

Je vous présente le travail effectué durant mon stage dans le but d'étudier la similarité acoustique entre différents instruments et différents instrument avec différents techniques de jeux en utilisant la représentation scattering.



# Sommaire

Je commence par une introduction, puis je présente l'algorithme de calcul, je l'explique un peu, je présente les résultats et je finis en posant le problème que j'essaye de résoudre maintenant.



# Introduction

## Motivations

Le web devient surchargé cordialement avec un nombre énorme de fichier audio dont un très grand nombre contient des fichiers musicale ; d'où la nécessité de les organiser d'une façon automatique ; et d'examiner les informations qui rends la recherche plus facile(exemple : la recherche de la partie joué par un saxophone dans une grande partition) et enfin pouvoir définir les descripteurs qui contribuent à définir le timbre de chaque instrument.

## Objectives

l'objective de ce stage est de trouver l'espace de descripteurs le plus approprié pour effectuer une discrimination entre les sons et que aidera à étudier la dis-similarité acoustique entre instruments et instruments+ technique de jeux.



# Algorithme

La procédure expérimentale est comme suit : je commence avec une base de données de 25119 fichier classé par instruments qui peuvent avoir des altérations comme avec ou sans sourdine par exemple, et par technique de jeux ; trémolo, vibrato, etc j'effectue ensuite l'extraction de descripteurs soit en utilisant MFCC soit en utilisant le scattering et on applique à ces descripteurs différents traitements ; puis je passe à un algorithme de metric learning le LMNN on est toujours dans un espace euclidienne, on calcule les distances deux à deux et enfin on utilise des paramètres de classements pour comparer les résultats.



# Extraction des descripteurs

les deux méthodes d'extractions de descripteurs que j'ai testé sont les MFCC et le scattering pour les MFCC beaucoup d'étude on été déjà effectuer et les paramètres optimales sont connus : un espace de 12 descripteurs calculé pour une fenêtre temporelle de 25ms ; avant de commencé l'étude j'ai testé différents paramètres et l'état de l'art à été validé avec ces même paramètres.

Par contre pour la représentation scattering est relativement nouvelle et le choix des paramètres n'est pas alors évidents. Notre but est alors de trouvé l'espace de descripteurs le plus compacte qui donne les meilleurs résultats avant d'appliquer la méthodes de metric learning. l'avantage de scattering est dans ces ordres supérieurs( pour l'audio c'est l'ordre 2). pour les MFCC on est obligé de prendre un fenêtre temporelle petite ce qui implique une perte d'information des structures temporelles de grande ordre mais pour les MFCC on peut prendre une fenêtre temporelle grande puisque dans les ordres supérieurs les informations seront ajouté.



## exemple de scattering

Deux exemples des descripteurs qu'on cherche à trouver pour effectuer la dissimilarité : dans la première figure on peut voir que l'ordre 1 de scattering il n'y a pas d'information sur l'attaque temporelle, mais dans l'ordre deux on peut voir que pour une certaine fréquence l'attaque est présent. de même pour le vibrato et plusieurs autres phénomènes liés à la technique de jeu.



# Traitement des descripteurs

## MFCC

Pour le traitement de descripteurs : dans les deux cas de MFCC et de scattering j'ai appliqué une standardisation.

## Scattering

et pour le scattering ; Vincent Lostanlen qui est un doctorant travaillant beaucoup sur le scattering m'a proposé d'éliminer les descripteurs qui cumulent une somme de variance inférieur de la somme cumulé de variance et de diviser par la médiane pour rendre les descripteurs symétrique et plutôt gaussienne. Dans mon cas j'ai trouver des meilleurs résultats en éliminant les grandes variances qui cumule une variance supérieur à 83%



# LMNN

Le but de LMNN (Large Margin Nearest Neighbors) est d'avoir une plus grande précision pour une classification basé sur les  $K$  plus proche voisins, On cherche alors avec les LMNN que les  $K$  plus proche voisins pour chaque observation appartiennent à la même classe, et d'introduire une marge entre chaque observation et toute autre observation n'appartenant pas à la même classe, dans l'exemple ici on peut voir que pour l'observation  $x_i$  on approche les trois voisins jaune appartenant à la même classe et on éloigne les individus n'appartenant pas à la classe jaune d'une certaine marge. (c'est un cas optimale en pratique on essaye de ce rapprocher de ce cas le plus possible) le problème est alors résolu en trouvant un espace de mahalanobis ou ces contrainte sont respecté ou bien en projetant vers un espace en multipliant les observations dans leur espace de descripteurs par une matrice  $L$  qui va nous projeté dans un espace toujours euclidienne et c'est ce qu'on cherche. Cela est fait grâce à une optimisation pénalisé contenant de terme de pénalisation le premier pénalise la grande distance entre les  $k$  plus proches observations de même classe et l'autre pénalise les petite distance entre les observations de différentes classe qui entre dans la marge.





# Résultats 1/2

## Ranking metrics pour 16 classe d'instruments

Déscripteurs	Traitements	MAP	Pat5	Temps
MFCC	Brut	22.66	85.93	77.10
MFCC	Stand.	24.22	86.89	67.33
MFCC	LMNN	23.80	86.77	744.96
scattering	Brut	16.49	70.40	148.96
scattering	Stand.	19.31	78.98	149.85
scattering	variance	30.73	93.87	147.08.00
scattering	IMNN	<b>60.07</b>	<b>99.99</b>	25849.98

## Ranking metrics pour 32 classe d'instruments avec variations

Déscripteurs	Traitements	MAP	Pat5	Temps
MFCC	Brut	20.84	83.98	58.11
MFCC	Stand.	22.29	85.12	55.56
MFCC	LMNN	22.85	84.47	1467.53
scattering	Brut	14.89	67.05	98.32
scattering	Stand.	17.52	75.13	97.21
scattering	variance	28.07	90.94	97.24
scattering	IMNN	<b>56.88</b>	<b>99.92</b>	23290.21



## Résultats 2/2

### Ranking metrics pour 498 classes d'instruments avec différentes techniques de jeux

Déscripteurs	Traitements	MAP	Pat5	Temps
MFCC	Brut	8.24	43.95	50.72
MFCC	Stand.	8.78	45.19	50.74
MFCC	LMNN	9.19	45.22	12367.32
scattering	Brut	6.46	37.07	96.35
scattering	Stand.	10.69	47.01	98.16
scattering	variance	20.41	57.98	94.22
scattering	IMNN	<b>44.42</b>	<b>88.08</b>	68156.67

### Ranking metrics pour 16 classes d'instruments avec division en train/test

Déscripteurs	MAP	MAP +LMNN	P@5	P@5 +LMNN
MFCC 80/20	23.24	24.54	73.90	75.14
MFCC 50/50	22.74	24.02	81.17	81.86
scat. 80/20	31.00	60.70	83.05	98.44
scat. 50/50	30.70	62.98	90.02	99.31

## Prochain travail

Toute cette partie à été bien résolu avec les résultats comme on l'a vu ensemble, pour le moment le problème à résoudre c'est de considérer un vérité terrain formé par plusieurs labeling donné par plusieurs utilisateur alors les labeling de la classe ne sont pas consistant il sont dans la plus part de cas divergents et il faut essayer de les combiner d'une façon à respecter le maximum tous les vérité terrain avec la même amplitude.

