Une Méthode Globale pour la Reconnaissance de Partitions musicales

Florence ROSSANT

ISEP, Institut Supérieur d'Electronique de Paris

21, rue d'Assas, 75006 PARIS, FRANCE

Florence.Rossant@isep.fr

Résumé – Cet article présente une méthode de reconnaissance automatique de partitions musicales. L'objectif est de permettre l'intégration d'un maximum de connaissances contextuelles, afin de réaliser une interprétation globale de haut niveau de l'ensemble de la partition. La méthode exposée procède en deux temps : la première étape analyse individuellement les objets et fournit des hypothèses de reconnaissance à la seconde étape qui, intégrant les règles d'écriture musicale, aboutit à une décision globale cohérente.

Abstract *– This paper presents a system that can automatically recognize a scanned paper-based music score. The purpose of the exposed method is to integrate as much contextual knowledge as possible in order to realize a high level and global interpretation of the whole music score. It is based on two major stages : the first one makes the analysis of the isolated objects and outputs some recognition hypotheses about them. The second one takes the final consistent decision through high-level processing including music writing rules.*

1. Introduction

Les caractéristiques de l'écriture musicale rendent l'automatisation de la lecture difficile. La difficulté se situe dès l'étape d'extraction des primitives. En effet, l'écriture musicale est très dense et il est également très fréquent, que des objets normalement séparés se touchent ou soient au contraire sub-divisés [1]. Les travaux de recherche sont donc allés au delà de la reconnaissance individuelle des symboles, par exemple en effectuant une reconnaissance globale basée sur un modèle probabiliste [2], ou sur une grammaire intégrant les interactions entre symboles et gérant tout le processus de reconnaissance [3]. Notre démarche va également dans le sens d'une interprétation globale de la partition, mais nous proposons de procéder en deux temps : la première étape analyse individuellement les objets et fournit des hypothèses de reconnaissance à la seconde étape qui, intégrant les règles d'écriture musicale, aboutit à une décision globale cohérente. Notre méthode présente plusieurs avantages : elle ne nécessite pas de nombreuses données d'apprentissage; elle permet d'intégrer dans un même processus des interactions de bas niveau de type graphique, et des interactions de plus haut niveau de type syntaxique, celles-ci pouvant par ailleurs être plus ou moins locales.

Après avoir présenté l'ensemble du programme, nous détaillerons les deux étapes permettant la reconnaissance des symboles, dites respectivement d'analyse et de décision. Enfin, nous présenterons les résultats obtenus, pour conclure sur les améliorations en cours d'étude.

2. Présentation du système

En entrée du programme, nous avons l'image binaire (1 pour un pixel noir, -1 pour un pixel blanc) de la partition scannée à 300 dpi, ainsi que des informations globales, la clé, la métrique et la tonalité. Pour l'instant, le système ne traite que le cas monodique, et reconnaît les symboles nécessaires à la reconstitution de la mélodie : notes (hauteur et durée), altérations, silences, barres de mesure. Un ensemble de pré-traitements permet de détecter les portées, de corriger l'inclinaison de l'image, et d'effacer les lignes de portée. Les algorithmes que nous utilisons (non présentés dans cet article) sont très simples mais peuvent introduire des défauts de segmentation. Des méthodes plus robustes sont exposées dans [4]. L'étape d'analyse extrait ensuite les différents objets et, à partir de calculs de corrélation avec des modèles de référence, produit un ensemble d'hypothèses de reconnaissance. Enfin, l'étape de décision évalue mesure par mesure toutes les combinaisons possibles et retient la solution la plus probable satisfaisant aux règles musicales.

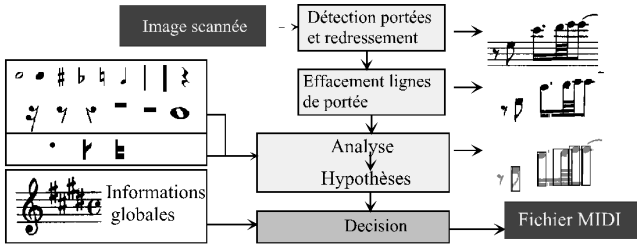

FIG 1 : Représentation de l'enchaînement des traitements

3. Analyse

Cette étape prend en entrée l'image *I* redressée, après effacement des lignes de portée. La technique utilisée est principalement basée sur le « pattern matching » .

3.1 Pattern matching

Définissons le score de corrélation entre la forme analysée à la position *(x,y)* dans l'image *I*, et les modèles de référence

M^k, de taille $d_x^k.d_Y^k$:

$$C^k(x,y) = \frac{1}{d_x^k.d_y^k} \sum_{(i,j)\in M^k} M^k(i,j) I(x+i, y+j) \qquad (1)$$

En cas de parfaite correspondance entre la forme testée et le modèle, le score atteint la valeur maximale de 1.0. Il décroît avec le nombre de pixels qui diffèrent. Ce calcul est effectué pour plusieurs positions *(x,y)*, afin de rechercher le score maximal et la position (x_k,y_k) correspondante.

D'autres techniques sont généralement préférées pour la reconnaissance des symboles musicaux, comme la morphologie mathématique [5] ou les réseaux neuronaux [6]. Mais le « pattern matching » est bien adapté à notre démarche, car le score de corrélation est relativement robuste aux défauts de segmentation et fournit une mesure de vraisemblance sur la nature de l'objet détecté.

Pour optimiser les calculs de corrélation, nous distinguons les symboles caractérisés par un segment vertical, plus long que 1.5 interligne (FIG 3), de tous les autres (FIG 5).

3.2 Objets caractérisés par un segment vertical

3.2.1 Localisation

Dans un premier temps, les segments verticaux plus longs que 1.5 interligne sont détectés dans chaque colonne de l'image. Une analyse des segments adjacents permet ensuite de ne retenir que le segment principal caractérisant une ligne verticale. On applique ensuite à ce dernier un algorithme de croissance de région, de sorte que les objets sont précisément localisés par une boîte englobante (FIG 2).

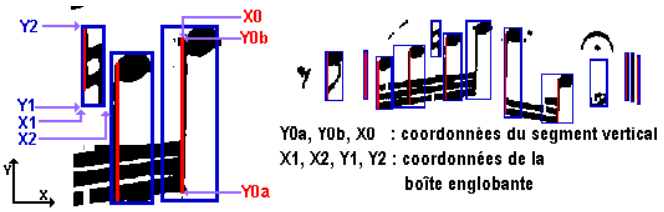

FIG 2 : Localisation des objets ayant un segment vertical

3.2.2 Caractéristiques géométriques et corrélations

Des caractéristiques géométriques intéressantes peuvent être déduites des résultats précédents. Par exemple, un objet entouré d'un cadre étroit situé entre la 1ère et la 5ème ligne de portée est sans doute une barre de mesure. Les classes sont regroupées en trois groupes principaux (FIG 3): les barres de mesure, les notes, les altérations. Cinq critères extraits des coordonnées des boîtes englobantes et des segments verticaux caractérisent chaque groupe.

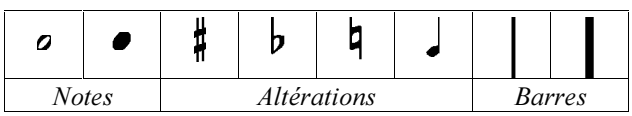

FIG 3 : Objets caractérisés par un segment vertical

Lorsqu'un objet satisfait à au moins 3 des 5 critères, le programme calcule sa corrélation avec tous les modèles du groupe, et mémorise chaque score $C^k(x_k,y_k)$ avec la position correspondante (x_k,y_k). Ainsi, le processus de localisation des symboles caractérisés par un segment vertical permet de réduire le coût de calcul et de fiabiliser les résultats, en éliminant les classes impossibles et en délimitant les zones de recherche.

3.2.3 Mémorisation d'hypothèses

Trois seuils ont été expérimentalement définis : un seuil de décision t_d, un seuil minimal t_m, un seuil d'ambiguïté t_a, correspondant à trois niveaux d'hypothèses, *L1*, *L2* et *L3*. Soit un objet *Oj* à identifier, et $C^{k1}(x_{k1},y_{k1})$ le score de corrélation le plus élevé obtenu pour cet objet. L'hypothèse « *Oj* est de classe M^{k1} » est stockée en *L1* si $C^{k1}(x_{k1},y_{k1})$ est supérieur à t_d ; si $C^{k1}(x_{k1},y_{k1})$ est inférieur à t_d, mais supérieur à t_m, cette hypothèse est stockée en *L2*, l'hypothèse *L1* étant alors « absence de symbole ». Enfin, si le deuxième plus haut score $C^{k2}(x_{k2},y_{k2})$ est supérieur à t_m, et est ambigu avec le premier (différence des scores inférieure à t_a), l'hypothèse « *Oj* est de classe M^{k2} » correspondante est stockée en *L3*. Ainsi, différentes hypothèses de reconnaissance sont émises pour chaque objet, mais aucune décision n'est prise.

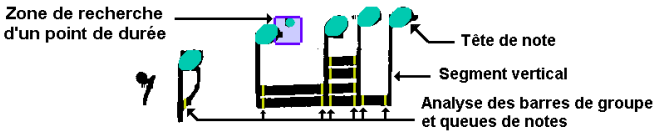

FIG 4 : Analyse de la durée des notes

La hauteur d'une note ou d'une altération est déduite de l'ordonnée y_k. La durée des notes est déduite du nombre de queues de note comptées à l'extrémité opposée du segment vertical, et de la recherche d'un point éventuel (FIG 4).

3.3 Analyse des autres objets

Cette étape concerne les silences et les rondes:

FIG 5 : Objets sans segment vertical

Ces symboles sont recherchés par calcul de corrélation entre les boîtes englobantes, uniquement le long de la 3ème ligne de portée pour les silences. Lorsqu'il y a correspondance entre un modèle et un objet, la fonction de corrélation présente un pic et le résultat est mis en mémoire. Comme précédemment, il peut y avoir ambiguïté lorsque plusieurs modèles présentent des pics à des positions voisines. C'est pourquoi un algorithme itératif examine par ordre décroissant les scores de corrélation obtenus par les différents modèles, les compare aux seuils t_m, t_d, et t_a tout en vérifiant leurs positions relatives, afin de les ranger en hypothèse de niveau *L1, L2* ou *L3*.

3.4 Analyse des groupements de notes

Il est d'usage dans la notation musicale de grouper les noires en fractions de temps, rendant la structure rythmique plus lisible (FIG 6). Ces conventions sont utilisées par notre

programme pour fiabiliser l'analyse de la durée des notes. Un algorithme basé sur une croissance de région extrait les notes groupées et compare leur structure rythmique aux organisations usuelles. S'il n'y a aucune correspondance, l'algorithme émet au maximum deux nouvelles hypothèses permettant d'atteindre la durée usuelle de groupe immédiatement supérieure et/ou immédiatement inférieure, tout en changeant un minimum de durées notes.

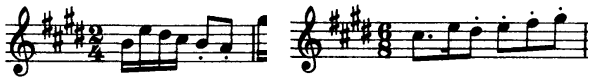

FIG 6 : Exemples de groupements usuels de notes

Les notes restantes sont maintenant supposées être non groupées, et le nombre de queues de note est plus précisément recompté sous cette nouvelle hypothèse.

4. Décision globale

4.1 Regroupement d'hypothèses

Les objets précédemment détectés et analysés sont réordonnés suivant l'axe horizontal. Afin d'introduire des informations globales telles que la métrique, la décision est prise mesure par mesure. Un maximum de 5 hypothèses peut être émis pour chaque objet:

TAB 1 : Définition des 5 niveaux d'hypothèses

L1	M^{k1} de score maximal Ou pas de symbole	: $C^{k1}(x_{k1},y_{k1}) \geq t_d$: $t_m < C^{k1}(x_{k1},y_{k1}) < t_d$
L2	M^{k1} de score maximal	: $t_m < C^{k1}(x_{k1},y_{k1}) < t_d$
L3	M^{k2} , 2ème score	: $t_m < C^{k2}(x_{k2},y_{k2})$ et $[C^{k1}(x_{k1},y_{k1}) - C^{k2}(x_{k2},y_{k2})] < t_a$
L4	1ère hypothèse de changement de durée	
L5	2ème hypothèse de changement de durée	

On construit alors pour chaque mesure un tableau à deux dimensions : horizontalement, les objets présents dans la mesure, verticalement, les hypothèses faites sur ces objets. Les points allongeant la durée des notes sont traités comme des objets à part entière et peuvent apparaître en *L1* ou *L2*.

4.2 Algorithme de décision

L'algorithme de décision calcule le score de corrélation moyen et la durée totale de toutes les combinaisons cohérentes, c'est-à-dire satisfaisant aux conditions suivantes:
- Un point doit être dans la zone de recherche d'une note qui le précède immédiatement.
- Une altération doit être suivie d'une note de même hauteur.
- Un changement de durée de note (choix du niveau *L4* ou *L5*) est possible si la combinaison respecte l'hypothèse sous-jacente de groupement de notes.

La combinaison retenue est celle qui satisfait au mieux aux critères de décision qui sont, par ordre de priorité :
- Le nombre total de temps dans la mesure est correct.
- Le score de corrélation moyen est maximal.
- Le nombre de corrections est minimal.

4.3 Exemple

Illustrons sur une mesure (FIG 7, TAB 2.) comment des ambiguïtés ont pu être levées et la solution correcte trouvée.

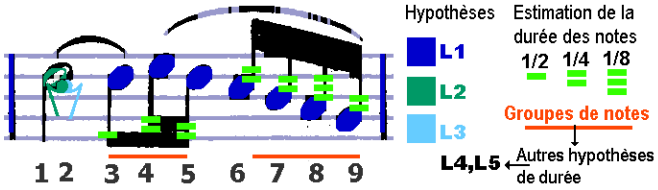

FIG 7 : Résultats de l'analyse

Pour l'objet 1, l'hypothèse *« blanche »* est de niveau *L2* car le score de corrélation entre l'objet et le modèle *« blanche »* est $t_m < 0.54 < t_d$; l'hypothèse L1 est donc *« pas de symbole »* (NS). Il en est de même pour l'objet 2, mais cette fois, le niveau *L3* est utilisé, l'hypothèse *« demi-soupir »* étant ambiguë avec l'hypothèse *« soupir »,* puisque $0.51 > t_m$ et $(0.55-0.51) > t_a$. Il n'y a que l'hypothèse « noire » pour les symboles 3 à 9, avec des scores supérieurs à t_d, donc de niveau *L1*. Pour que le groupe de notes 6,7,8,9 atteigne la durée usuelle de 1 temps, il faut changer la durée de la noire n°8 de 1/8 vers 1/4, d'où l'hypothèse *L4*. Finalement, il y a 12 combinaisons possibles, toutes cohérentes. L'algorithme choisit la solution représentée en gris clair (TAB 2), validant l'objet 1 en *« blanche »,* éliminant l'objet 2, et corrigeant la durée de la note 8. Le nombre de temps de la mesure est correct (4 temps), avec un score moyen de 86.75%.

TAB 2 : Tableau des hypothèses (classe, durée, score)

	1	2	3	4	5	6	7	8	9
L1	NS 0/1 -	**NS 0/1** -	**1/2** **0.94**	**1/4** **0.90**	**1/4** **0.94**	**1/4** **0.87**	**1/4** **0.92**	1/8 0.88	1/4 0.95
L2	2/1 0.54	1/1 0.55							
L3		1/2 0.51							
L4								**1/4** **0.88**	

5. Résultats et améliorations

5.1 Taux de reconnaissance

Les résultats suivants ont été obtenus sur une base de 50 pages de musique, provenant d'éditions variées, et sans aucun ajustement de modèles ou de paramètres.

TAB 3 : Taux de reconnaissance

Barres de mesure: **r=99.5%**					Notes: r=**97.2**				
R1	R2	R3:	R4	R5	R1	R2	R3:	R4	R5
99.5	0.0	0.0	0.0	0.5	94.1	3.1	1.6	1.1	0.1
Altérations: r=**92.7**					Silences: r=**84.7**				
R1	R2	R3:	R4	R5	R1	R2	R3:	R4	R5
90.7	2.0	4.2	0.9	2.2	75.2	9.5	1.7	2.1	11.5

R1 : taux de symboles initialement corrects (choix L1 juste)
R2 : taux de symboles bien corrigés (choix L2 à L5 juste).
R3 : taux de décisions initiales (L1) fausses non corrigées.

R4 : taux de corrections fausses.
R5 : taux de symboles manquants ou ajoutés.

5.2 Discussion

Toutes classes confondues, le taux moyen de reconnaissance est au dessus de 97%, Il est important de noter que le nombre de corrections fausses (R4) est bien plus faible que le nombre de corrections justes (R2). Bien que la solution correcte soit rarement absente des hypothèses, l'algorithme de décision ne permet pas encore de résoudre toutes les ambiguïtés, car les contraintes exprimées sont insuffisantes. La figure 8 montre 3 cas typiques d'erreurs:

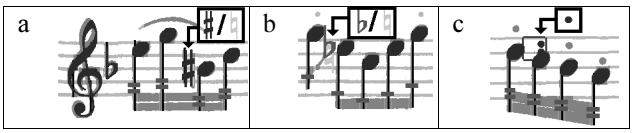

FIG 8 : Cas typiques de confusions

La contrainte essentielle étant la durée totale de la mesure, l'algorithme de décision ne peut résoudre les ambiguïtés entre altérations (a), ou la fausse détection d'altérations (b). De plus, cette contrainte peut actuellement être elle-même génératrice d'erreurs, par exemple, par le choix d'un silence plus court pour compenser la fausse détection d'un point (c). Enfin, les erreurs sur les barres de mesure rendent au mieux l'algorithme de décision inopérant, introduisent au pire de fausses corrections. Les résultats insuffisants sur les altérations et les silences s'expliquent ainsi.

5.3 Améliorations en cours d'étude

Ces premiers résultats montrent bien sûr qu'il faudrait fiabiliser la détection des barres de mesure, en utilisant un détecteur de segment plus robuste [7]. Mais surtout, il est nécessaire d'introduire dans l'algorithme de décision davantage de connaissances contextuelles, notamment au niveau des positions graphiques relatives entre les objets, et des règles d'utilisation des altérations. La modélisation floue semble être une approche intéressante, car elle permet de modéliser et de fusionner des informations aussi bien numériques que syntaxiques. De premiers essais ont déjà été réalisés pour améliorer la reconnaissance des altérations [8], en exprimant un degré de compatibilité graphique entre une altération et la note suivante, et un degré de compatibilité syntaxique entre la tonalité et les altérations dans ou hors de la mesure. Ainsi, le taux de reconnaissance moyen a été amélioré sur ces symboles de 4.9%. La méthode est actuellement étendue aux groupements de notes.

5.4 Comparaison

Bien que les logiciels du commerce proclament de très bons taux de reconnaissance, il reste en pratique beaucoup trop d'erreurs. Nous avons comparé (FIG 9) les résultats produits par notre programme (à gauche) avec ceux de SmartScore [9] pour Windows (à droite). Les erreurs sont indiquées par une flèche pleine. Il est évident que SmartScore ne satisfait pas à la règle élémentaire de la métrique alors que notre programme utilise cette règle pour réintroduire des silences et corriger des erreurs de durée note (corrections en rouge à gauche, flèche simple ↓).

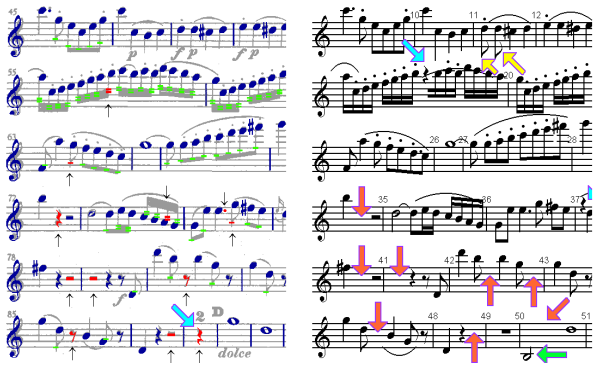

FIG 9 : Comparaison de résultats

6. Conclusion

Nous avons présenté un système de reconnaissance de symboles musicaux procédant en 2 étapes. La première extrait les symboles et produit un ensemble d'hypothèses de reconnaissance. La seconde réalise sur chaque mesure une décision globale devant satisfaire à des contraintes d'écriture musicale. Les résultats présentés montrent l'efficacité de cette méthodologie. Cependant, il est nécessaire d'introduire d'avantage de connaissances contextuelles dans l'algorithme de décision. Ceci est actuellement en cours d'étude.

Références

[1] D. Blostein, H. Baird. *A critical survey of music image analysis.* Structured Document Image Analysis, 405-434, éd. H.S.Baird at al., Springer Verlag, 1992.

[2] M.V. Stückelberg, D. Doermann. *On musical score recognition using probabilistic reasoning*. Proc. of ICDAR, 115-118, Bangalore, India, 1999.

[3] B. Couasnon, B. Rétif. *Using a grammar for a reliable full score recognition system*. Proc. of ICMC, 187-194, Banff, Canada, 1995.

[4] N. Carter, R. Bacon. *Automatic recognition of printed music*. Structured Document Image Analysis, 456-465, éd. H.S. Baird at al., Springer-Verlag, 1992.

[5] B. Modayur. *Music score recognition - A selective attention approach using mathematical morphology.* Tech. report, University of Washington, Seattle, 1996.

[6] H. Miyao, Y. Nakano. *Head and stem extraction from printed music scores using a neural network approach.* Proc. of ICDAR, 1074-1078, Montreal, Canada, 1995.

[7] V. Poulain d'Andecy, J. Camillerapp, I. Leplumey. *Kalman filtering for segment detection: application to music scores analysis,* Proc. of ICPR, 301-305, Jerusalem, Israel, 1994.

[8] F. Rossant, I. Bloch. *Reconnaissance de Partitions Musicales par Modélisation Floue et Intégration de Règles Musicales*, GRETSI, Toulouse, France, 2001.

[9] SmartScore Demo for Win95/98/NT Version 1.3: http://www.musitek.com/demopage.html