



Nouveau paradigme d'étiquetage de morceaux dédié à la génération de listes de lecture musicale

Thèse d'Informatique

Groupe de Travail du Thème « Son et Interaction »

Directeurs de Thèse :

Pierre Hanna
Matthias Robine

Doctorant et présentateur :

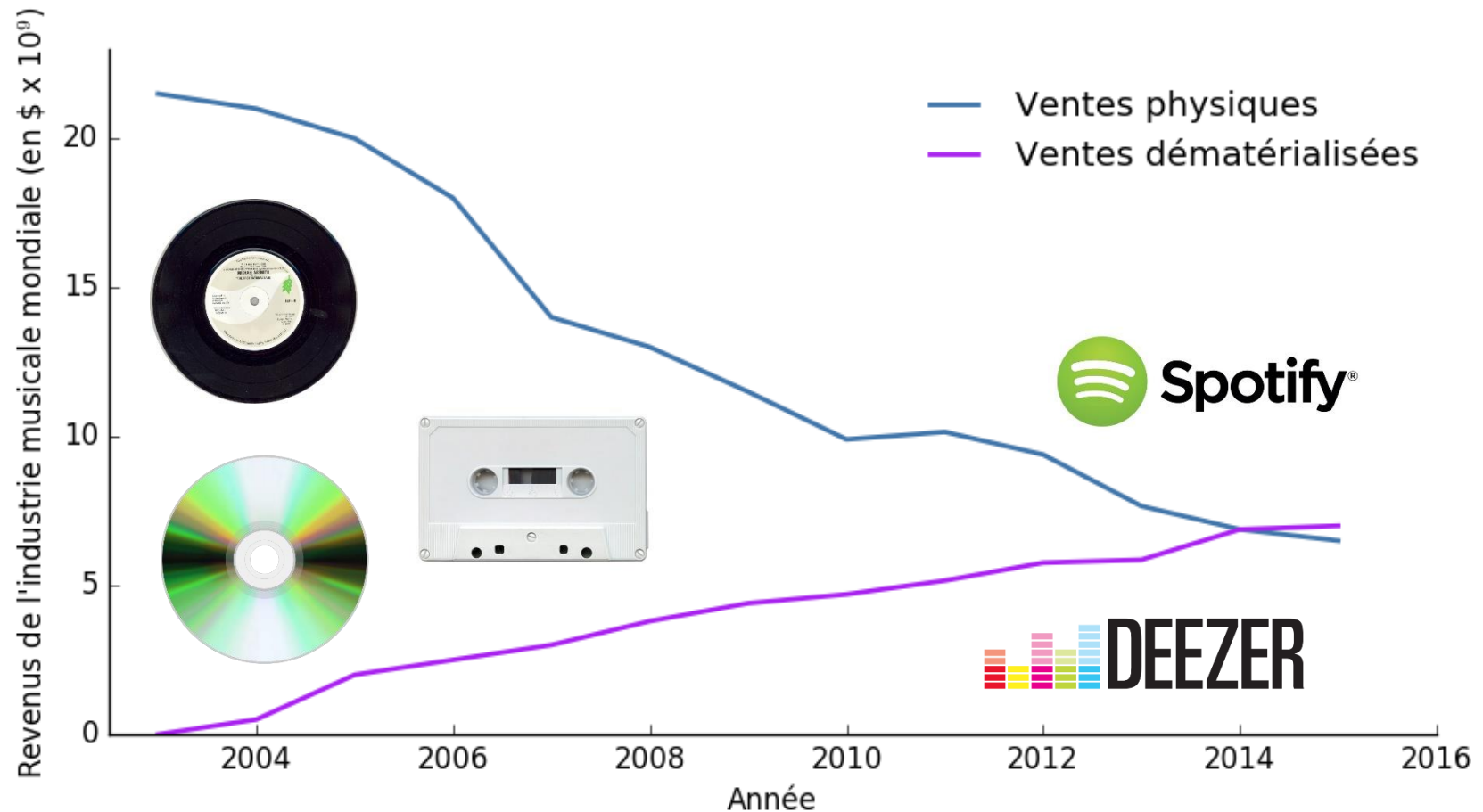
Yann Bayle

Financement :

Bourse blanche MENRT

Jeudi 24 Novembre 2016

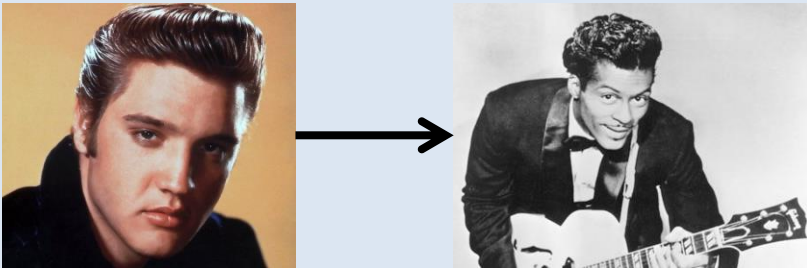
Musique



Source : IFPI

Streaming

Recommandation



Listes de lecture

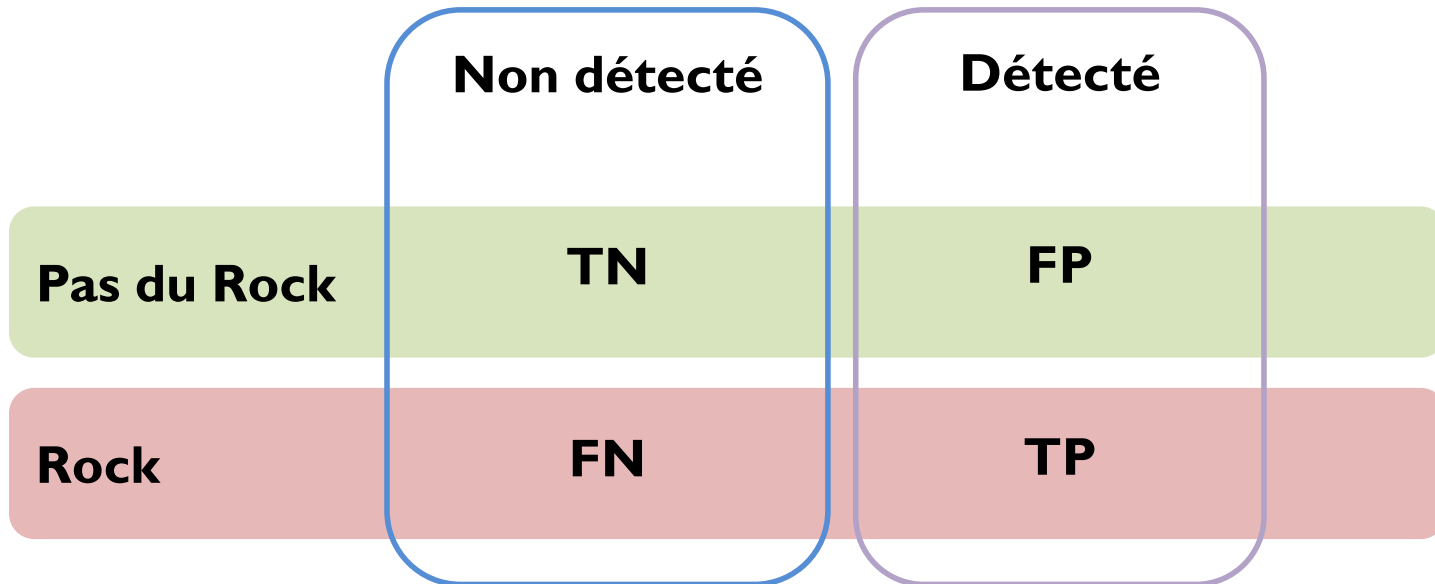
- ♪ **Genre** (Rock, Blues, ...)
- ♪ **Humeur** (Joie, Nostalgie, ...)
- ♪ **Activité** (Sport, Travail, ...)
- ♪ **Top 100**
- ♪ **À thème** (« Obama », ...)

Marquer les chansons

Marquage

Méthode	Avantages	Inconvénients	Exemples
Manuelle (éditeur)	Précis	Peu	PANDORA®
Manuelle (communauté)	Nombreux	Incorrect Ambigu	 
Automatique (goût utilisateurs)	Précis	Couverture	 
Automatique (marquage automatique)	Couverture	Précis	

Métriques

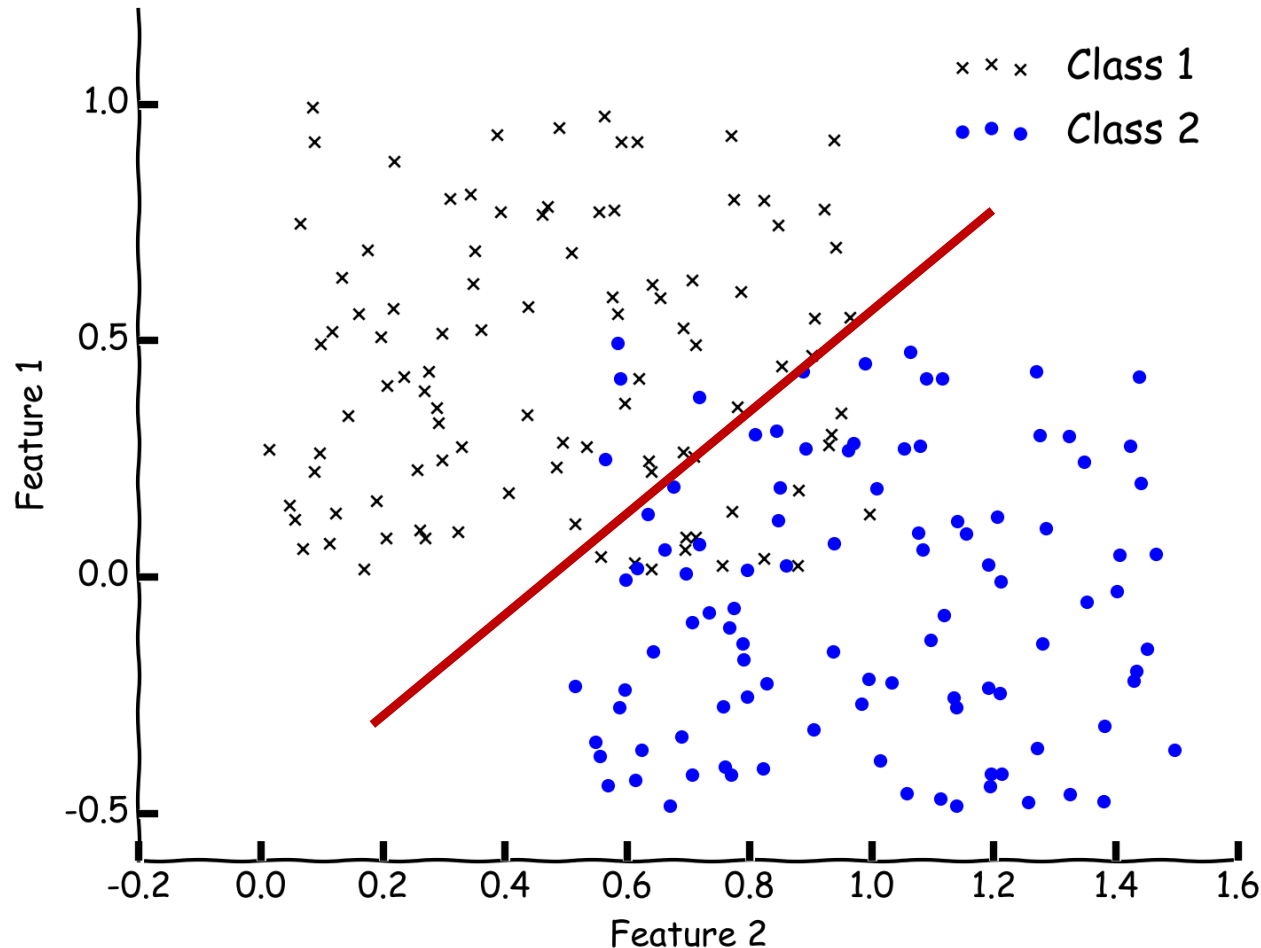


$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F-Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

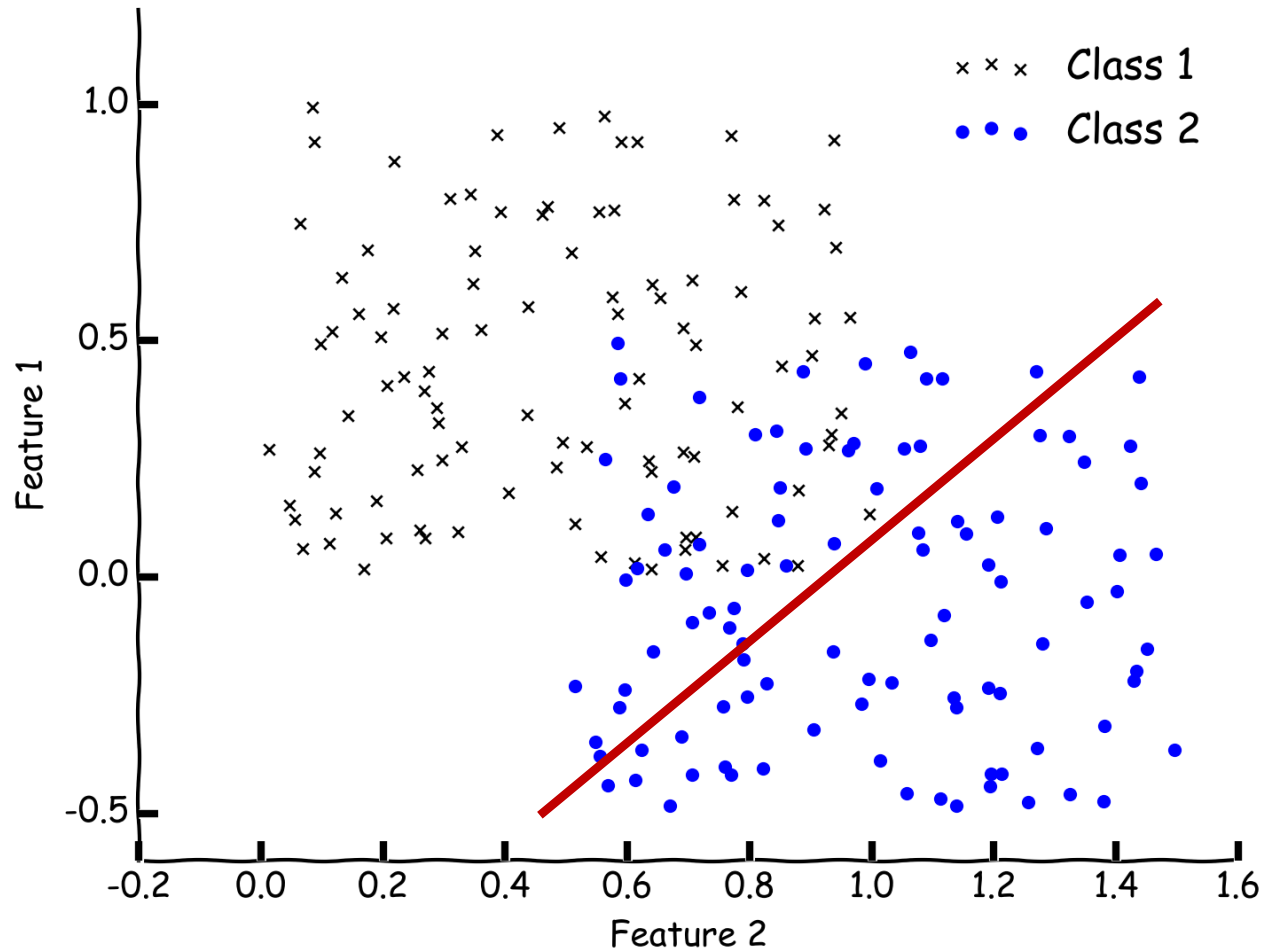
Métriques



Maximisation de la
F-Measure
pour les 2 classes

80 % Précision
80 % Recall

Métriques



Maximisation de la
Précision
pour la classe 2

100 % Précision
60 % Recall

Tâche

- ♪ Au sein d'une base de données musicales
- ♪ Identifier (et étiqueter) les morceaux qui :
 - ♪ contiennent du chant
 - ♪ sont purement instrumentaux
- ♪ Définitions
 - ♪ Voix = Voice
 - ♪ Chant = Singing Voice
 - ♪ Chanson = Song
 - ♪ Morceau (titre, pièce ?) instrumental = Instrumental

Existant

- ♪ Ghosal et al. 2013
- ♪ 95 % de précision
- ♪ Base de données musicale équilibré de 540 ...
- ♪ ... extraits de 30 secondes
- ♪ Morceaux inconnus
- ♪ 5-folds Cross-validation

Existant

♪ Gouyon et al. 2014

♪ 3 algorithmes

♪ Cross validation (CV)

♪ 3 datasets

♪ CAL500 58 non-vocal and 444 vocal (2-folds CV)

♪ MSD24k (30s snippet) 531 non vocal (3-folds CV)

♪ MagTag5k 2349 (3-folds CV)

♪ Dispo chercheurs

♪ Seulement features

♪ Plus disponible

♪ Annoté et vérifié à la main

Existant

- ♪ AcousticBrainz (non publié)
- ♪ 5-folds cross-validation
- ♪ Non reproducible

voice_instrumental

In-house MTG collection

Use: classification into music with voice/instrumental

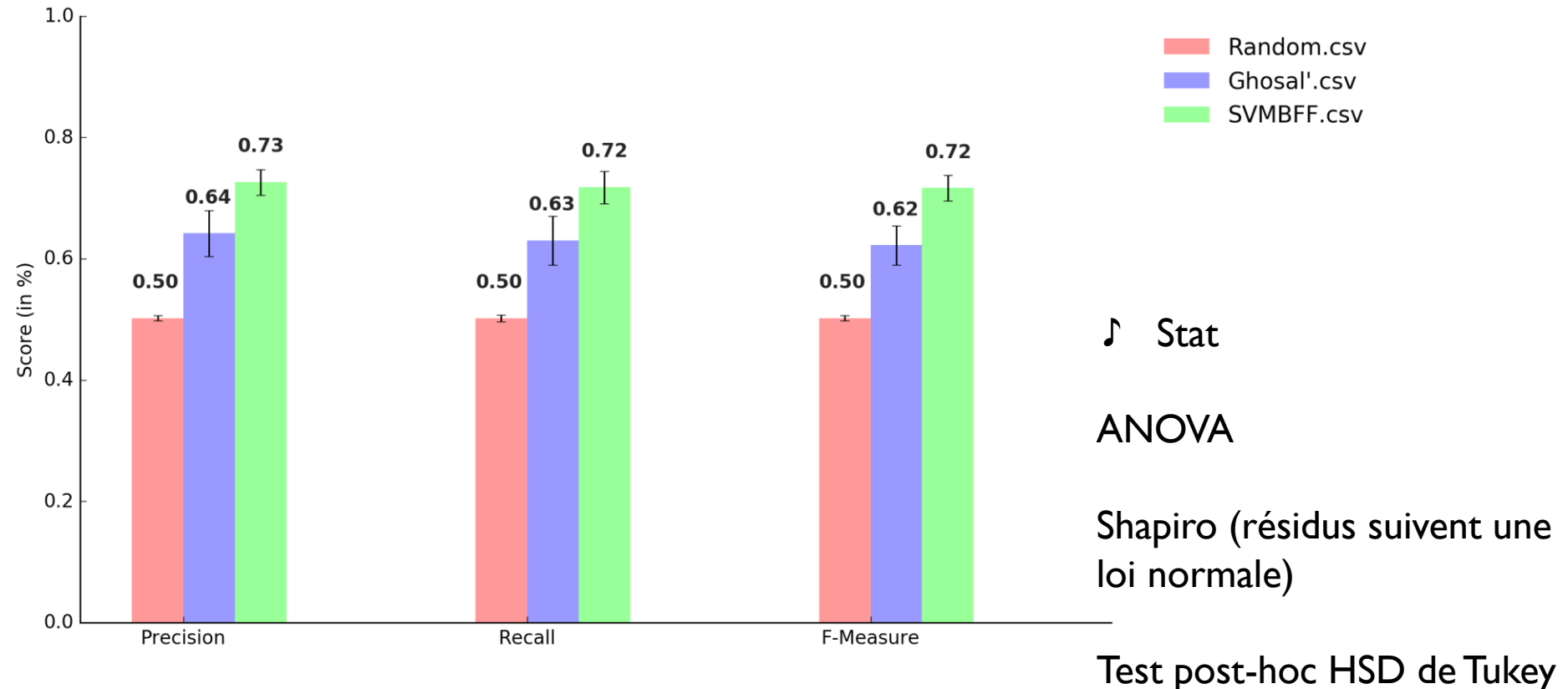
Size: 1000 track excerpts, 500 per class

Accuracy: 93.800000%

Predicted (%)					Actual (%)
	instrumental	voice		Proportion	
instrumental	94.20	5.80	instrumental	50.00 %	
voice	6.60	93.40	voice	50.00 %	

Résultats de l'existant

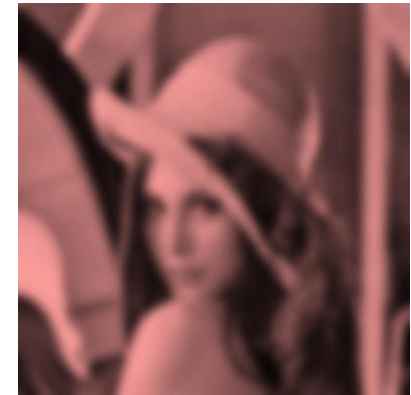
- ♪ Réimplémentation et récupération des algorithmes existant
- ♪ BDD 200 morceaux open-source et annotés par la communauté



Problème I

- ♪ Comment augmenter la base de données musicale ?
- ♪ Bases actuelles pas représentative

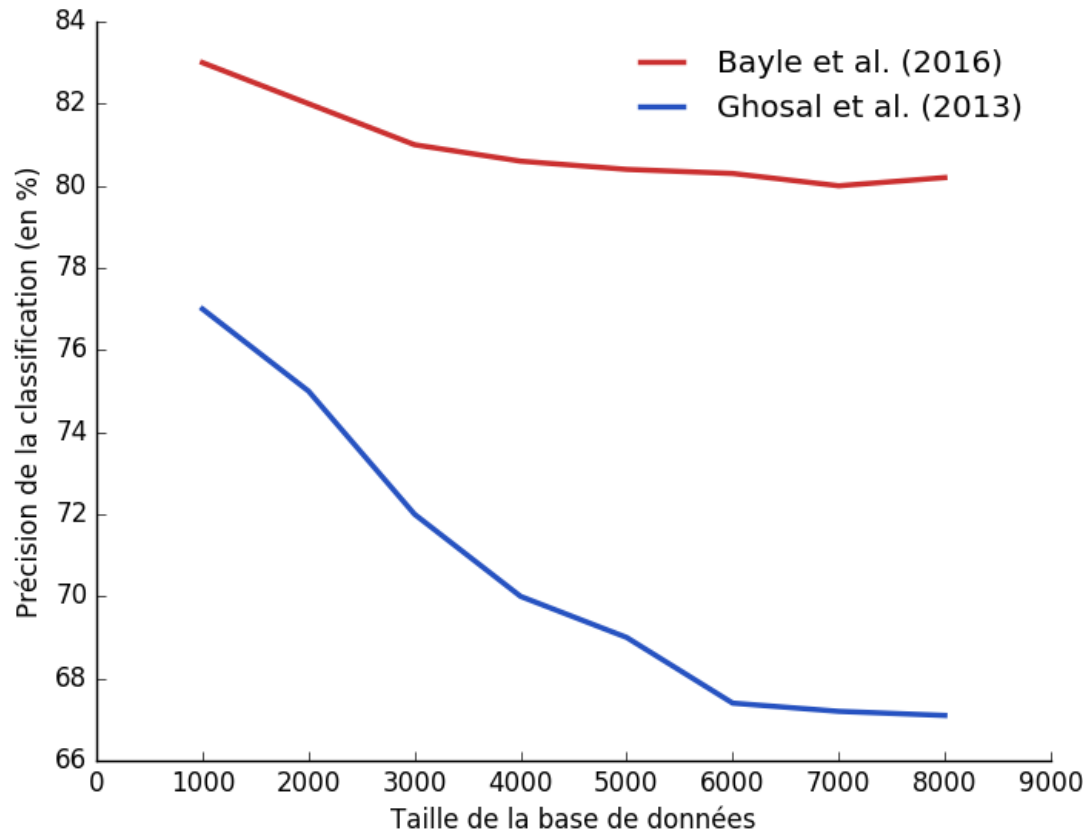
« Data augmentation »



J. **Schlüter** and T. Grill,
“Exploring data
augmentation for improved
singing voice detection with
neural networks,” in *ISMIR*,
2015.

Article publié

Y. Bayle, P. Hanna, M. Robine, “**Classification à grande échelle de morceaux de musique en fonction de la présence de chant**” in *Journées d’Informatique Musicale*, 2016, vol. 1, pp. 144-152.



Problème I bis

♪ Comment diversifier la base de données musicale
et
avoir la base de vérité correspondante ?

Morceaux

Reproductibilité (tendre vers Réplicabilité)

- ♪ ISRC (ex : GBAAM0201110)
- ♪ International Standard Recording Code
- ♪ L'ISBN de la musique



- ♪ Plusieurs millions de titres référencés par un ISRC
- ♪ Détails des résultats pour chaque morceau !

Base de vérité



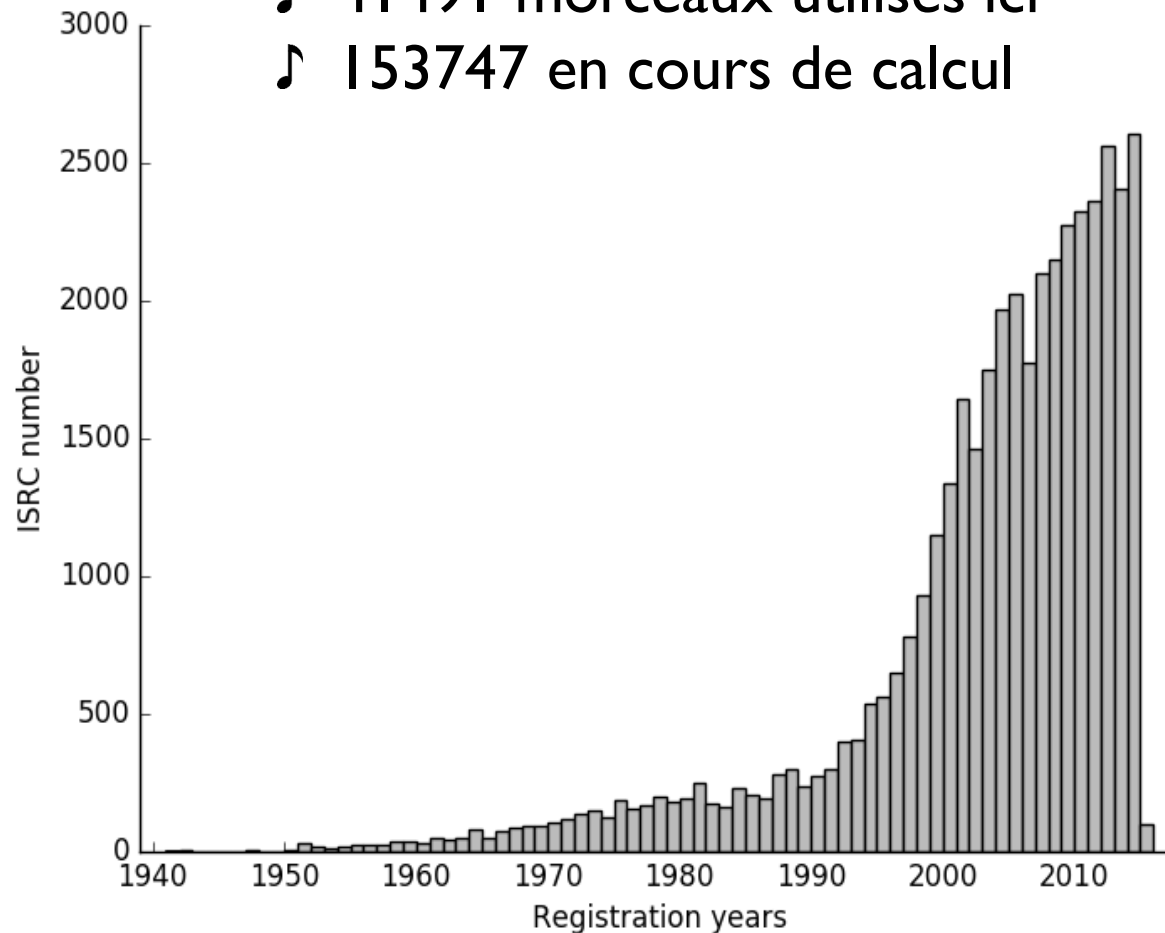
musixmatch



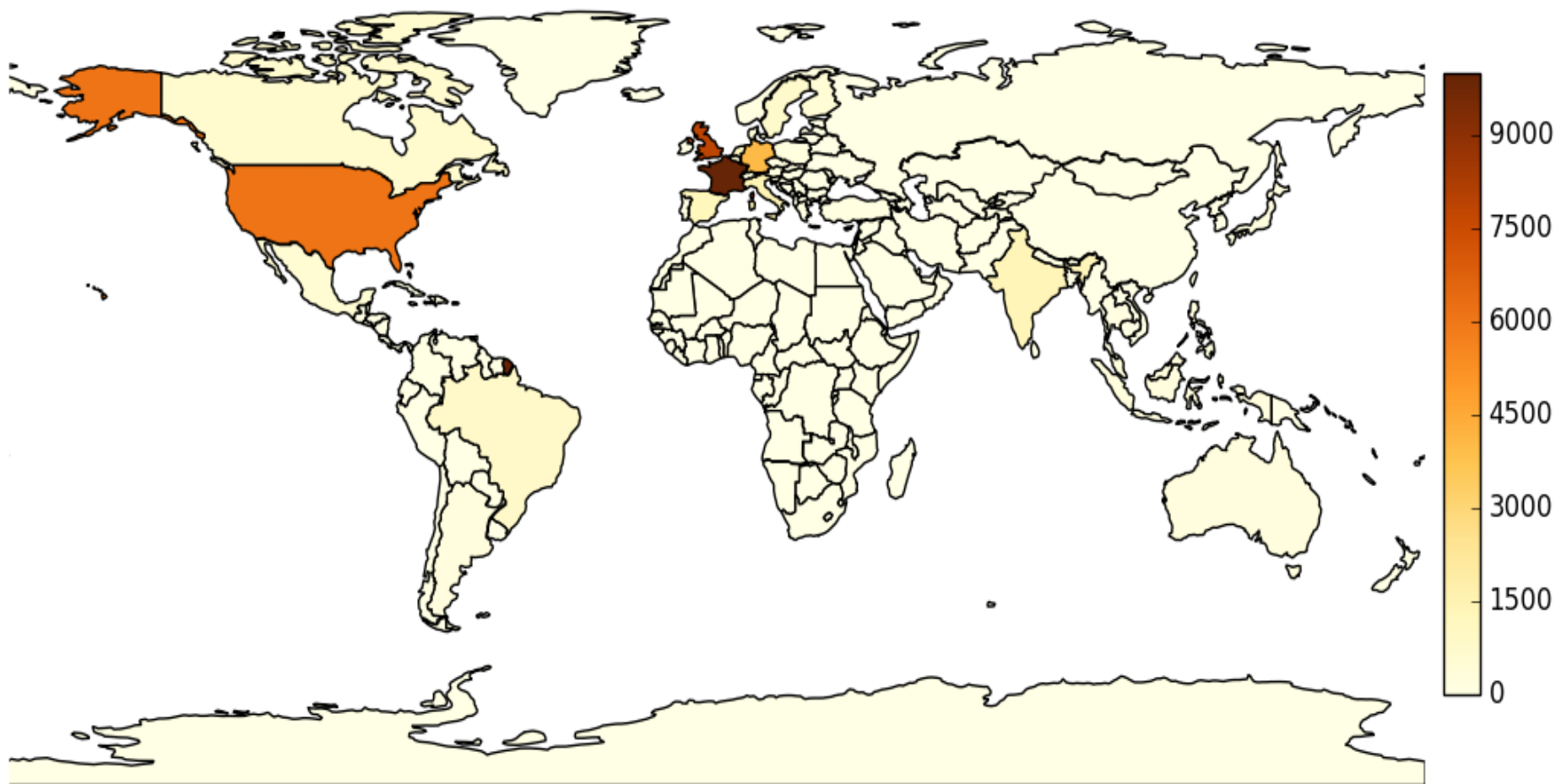
- ♪ Paroles
- ♪ ~30 millions de titres référencés
- ♪ +14 millions de paroles de chansons (58 langues)
- ♪ Plusieurs millions de titres référencés par un ISRC
- ♪ Exemple Joe Satriani – Super Colossal – Crowd chant

Base de données musicale

♪ 41491 morceaux utilisés ici
♪ 153747 en cours de calcul



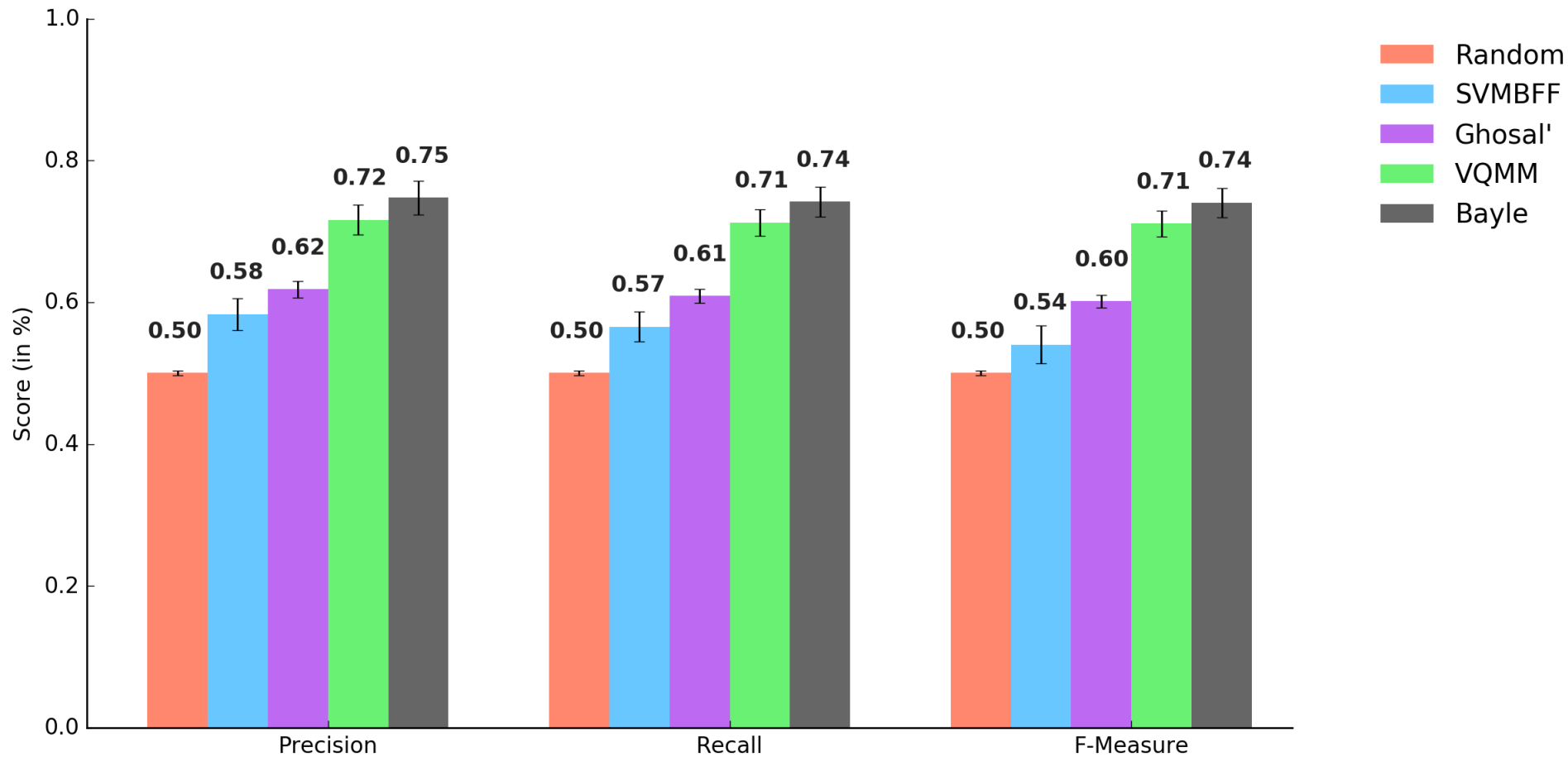
Base de données musicale



Problème 2

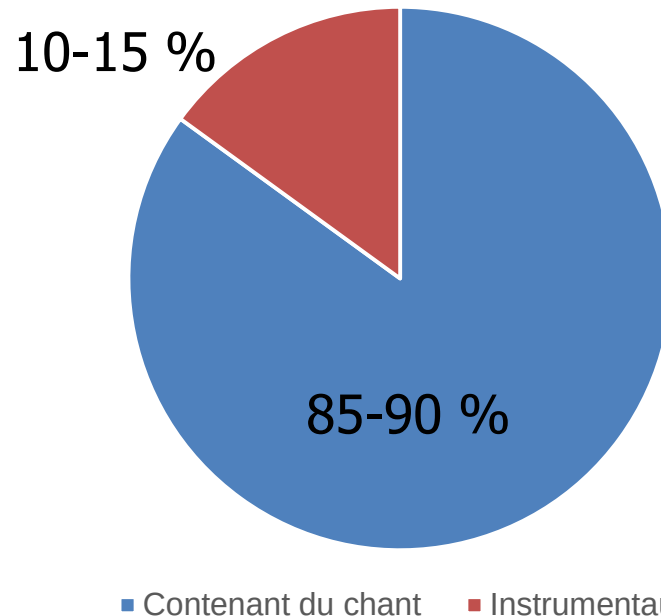
- ♪ Quelles bases d'apprentissage et de test ?
- ♪ Jusqu'à présent N-folds cross validation
- ♪ Mais contraintes pour l'utilisation d'une grande base de train :
 - ♪ Temps de calcul SVM
 - ♪ Besoin de base de vérité locale seulement disponible pour quelques centaines de chansons
- ♪ Base de train 200 morceaux
- ♪ Base de test 41491 morceaux

Résultats globaux base équilibrée



Problème 3

Répartition des morceaux



♪ 41491 morceaux utilisés

♪ 4456 Instru

♪ 37035 Song



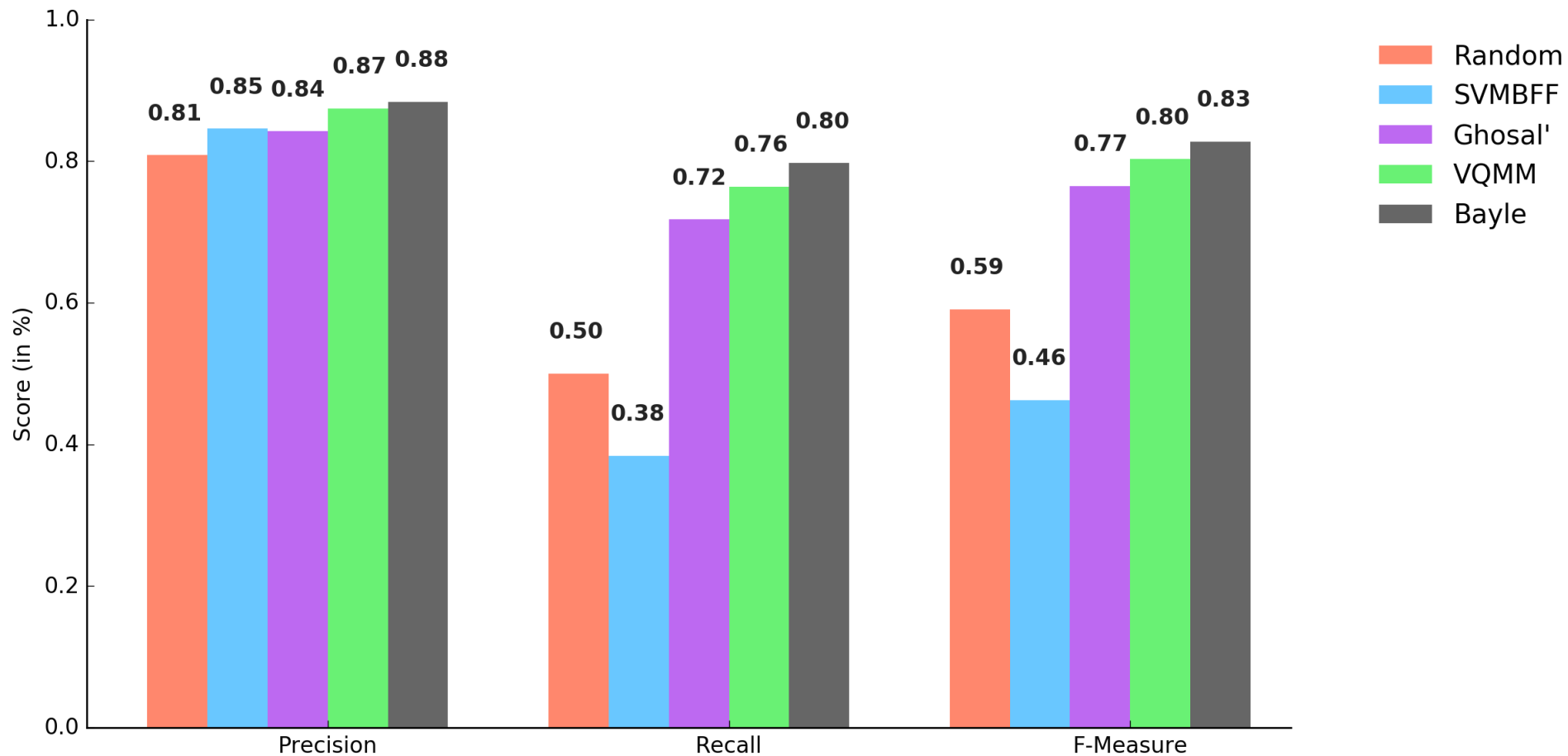
PANDORA



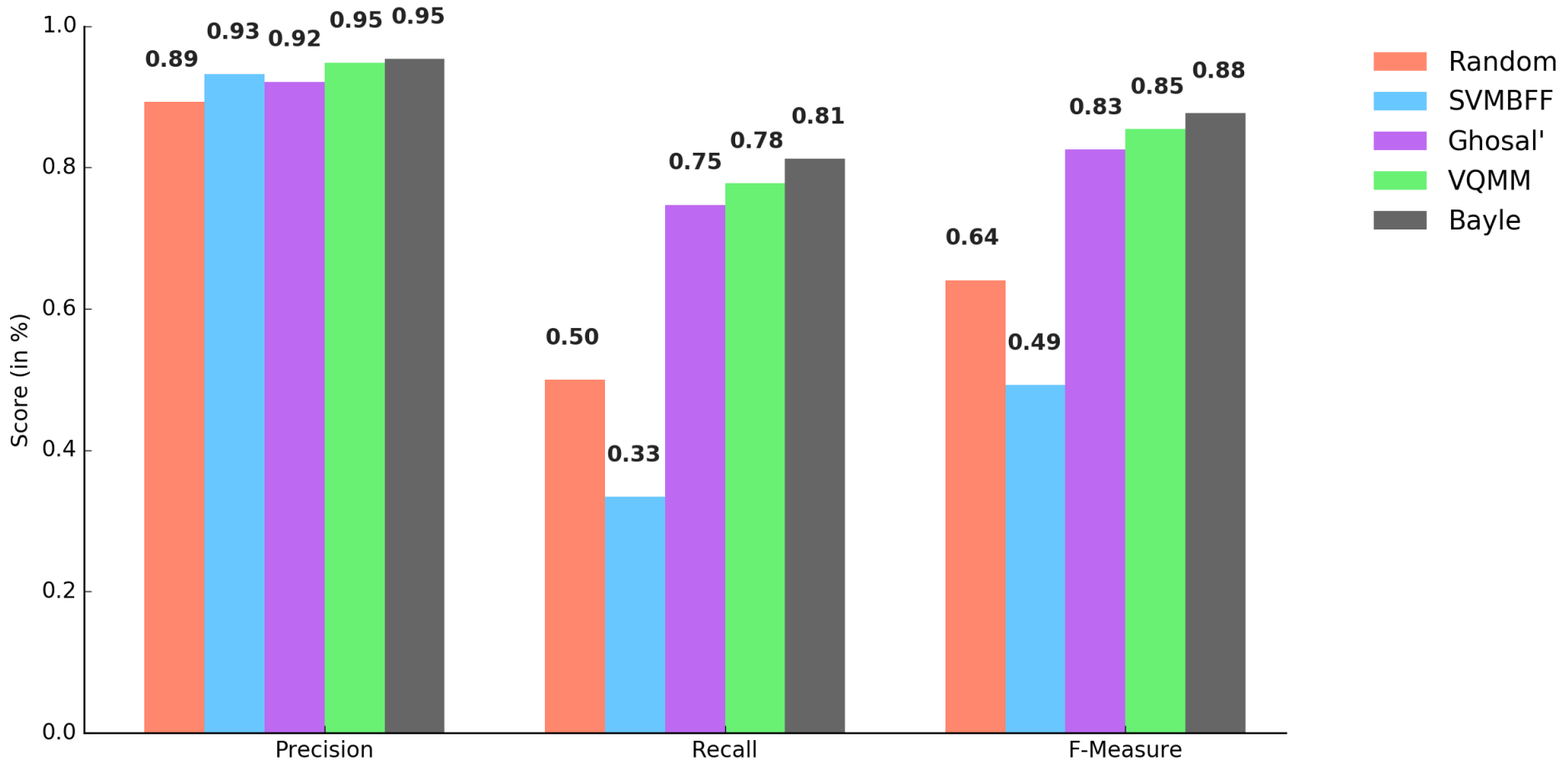
last.fm



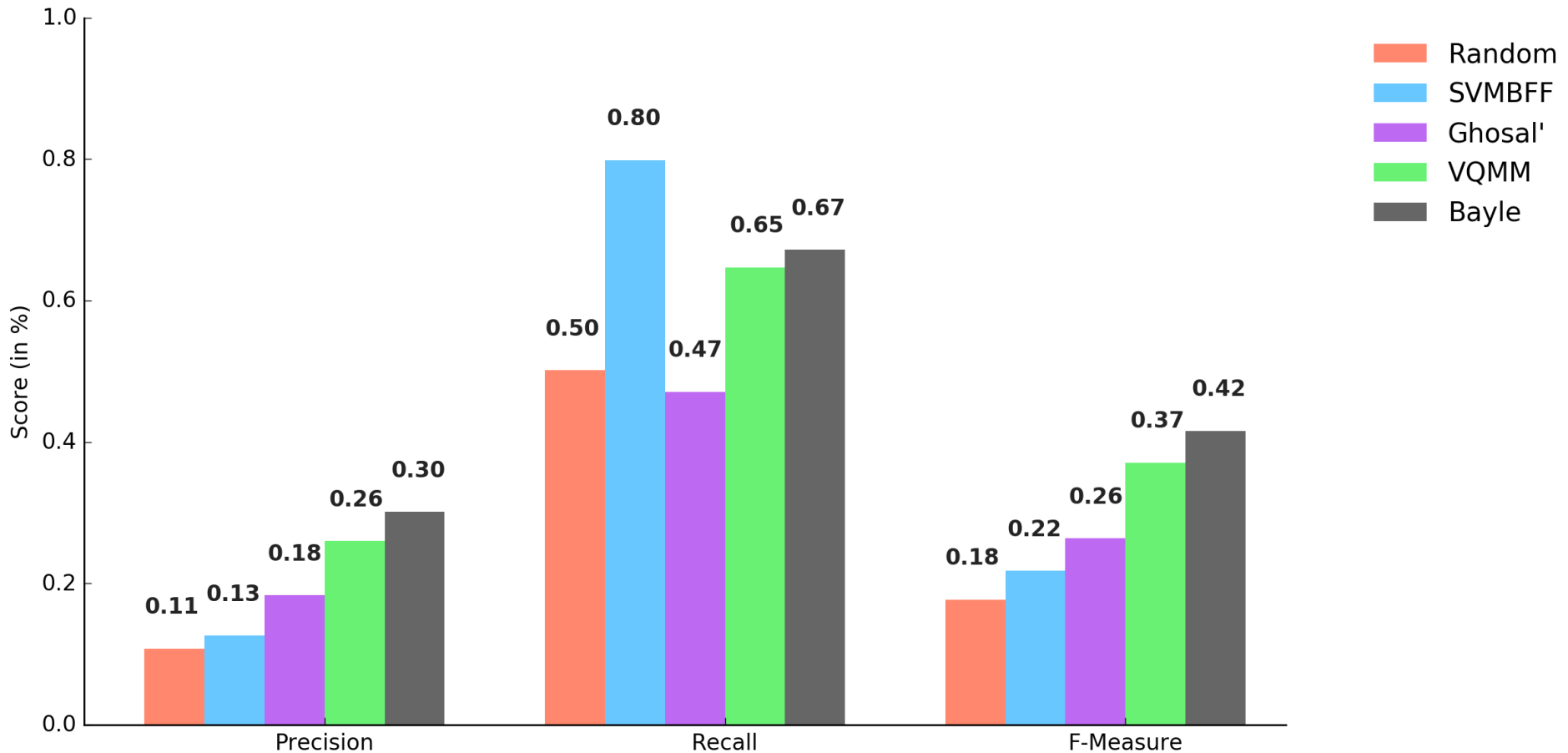
Résultats globaux base réelle



Résultats base réelle pour Chansons



Résultats base réelle pour Instrumentaux



Problème 4

- ♪ Pour une grande BDD, les algorithmes existants
 - ♪ voient leurs métrique (précision/recall) chuter

Problème 5

- ♪ imposent un temps de calcul trop important
- ♪ Ex : Apprentissage SVM $O(n_features * n_observations^2)$
- ♪ 48 ans sur ma machine (8 cœurs 16Go RAM)

Objectif

- ♪ Étiqueter les morceaux Chantés/Instrumentaux
- ♪ Générer liste de lecture musicale parfaite

Solution

- ♪ Playlist parfaite
 - ♪ On ne veut pas optimiser la f-measure
 - ♪ Optimisation de la précision
 - ♪ Voire assurer 100% précision et donc maximiser Recall
- ♪ Traiter chaque classe indépendamment

Détail et comparaisons des algorithmes

♪ Ghosal et al. 2013

♪ Gouyon et al. 2014

♪ SVMBFF

♪ VQMM

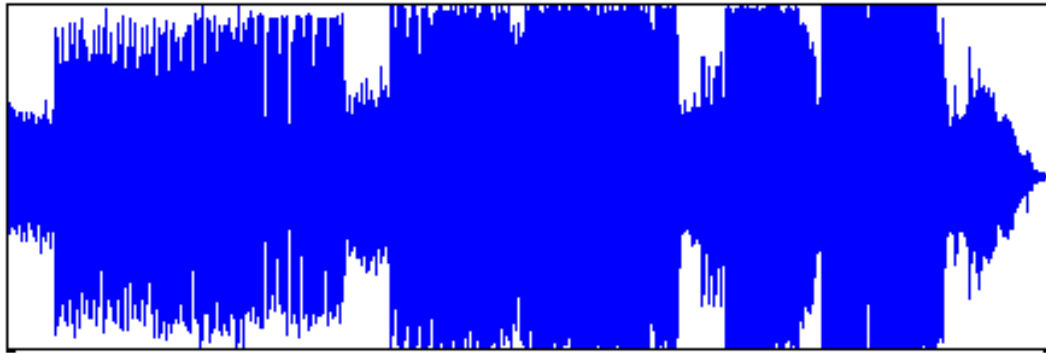
♪ Bayle

♪ Bayle et al. 2016

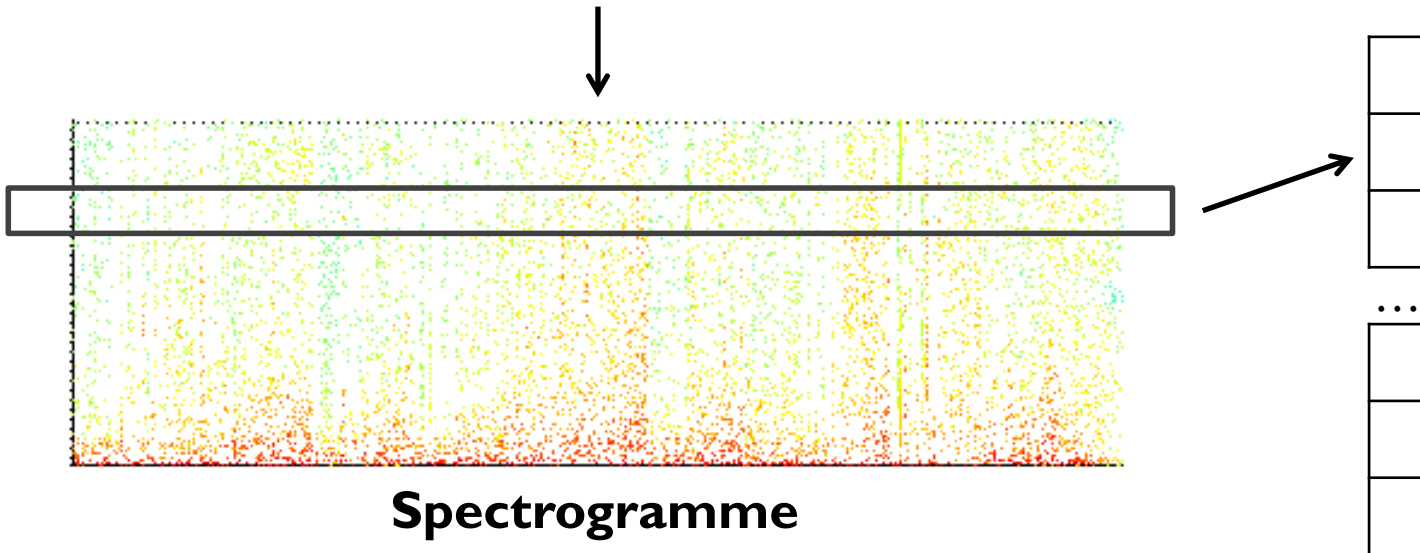
♪ Nouvel algorithme ?

Ghosal et al. 2013

Signal audio



**Vecteur de
coefficients MFCC
13 features par chanson**



Spectrogram

Gouyon et al. 2014 (SVMBFF)

- ♪ Même features que Ghosal et al. 2013 avec en plus
 - ♪ Moyenne et variance des MFCC
 - ♪ Taux de passage à zéro du signal temporel
 - ♪ Spectral centroid (brightness)
 - ♪ Roll-off (Upper Frequency bound for 95% of energy)
-
- ♪ 68 features par chanson
-
- ♪ Classification à l'aide de SVM

Gouyon et al. 2014 (VQMM)

- ♪ 13 MFCC par trames de 93ms
- ♪ Construit un signal de MFCC *via* un vector quantization (VQ) pour chaque chanson
- ♪ Construit un codebook des VQ pour toute la base
- ♪ Estime les probabilités conditionnelles avec un Modèle de Markov du premier ordre

Singing Voice Detection

Regnier, L., & Peeters, G. (2009, April). Singing voice detection in music tracks using direct voice vibrato detection. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 1685-1688). IEEE.

Nwe, T. L., & Li, H. (2008, March). On fusion of timbre-motivated features for singing voice detection and singer identification. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on* (pp. 1685-1688). IEEE.

Nwe, T. L., Shenoy, A. P., & Li, H. (2008, March). On fusion of timbre-motivated features for singing voice detection and singer identification. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on* (pp. 1685-1688). IEEE.

**40 % de Faux Positifs !
sur les pistes instrumentales**

Lu, S., & S. (2008, March). On fusion of timbre-motivated features for singing voice detection and singer identification. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on* (pp. 1685-1688). IEEE.

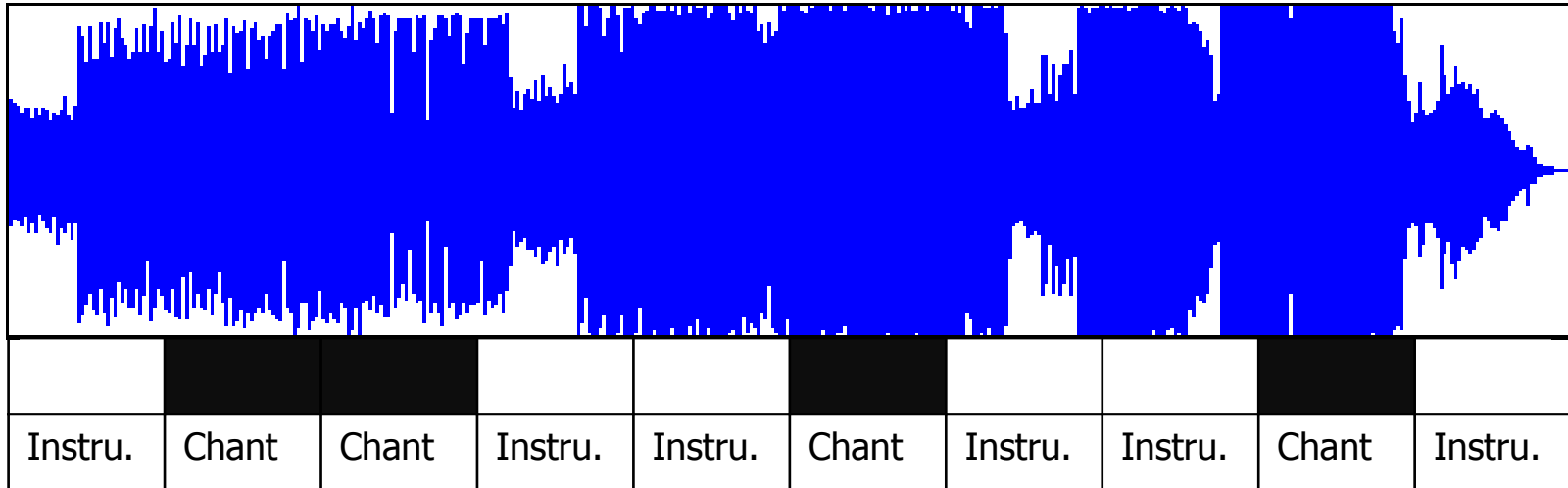
Shenoy, A. P., & Vang, Y. (2005, July). Singing voice detection for karaoke application. In *Visual Communications and Image Processing 2005* (pp. 596028-596028). International Society for Optics and Photonics.

Rao, V., Ramakrishnan, S., & Rao, P. (2009). Singing voice detection in polyphonic music using predominant pitch. In *INTERSPEECH* (pp. 1131-1134).

Lehner et al. 2014

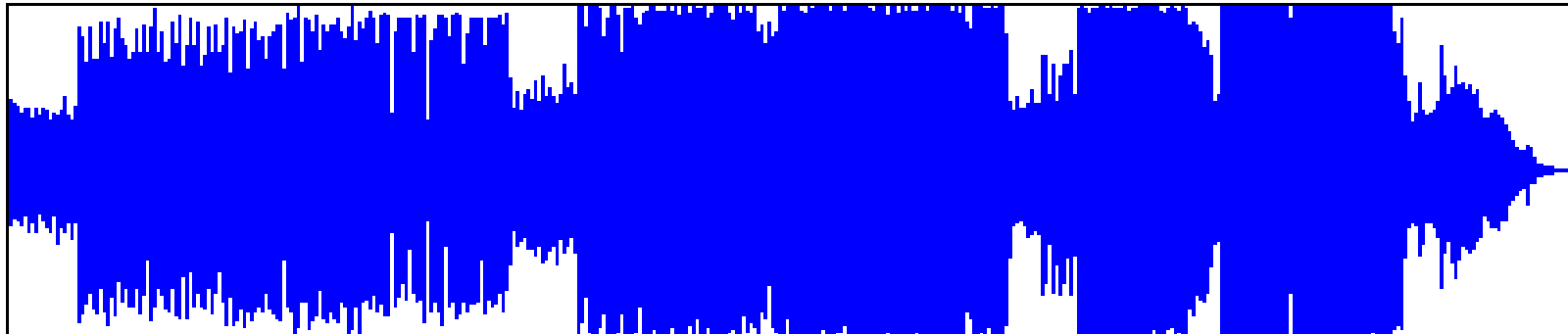
- ♪ Extraction des features pour chaque trame :
 - ♪ 60 MFCC features (30 MFCC + corresponding deltas)
 - ♪ 17 Spectral Flatness means
 - ♪ 17 Spectral Contraction variances
 - ♪ 17 Fluctogram variance features
 - ♪ Coefficient de fluctuations des demi-tons à l'aide d'une corrélation croisée
 - ♪ 5 Vocal Variance features
 - ♪ Coefficient de variation à long terme du contenu fréquentiel
- ♪ 116 features par trame de 200 ms

Prise de décision



Classification de la chanson

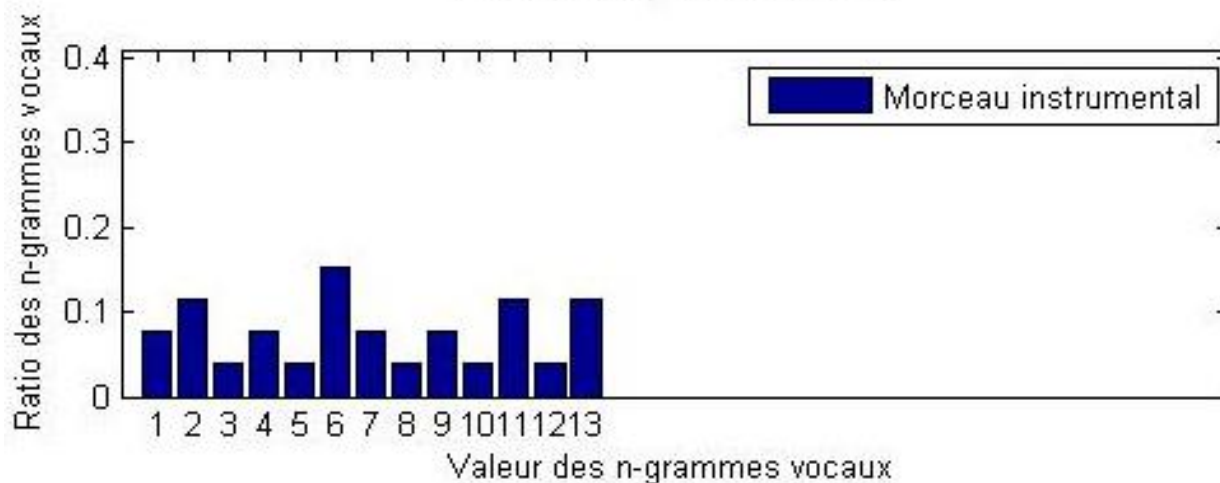
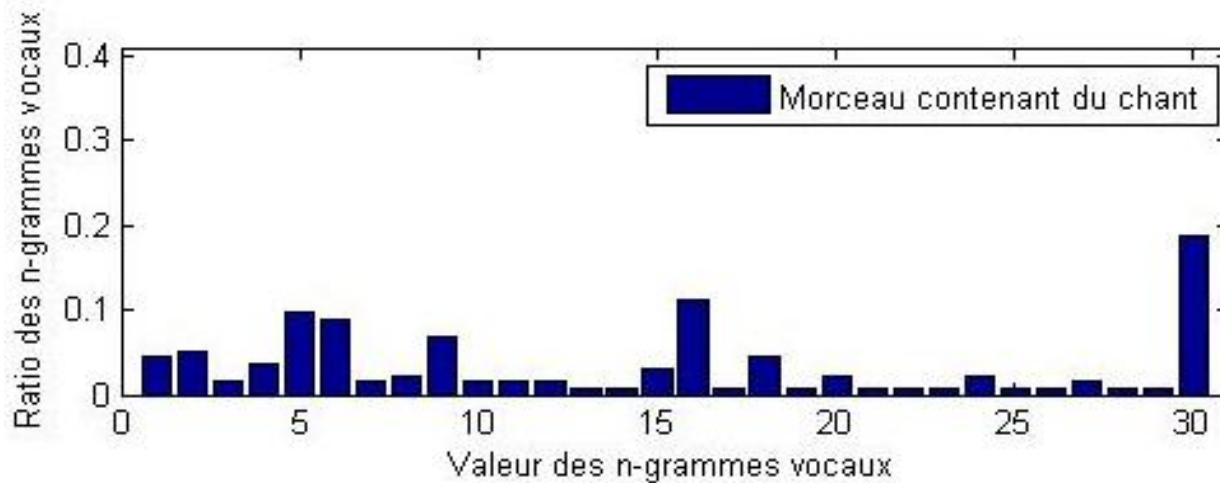
N-grammes



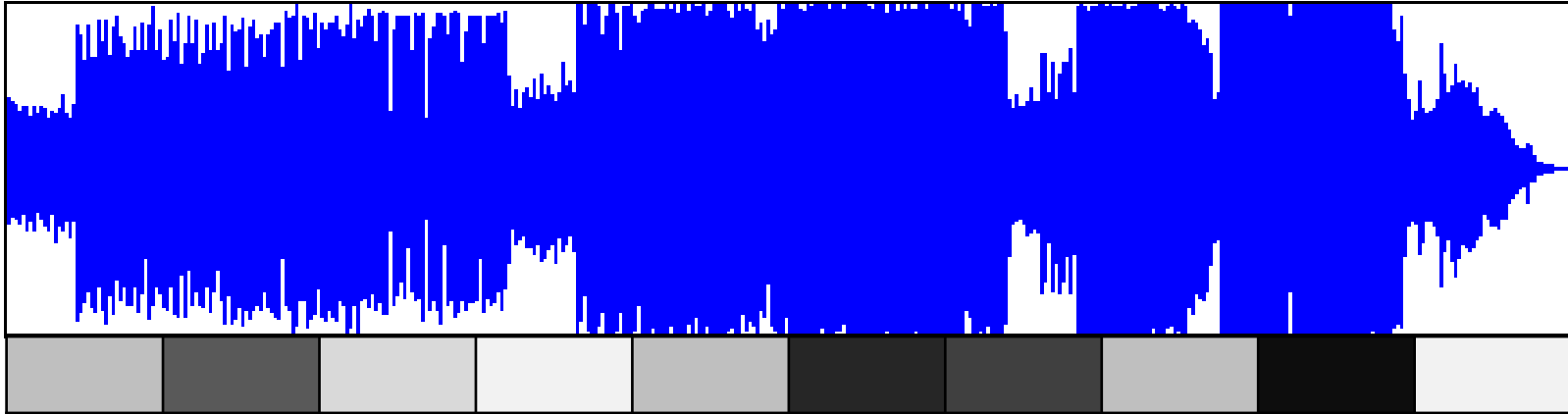
Instru.	Chant	Chant	Instru.	Instru.	Chant	Instru.	Instru.	Chant	Instru.
	2					1			1

2	1	1	...
---	---	---	-----

N-grammes

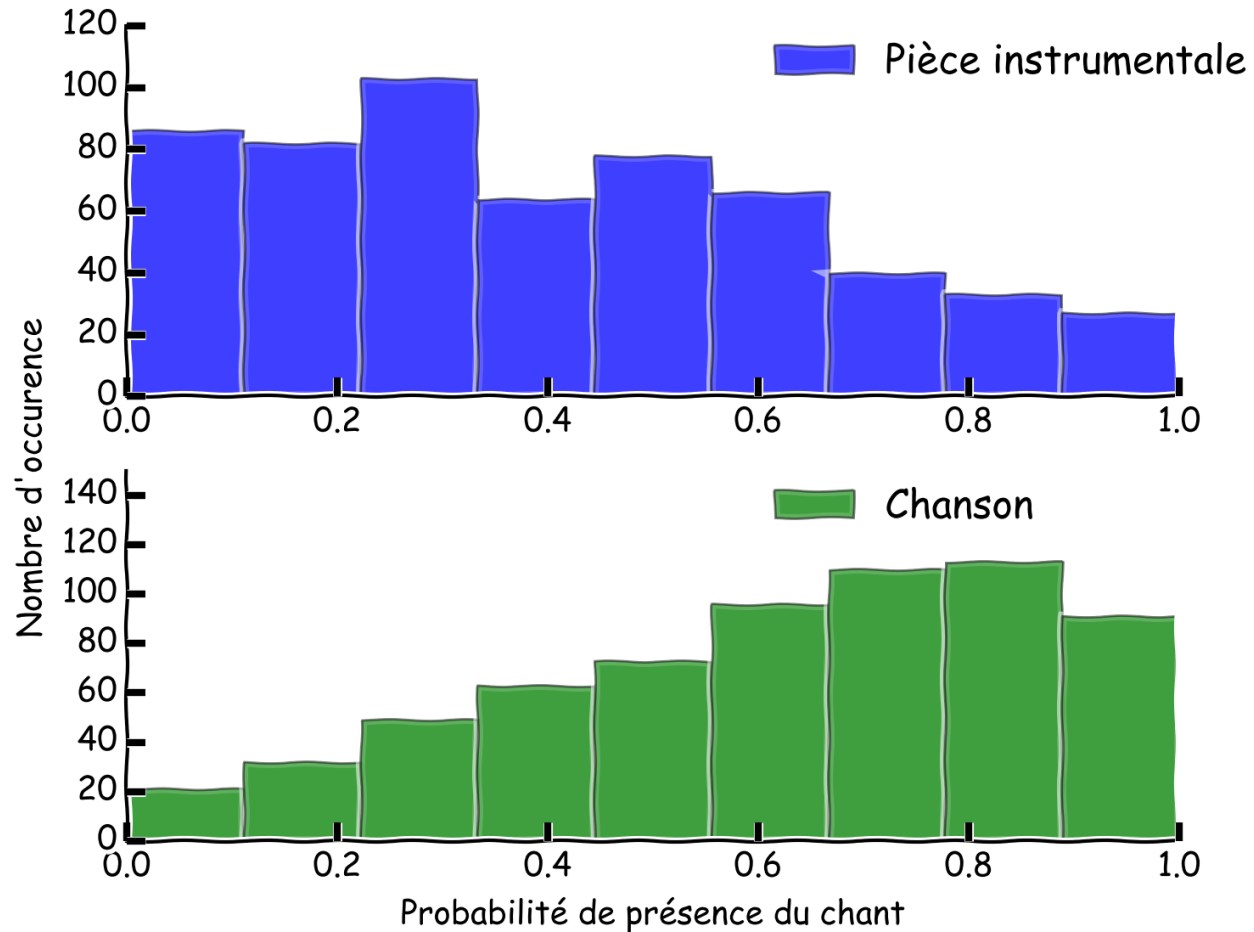


Histogramme



Estimation de la présence de la voix comprise entre [0; 1]

Histogramme



Problème 6

- ♪ Algorithme de Lehner en Matlab
- ♪ Conversion actuelle en python pour exécution sur base de données Simbals
- ♪ Pas d'autres algorithmes disponibles
 - ♪ Code non disponible
 - ♪ Code disponible mais pas adaptable sauf modification profonde de l'algorithme
 - ♪ Code disponible mais nécessité d'optimisation

Problème 7

- ♪ ISRC (International Standard Recording Code)
- ♪ Années = « Registration Year »

FR-26S-**20**-04007

Pays-Entreprise-**Année**-ID

<https://api.deezer.com/2.0/track/isrc:FR26S2004007>

Bruno Coulais – Caresse sur l'Océan (**30 janvier 2015**)

- ♪ Pays d'enregistrement != pays de l'artiste
- ♪ Mauvaises références
- ♪ ISWC (International Standard Work Code)
- ♪ Identifiants industriels

Conclusion et Perspectives

- ♪ Features qui impacte plus les résultats que le Machine Learning
- ♪ Analyse multirésolution
- ♪ Biais du genre musical
- ♪ Deep Learning
 - ♪ Nécessite beaucoup de données en entrée
- ♪ Lead/Background separation
- ♪ Définition du chant ? (Cf Talking box de Gilmour)

Apports

- ♪ Remise en question
 - ♪ de la pertinence du rapport train/test
 - ♪ de la capacité des algorithmes existants à traiter de gros volume de données musicale
 - ♪ au niveau du temps calcul
 - ♪ et des résultats obtenus
 - ♪ Optimisation F-Measure plutôt que précision (en fonction des objectifs)
- ♪ Proposition
 - ♪ Base de données sous forme d'identifiants
 - ♪ Base de vérité correspondante
- ♪ Formalisation
 - ♪ Nouvelle tâche *via* la création d'une playlist parfaite
 - ♪ Paradigme d'une tâche existante : tag chant/instru (jamais fait)
- ♪ Indications (et explications) de features et de classifieurs qui ne fonctionnent pas !

Merci

**Questions
&
Suggestions**