Amélioration de la reconnaissance de partitions musicales par modélisation floue et indication des erreurs possibles

F. ROSSANT[1], I. BLOCH[2]

[1]ISEP, 28 rue Notre-Dame des Champs, 75006 Paris, [2]ENST CNRS UMR 5141, 46 rue Barrault, 75634 PARIS Cedex 13

florence.rossant@isep.fr , isabelle.bloch@enst.fr

Résumé – Nous proposons une méthode de reconnaissance de partitions musicales fondée sur la modélisation floue des informations extraites de la partition scannée et des règles de musique. L'objectif est de prendre en compte les ambiguïtés qui subsistent à l'issue de l'étape d'analyse individuelle des symboles, et d'obtenir une interprétation globale cohérente. Le formalisme proposé permet de modéliser l'imprécision sur la détection des symboles, la variabilité des polices, la souplesse des règles musicales, et ainsi de fiabiliser les résultats de reconnaissance. Nous montrerons qu'il permet également d'indiquer les erreurs potentielles, de manière à faciliter la correction manuelle.

Abstract *– We propose an OMR method based on fuzzy modeling of the information extracted from the scanned music score and of musical rules. The aim is to disambiguate the recognition hypotheses output by the individual symbol analysis process. Fuzzy modeling allows to account for imprecision in symbol detection, for typewriting variations, and for flexibility of rules. The reliability of the recognition is increased, and the results are also used in order to indicate the possible errors, and thus to facilitate the manual correction.*

1. Introduction

La recherche dans le domaine de la reconnaissance optique de la musique (OMR) a commencé dès les années 1970. Cependant, la difficulté d'extraction des symboles musicaux, due au haut degré de connectivité entre primitives aussi bien qu'aux défauts d'impression, ainsi que la complexité et la souplesse des règles d'écriture musicales [1], ne permettent pas encore d'obtenir des résultats de reconnaissance suffisamment fiables. Les efforts se sont concentrés depuis les années 1990 sur l'introduction d'informations structurelles. Cependant, les méthodes présentées se limitent souvent à la description des symboles [2], à la formalisation de règles graphiques strictes et locales [3,4], ou bien sont fondées sur des modèles probabilistes [5], nécessitant de nombreuses données d'apprentissage.

L'objet de notre recherche est de prendre en compte les sources d'incertitude et d'imprécision (variabilité des polices, défauts de segmentation), de modéliser et d'intégrer l'ensemble des règles d'interprétation de haut niveau, strictes ou souples, dans un processus global de décision, afin d'aboutir à une reconnaissance totalement cohérente de la partition. La théorie des ensembles flous et des possibilités offre un formalisme bien adapté à la modélisation et l'intégration de contraintes souples [7], ainsi qu'à la prise en compte de l'imprécision sur la forme et la localisation [8].

Une autre contribution de cet article est l'aide à la correction. Certains auteurs tentent de repérer des incohérences afin de tenter de corriger certains types d'erreurs (sur les durées des noires le plus souvent) par des post-traitements spécifiques [3,9]. Ici, nous utilisons directement certains résultats de la modélisation floue, exprimés sous forme de degrés de possibilité, et ajoutons d'autres critères portant sur la décomposition rythmique des mesures, afin d'indiquer les symboles ou groupes de symboles qui pourraient être erronés, sans restriction sur la nature de l'erreur, de manière à faciliter la vérification et la correction manuelle.

Nous débuterons par une brève présentation du système global [10]. Puis nous présenterons les éléments de la modélisation floue qui servent d'indicateurs pour la correction. Nous indiquerons ensuite la règle de décision globale, la méthode d'indication des erreurs, avant de conclure sur les performances obtenues en reconnaissance et en détection des erreurs.

2. Système global

En entrée du programme, nous avons l'image binaire (1 pour un pixel noir, 0 pour un pixel blanc) de la partition scannée à 300 dpi, ainsi que des informations globales, la clef, 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.

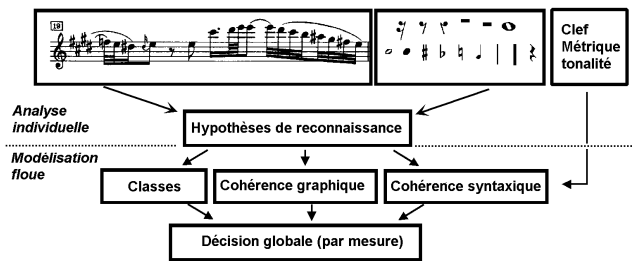

FIG. 1 : Système global

L'algorithme procède en deux étapes. Une première étape d'analyse [6], réalise la détection des objets, et, par corrélation avec des modèles de référence, propose pour chacun plusieurs hypothèses de reconnaissance. La théorie des ensembles flous et des possibilités permet dans une

seconde étape de combiner les informations de position et de corrélation fournies par la première, de modéliser les règles graphiques et syntaxiques de la musique, afin d'aboutir à une décision finale par optimisation de tous les critères.

3. Modélisation floue

3.1 Résultats initiaux exploités

L'étape d'analyse [6] fournit les positions (x_k,y_k) et les scores de corrélation $C^k(s)$ des objets s avec les modèles de référence M^k. Trois hypothèses de reconnaissance sont retenues, au maximum, si les scores de corrélation correspondants sont supérieurs à un seuil t_m (fixé à 0.3 dans nos expérimentations). On ajoute également la possibilité qu'il n'y ait pas de symbole ('-' dans les tableaux), lorsque le plus haut score de corrélation est inférieur au seuil de décision $t_d(k)$. Ces seuils ont été définis expérimentalement. Le seuil $t_d(k)$ est élevé lorsqu'on observe une grande probabilité de fausse détection pour la classe k, et plus faible lorsque le score de corrélation est très sensible aux variations de police de symboles. La figure 2 montre deux barres de mesure (a) et (b) extraites d'une même partition, et, en superposition, les hypothèses de reconnaissance. Le tableau 1 indique certains scores de corrélation obtenus.

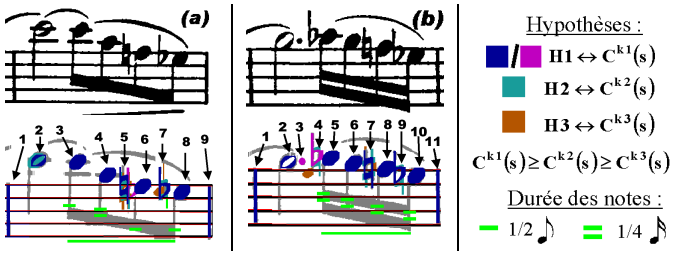

FIG. 2 : Image originale et hypothèses de reconnaissance

TAB. 1 : Scores de corrélation

(a)	2	5	7	(b)	3	4	7
H1	0.59	ou 0.56	0.72	H1	ou 0.74	ou 0.67	0.71
H2	0.38	0.49	0.54	H2		0.52	0.56
H3		0.42	0.43	H3		0.40	0.47

On voit sur cet exemple les limites de l'analyse individuelle des symboles : les scores de corrélation peuvent être très ambigus, et le plus haut ne correspond pas toujours à l'hypothèse correcte. Ces ambiguïtés sont dues aux défauts de segmentation et aux variations dans les polices de symboles inter et intra partition (voir par exemple les bécarres 5(a) et 7(b) ou les bémols 7(a) et 4(b)). Pour les résoudre, nous modélisons les classes de symboles par des distributions de possibilité apprises à partir des scores de corrélation, et nous modélisons les règles graphiques et syntaxiques de l'écriture musicale par des coefficients de compatibilité entre symboles.

3.2 Modélisation des classes de symboles

Les scores de corrélation obtenus indiquent la similarité entre les symboles de la partition et les modèles de classe. Nous définissons donc le degré de possibilité $\pi_k(s)$ d'appartenance de l'objet s à la classe k comme une fonction croissante du score de corrélation $C^k(s)$ (Fig. 3) :

$$\pi_k(s) = f_k\left(C^k(s)\right) \qquad (1)$$

Le paramètre D de la fonction f_k représente la zone d'incertitude et il est constant (toujours 0.3). En revanche, S_k est fonction de la ressemblance entre le modèle de référence M^k et les symboles de la partition, et il est déduit des scores de corrélation résultant de l'analyse individuelle. Soient n_k le nombre d'objets obtenant le plus haut score de corrélation pour la classe k, supérieur à $t_d(k)$, et $m(k)$ la moyenne de ces scores. On définit S_k par :

$$S_k = \frac{t_d(k) + n(k)m(k)}{n(k)+1} \qquad (2)$$

Lorsque le modèle de référence M^k ne correspond pas bien aux symboles de cette classe dans la partition, à cause des variations de police, alors $m(k)$ est proche de $t_d(k)$. Dans le cas contraire, il prend des valeurs élevées, translatant la distribution de possibilité vers la droite.

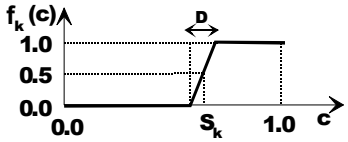

FIG. 3 : Distribution de possibilité de la classe k

Notons que la forme des distributions de possibilité n'a pas besoin d'être estimée de manière précise, et on constate expérimentalement une bonne robustesse par rapport à cette forme. L'essentiel est que ce n'est pas une fonction binaire, et qu'elle est croissante, c'est-à-dire que le degré de possibilité d'appartenance à la classe k est d'autant plus élevé que le score de corrélation l'est.

Le tableau 2 montre les résultats obtenus pour les deux barres de mesures de la figure 2. Les degrés de possibilité présentent généralement moins d'ambiguïté que les scores de corrélation du tableau 1. Notons par ailleurs que l'ordre peut changer, par exemple, l'objet 5 mesure (a) obtient un degré de possibilité maximal pour la classe bécarre, alors que le degré de possibilité pour la classe bémol est maintenant nul.

TAB. 2 : Degrés de possibilité d'appartenance aux classes

(a)	2	5	7	(b)	3	4	7
H1	0.35	ou 0.00	0.40	H1	ou 0.43	ou 0.28	0.68
H2	0.00	0.13	0.25	H2		0.20	0.35
H3		0.00	0.00	H3		0.00	0.00

3.3 Modélisation des règles graphiques

Elles permettent d'exprimer la cohérence sur les positions relatives des symboles. Aux règles déjà définies [10], concernant les relations entre les altérations et les notes, les points et les notes, a été ajoutée la définition d'un coefficient de compatibilité entre deux symboles voisins quelconques. Cette règle exprime que deux symboles ne devraient pas se superposer. On estime les positions des boîtes englobantes de deux symboles successifs s_n et s_m $(m>n)$, à partir des dimensions typiques des deux classes k_n et k_m considérées (Fig. 4a). Soient Δl et Δh les décalages entre les bords horizontaux et verticaux. Les valeurs admissibles dans les deux directions sont définies par la fonction f (Fig. 4b), et le coefficient de compatibilité entre s_n et s_m par :

$$C_p\left(s_n^{k_n}, s_m^{k_m}\right) = Max\left[f(\Delta l), f(\Delta h)\right] \qquad (3)$$

La fonction *f* autorise des décalages négatifs, afin de prendre en compte l'imprécision de l'estimation, et afin de traiter les partitions de forte densité.

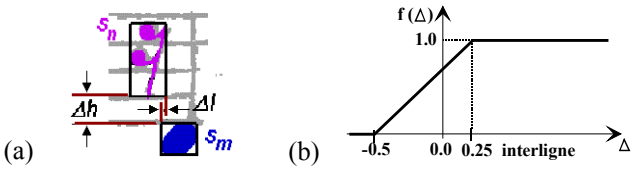

FIG. 4 : Compatibilité graphique entre deux symboles

Le coefficient de compatibilité étant calculé pour toutes les paires de symboles voisins, on peut maintenant en déduire pour chaque symbole son coefficient de compatibilité avec tous ses voisins, antérieurs et postérieurs :

$$C_p\left(s_n^{k_n}\right)=\left[\min_{j<n} C_p\left(s_n^{k_n}, s_j^{k_j}\right)\right]\cdot\left[\min_{l>n} C_p\left(s_n^{k_n}, s_l^{k_l}\right)\right] \quad (4)$$

Cette règle s'est révélée très efficace en cas de partitions de forte densité, car elle permet de ne pas éliminer des configurations qui théoriquement seraient inacceptables. Le bémol 9 de la mesure (b) par exemple, qui se superpose légèrement à la note précédente, obtient un coefficient de compatibilité graphique de 0.5.

3.4 Modélisation des règles syntaxiques

Nous appelons règles syntaxiques les règles musicales qui concernent les altérations et la tonalité [10], ou la métrique. Concernant la métrique, une règle stricte est celle du nombre de temps par mesure. Les méthodes de groupement des notes en temps, multiple ou fraction de temps, permettant de faciliter la lecture rythmique, sont au contraire des règles d'usage souples. En effet, il n'y a pas une décomposition unique pour une mesure donnée, et ces règles peuvent être relâchées, pour indiquer un phrasé par exemple.

L'étape première d'analyse individuelle des symboles a fourni pour chaque note noire sa durée [6], qui va maintenant être confirmée ou révisée en considérant les groupes de notes. Ceux-ci sont extraits par un algorithme de croissance de région. Lorsque la durée du groupe n'est pas usuelle, deux nouvelles hypothèses au maximum sont générées, augmentant et/ou diminuant la durée du groupe *g*, en changeant un minimum de durées dans le groupe. Le degré de possibilité affecté à chaque hypothèse H_l est fonction du nombre de corrections de durée de note *l(g)* et du nombre total de notes *L(g)* dans le groupe *g :*

$$C_l^{H_l}(g)=1.0-l(g)/L(g) \quad (5)$$

Dans la mesure (a) de la figure 2, le groupe de 4 notes a une durée inhabituelle de 1,75 temps (1/2, 1/4, 1/2, 1/2). Les deux hypothèses générées sont donc (1/2, 1/2, 1/2, 1/2), ou (1/2, 1/4, 1/8, 1/8), qui mènent le groupe respectivement à 2 temps ou 1 temps. Les degrés de possibilité sont 1.0 pour la configuration initiale, 0.75 et 0.5 pour les propositions de correction. Ces degrés n'évaluent pas les hypothèses par rapport aux règles d'usage de groupement, mais par rapport à l'interprétation initiale des durées, qui est supposée fiable.

4. Fusion et décision

La décision globale est réalisée mesure par mesure, en évaluant toutes les combinaisons possibles d'hypothèses. Pour chaque combinaison de symboles, indicée par *j*, il peut y avoir plusieurs combinaisons de durée H_l. Toutes les combinaisons de symboles qui contiennent au moins un coefficient de compatibilité graphique nul sont éliminées. Pour les autres, on moyenne les degrés de possibilité d'appartenance aux classes (1), les coefficients de compatibilité graphique (4) et les coefficients de compatibilité syntaxiques portant sur des altérations [10]. Le coefficient résultant *Conf$_s$(j)* indique la cohérence mutuelle des symboles. Les différentes combinaisons de durées de notes sont ensuite évaluées en fusionnant les coefficients (5) en un coefficient *Conf$_l$(j,H$_l$)* [10], exprimant le degré de possibilité des durées des notes pour la configuration H_l. On obtient le degré de possibilité final par multiplication de *Conf$_s$(j)* et *Conf$_l$(j,H$_l$)*, de sorte que les critères sur les symboles et sur les durées de notes doivent être simultanément respectés. L'algorithme retient la configuration maximisant ce degré de possibilité, avec priorité aux configurations qui satisfont à la contrainte stricte du nombre de temps par mesure.

5. Indication des erreurs

Les erreurs sont de quatre sortes : symbole ajouté, confusion, symbole manquant, erreur de durée de note. Nous proposons d'analyser la solution retenue par l'algorithme de décision, afin d'indiquer à l'utilisateur les symboles potentiellement erronés. Les critères utilisés sont le degré de possibilité d'appartenance aux classes, la compatibilité graphique, la décomposition rythmique de la mesure.

5.1 Symboles et relations graphiques

Considérons de nouveau chaque symbole s_n^k, classé en classe *k* par l'algorithme de décision, avec un degré de possibilité $\pi_k(s_n^k)$ d'appartenance à la classe *k* et une compatibilité de position $C_p\left(s_n^k\right)$ avec les autres symboles de la mesure. Des valeurs faibles peuvent être révélatrices d'une erreur de classification du symbole s_n. La règle suivante est donc appliquée : si $\pi_k(s_n^k)<t_s^k$ et $C_p\left(s_n^k\right)<t_g^k$, alors le symbole s_n est indiqué comme potentiellement faux. Les seuils t_s^k et t_g^k ont été déterminés par apprentissage, de façon à minimiser sur chaque classe *k* la somme des occurrences « non détection » et « fausses alarmes ». La base d'apprentissage comprend la moitié des images de notre base de données et est représentative des différents types d'édition.

Il n'y a pas de règle permettant, sur la base des hypothèses non retenues par l'algorithme de décision, d'indiquer les symboles manquants. En effet une règle du type $\pi_k(s_n)>t'^k_s$ ou $C_p\left(s_n^k\right)>t'^k_g$, conduit essentiellement à des fausses alarmes, les symboles manquants étant en fait généralement absents de l'ensemble des hypothèses. Cependant, les autres indicateurs permettent également de localiser grossièrement des symboles manquants, car les erreurs dans une mesure sont souvent corrélées.

5.2 Métrique et décomposition rythmique

Nous indiquons tout d'abord toutes les mesures qui ne satisfont pas à la contrainte stricte de métrique. Nous analysons ensuite la décomposition rythmique de la mesure. La durée minimale des groupes de notes dans la mesure nous permet de fixer un pas de découpage, égal à 1.0 ou 0.5 dans

une métrique binaire, 1.5 ou 0.5 dans une métrique ternaire. Les groupes de notes sont ensuite associés aux silences voisins, de manière à ce que la durée totale de chacune des associations soit égale à un multiple du pas. Celles qui ne satisfont pas à ce découpage idéal sont affectées d'un degré de possibilité nul. Pour toutes les autres, nous affectons un degré de possibilité égal à 0.5 pour les groupes peu courants mais possibles, et 1.0 pour les groupes tout à fait possibles. Les critères portent sur la répartition des durées et sur le type de métrique. Par exemple le degré de possibilité d'un groupe croche pointée/double/croche est de 1.0 dans une métrique ternaire, 0.5 dans une métrique binaire avec un pas de 0.5.

6. Résultats et conclusion

Le taux de reconnaissance des symboles, évalué sur une centaine de partitions monodiques (environ 42500 symboles) est maintenant de 98.6%. Ce résultat a été amélioré par rapport à [10] (+0.2%), malgré l'introduction dans la base de partitions plus difficiles, et cela en particulier grâce au compléments sur les règles graphiques. 98.4% des notes (noires, croches, etc.) ont une durée correcte. De plus, notre méthode permet de résoudre des cas pour lesquels SmartScore [11] échoue (Fig. 5). Les ambiguïtés sur la classe des symboles, les symboles connectés, et les durées des croches sont bien résolues. Les taux de reconnaissance sur les deux pages complètes sont de 98.7% pour notre programme, 92.0% pour SmartScore.

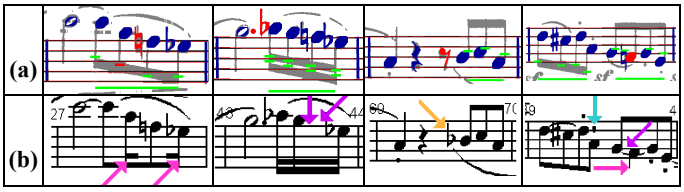

FIG. 5 : Résultats de reconnaissance
(a) notre méthode : 0 erreur, (b) Smartscore : 8 erreurs

Sur la base de test, 76% des confusions et ajouts sont bien indiqués, avec 30% des indications qui sont de fausses alarmes (0.4% de l'ensemble des symboles). 67% des erreurs de durée de note sont bien localisées : 40% directement par l'indication sur le groupe de notes, 27% par leur voisinage (mesure indiquée fausse ou symbole erroné dans la même mesure bien indiqué). Enfin, 64% des symboles manquants sont approximativement localisés par leur voisinage.

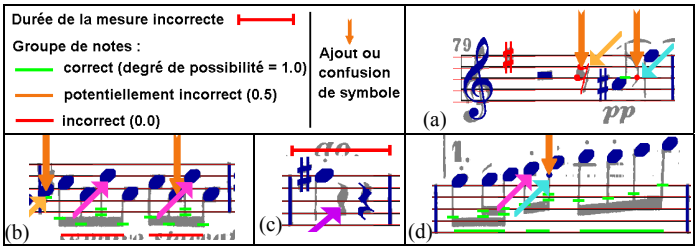

FIG. 6 : Résultats en détection

Sur la figure 6, on peut constater que les confusions et les ajouts de symboles sont bien indiqués (a)(b)(d) à l'exception d'une fausse alarme (croche en (b)); en (b), les deux erreurs de durée sont bien indiquées par l'indication de groupe erroné, et en (c), le soupir manquant est localisé par l'indication de mesure fausse; enfin, en (d), l'erreur sur la durée de la croche ne peut être détectée via le groupe car celui-ci est parfaitement cohérent (croche pointée + double croche) ; mais la détection du point ajouté permet de repérer rapidement cette erreur.

Ces résultats montrent l'intérêt de la méthode proposée, qui permet d'aboutir à une bonne fiabilité tant au niveau de la reconnaissance que de l'indication des erreurs possibles. On peut penser que cette méthodologie permet également d'adapter l'algorithme de reconnaissance à la partition traitée. En effet, l'analyse des degrés de possibilité d'appartenance aux classes et des coefficients de compatibilité graphique obtenus sur les premières pages permet certainement d'affiner la règle de décision, et ainsi d'obtenir une reconnaissance plus fiable sur le reste de la partition. D'autre part, les critères utilisés pour la détection des erreurs de durée pourraient être également intégrés dans l'algorithme de décision, de manière à obtenir un degré de possibilité exprimant simultanément le degré de possibilité des groupes de notes par rapport à l'interprétation initiale et la cohérence de la répartition des durées dans la mesure. Des cas d'erreurs qui se compensent (figure 6(b) par exemple) pourraient ainsi être éliminés au profit de corrections de durées conduisant à une décomposition rythmique correcte et à la bonne solution.

Références

[1] D. Blostein, H. Baird, *A Critical Survey of Music Image Analysis*. Structured Document Image Analysis, Eds H.Baird et al., Springer, Verlag, pp 405-434, 1992.

[2] G. Watkins, *The Use of Fuzzy Graph Grammar for Recognising Noisy Two-Dimensional Images*. NAFIPS Conference , pp 415-419, 1996.

[3] B. Coüasnon, B. Rétif, *Using a Grammar for a Reliable Full Score Recognition System*. International Computer Music Conference, Banff, Canada, pp 187-194, 1995.

[4] H. Fahmy, D. Blostein, *A graph-rewriting paradigm for discrete relaxation : application to sheet-music recognition*. Int. Journal of Pattern Recognition and Artificial Intelligence, Vol. 12, No. 6, pp 763-799, 1998.

[5] M.V. Stückelberg, C. Pellegrini, M. Hilaro, *An Architecture for Musical Score Recognition using High-Level Domain Knowledge*. 4th ICDAR Conference, vol. 2, pp 813-818, 1997.

[6] F. Rossant, *A Global method for music symbol recognition in typeset music sheets*. Pattern Recognition Letters 23 (10), pp. 1129-1141, 2002.

[7] D.Dubois, H. Prade, *Fuzzy Sets and Systems : Theory and Applications*. Academic Press, New-York, 1980.

[8] I. Bloch, H. Maître, *Fusion of Image Information under Imprecision*, In B. Bouchon-Meunier, Ed. , Aggregation and Fusion of Imperfect Information, Series Studies in Fuzziness, Physica Verlag, Springer, pp 189-213,1997.

[9] D. Blostein, L. Haken, *Using diagram generation software to improve diagram recognition : a case study of music notation,* IEEE Trans. on PAMI, Vol.21, No.11, pp 1121-1135, 1999.

[10] F. Rossant, I. Bloch, *A fuzzy model for optical recognition of musical scores*. Fuzzy Sets & Systems 141, pp165-201, 2004.

[11]SmartScore 3.2 Pro Demo, http://www.musitek.com/