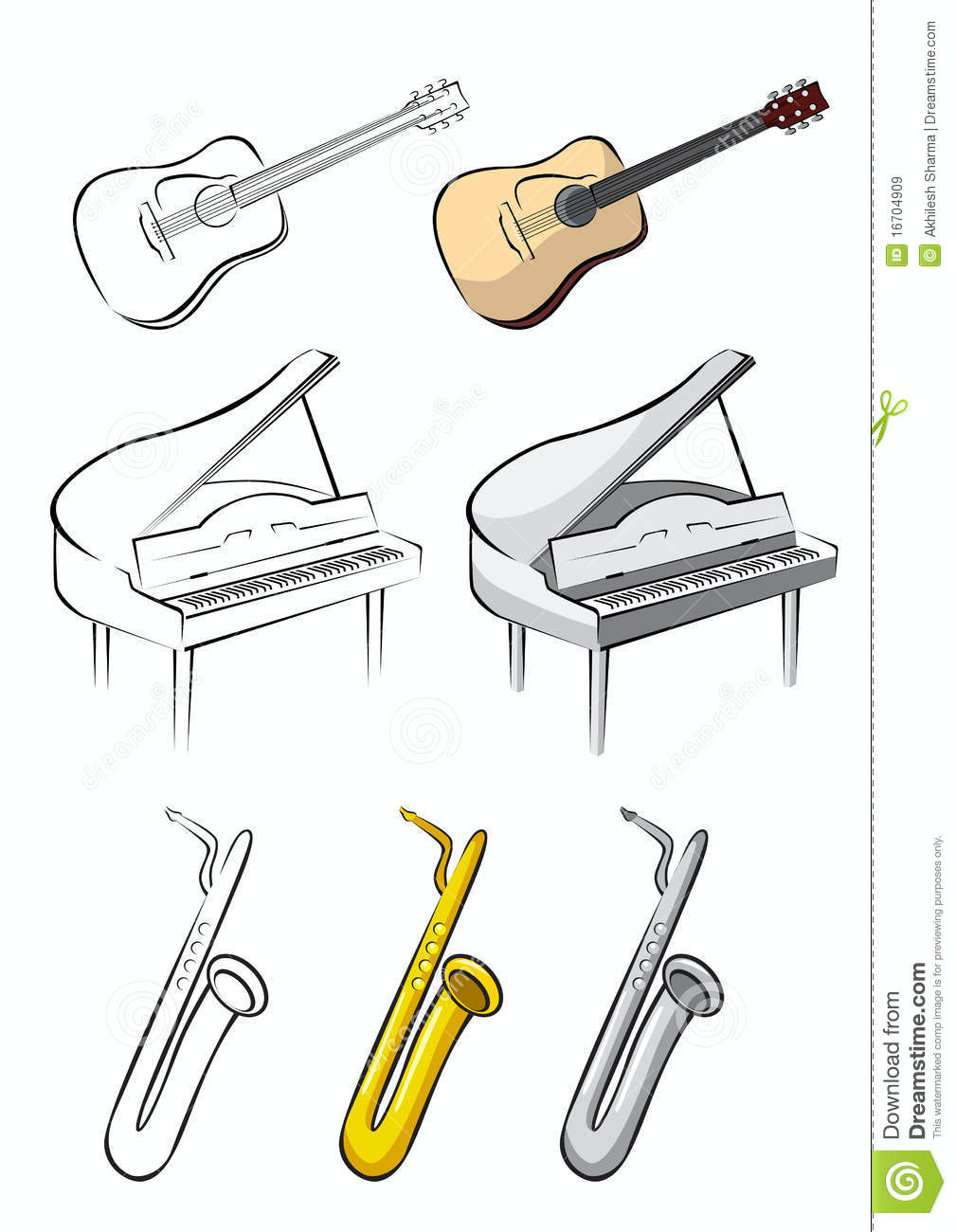
PROJET ANNUEL IABD

**Sujet** : Classification d’instruments (Piano, Saxophone, Guitare)



**Participants** : KHAOUANI Ahmed / RENOIR Marvin / CLODION Quentin

**Rapport intermédiaire** : 28/04/2020

L’objectif du projet étant porté sur la classification de son provenant de trois instruments, le piano, la guitare et la saxophone. L’élément qui a influencé notre choix est la popularité des instruments, car le but étant d’avoir possibilité de récupérer suffisamment de matière afin de constituer un Dataset suffisant riche et diversifié.

Il faut savoir que ce sont trois instruments de familles distinct (léger rapprochement entre le piano et la guitare)

**Piano** : Famille des instruments à cordes

**Guitare** : Famille des instruments à cordes pincées

**Saxophone** : Instruments à vent de la famille des bois

Après avoir choisi nos trois instruments, il fallait trouver un choix de constitution de notre dataset. Nous avons donc décidé de récupérer un maximum de piste audio pour chacun des instruments et d’ensuite de réaliser des cuts toutes les 2sec. Pour cela nous avions trois choix qui s’offrait à nous comme source de données : Youtube, Spotify, Soundcloud

Nous nous sommes concentrés dans un premier temps sur SoundCloud avant de se rendre compte qu'un trop grand nombre de piste audio comprenait des droits d’auteur dont il fallait s'acquitter. Nous nous sommes donc tournée vers Youtube.

Afin de récupérer nos pistes audio nous avons eu l’idée de créer un scrapper en

Python qui serait capable de récupérer les X premières vidéos d’une recherche Youtube. Ce procédé est très intéressant mais malheureusement il comporte deux problèmes majeur. Le premier étant son temps de réalisation, le second est que nous n’avons pas la visibilité de ce que le scrapper va récupérer. En tapant sur youtube “Piano” nous pouvons avoir comme résultat des tutos, des présentations de piano etc.. ce qui ne nous intéresse pas du tout. Nous avons donc décidé au final de récupérer à la main des vidéos Youtube qui nous semblaient les plus pertinentes.

Pour récupérer les vidéos nous avons mis en place un système permettant d’avoir un Dataset le plus diversifié possible.

* Pour chaques instruments, identifié trois sous familles.
* Pour chaque sous famille, prendre en compte trois types d'enregistrements (studio, webcam, en public).
* Pour chaque types d'enregistrement de chaques sous familles, récupérer trois vidéos d’au moins 5 minutes chacune.

En respectant ce système de sélection, nous avons donc atteint un total de 81 vidéos.

Après avoir récupérer toutes ces vidéos il a fallu faires un traitement car elles ne sont pas directement exploitables.

**1ère étape** : Identifier les passages de la vidéos à exclure (voix, générique, blanc...)

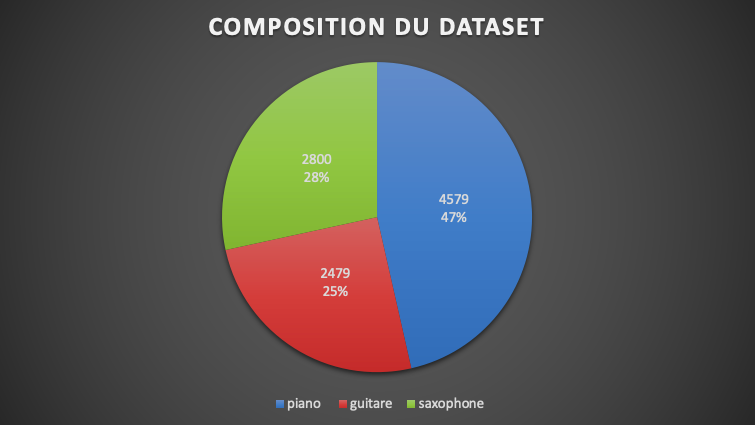
**2ème étape** : Réaliser des cuts de 2 sec sur chaque vidéos, ce sont ces cuts qui composeront le Dataset. Pour réaliser les cuts, CLODION Quention a développé un application qui permet de réaliser des cuts de la durée que l’on souhaite sur une piste audio.

**3ème étape** : Exclure les cuts de 2 sec qui correspondent aux passage inutilisable repérer lors de la première étape.

Afin d'accroître notre efficacité nous avons chacun choisit un instrument :

* ***KHAOUANI Ahmed*** *: Piano*
* ***CLODION Quentin*** *: Guitare*
* ***RENOIR Marvin*** *: Saxophone*

Voici un schéma représentant notre Dataset composé d'extraits audio de 2 sec :



Nous avons à l’heure actuelle, 3 pistes pour la formation du dataset. Ces 3 pistes ont été implémentés pour que l’on puisse rapidement adapter le dataset en fonction de la piste la plus concluante.

Des tests sur les datasets sont actuellement en train d’être réalisés en python avec Tensorflow.

1 - Utiliser les données des fichiers wav :

La première idée et la plus simple consiste à utiliser les données d’un fichier .wav de 2 secondes en mono et avec une fréquence de 20000HZ pour diminuer la taille des données en entrées. On pense que cette solution n’est pas pertinente et ne permettra pas d’obtenir les meilleurs résultats possibles.

2 - Transformation de Fourier :

La seconde hypothèse consiste à utiliser la transformation rapide de Fourier pour récupérer l’amplitude associée à chaque fréquence de l’enregistrement. Cette technique devrait plutôt bien fonctionner car la quantité des données utilisées en entrée du modèle seront bien moins astronomiques.

3- Spectrogramme :

Enfin la dernière idée consistait à utiliser le spectrogramme qui comporte les mêmes données que le spectre de Fourier mais cette fois-ci, en fonction du temps. Aujourd’hui, on ne pense pas que cette donnée (le temps) sera vitale. De plus, cela ajouterait un paramètre à prendre en compte ce qui allongerait le temps de calcul.

Ces 3 idées peuvent aussi être traiter comme de la classification d’image en prenant respectivement, le spectre d’un fichier audio, le spectre de Fourier du morceau et enfin, le spectrogramme. Cela ne nous semble pas particulièrement être une bonne idée dû au manques de détails de la photo. Mais cette piste a tout de même été envisagée par le passé.

Music Creator :

Cette application a évolué au fur et à mesure du projet pour aujourd’hui, permettre la modification en masse de fichiers pour constituer le dataset.

Au départ, il s’agissait d’une application qui permettait de comprendre les mécanismes des fichiers .wav en proposant la lecture et la réécriture de fichiers sans s’appuyer sur des frameworks existants.

Par la suite, on a commencé à ajouter plusieurs opérations sur cette application tel que splitter les canaux d’un fichier wav pour le découper en plusieurs fichiers, découper un fichiers en plusieurs fichiers plus petits ou superposer des morceaux sur un même fichier.

Cette application nous a donc permis dans un premier temps, de mieux comprendre le fonctionnement d’un fichier .wav pour par la suite, pouvoir les manipuler plus facilement si besoin.

Etant donné que l’application commençait à avoir de nombreuses fonctionnalités, on a décidé de l’adapter en winform pour pouvoir l’utiliser plus simplement pour la construction du dataset.

Aujourd’hui on l’utilise donc pour écrire le contenu des fichiers .wav dans des fichiers .csv, scinder les canaux, changer la fréquence d’un morceau ou bien couper un morceaux en différents morceaux d’un durée donnée.