



HAL
open science

Reconnaissance de la musique basée sur la perception artificielle

Andranik Tanguiane

► **To cite this version:**

Andranik Tanguiane. Reconnaissance de la musique basée sur la perception artificielle. First General Assembly of the European Society for the Cognitive Science of Music, 1991, Trieste, Italie. pp.24. hal-00910608

HAL Id: hal-00910608

<https://hal.science/hal-00910608>

Submitted on 3 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reconnaissance de la musique basée sur la perception artificielle

A. S. Tanguiane

ACROE-LIFIA, 46 av. Félix Viallet, 38031 Grenoble, France
Tel. (33) 76.57.46.70. Fax. (33) 76.57.46.02

Abstract

L'objectif est de développer un système informatique pour la notation automatique de la musique. En plus des méthodes déjà connues, il est proposé d'implémenter l'approche "perception artificielle". Les données obtenues par traitement du signal sont autorganisées afin de séparer les formes avant leur reconnaissance par des méthodes d'intelligence artificielle. Dans un certain sens, la fonction du modèle est similaire à la perception d'objets en peinture abstraite sans reconnaissance explicite.

L'information d'entrée est représentée en termes de formes génératives et de leurs transformations. Parmi toutes les représentations possibles nous choisirons la représentation la moins complexe. La complexité est comprise au sens de Kolmogorov, i.e. comme la totalité de la mémoire requise. Par la représentation la plus efficace, nous obtenons la description des données sémantiques, révélant certaines relations causales.

Notamment, l'approche est appliquée à la séparation des voix et au suivi de rythme/tempo. Les spectres sonores sont représentés en termes des sous-spectres génératifs, provenant de la décomposition de musiques polyphoniques en parties, et des accords en notes. Les événements temporels sont aussi représentés en termes de formes rythmiques génératives. Le critère de la représentation la moins complexe casse l'ambiguïté qu'il y a à reconnaître le rythme relativement au tempo, qui est reconnaissable relativement au rythme.

Le modèle explique aussi certains phénomènes dans la perception du ton et du temps musical. Les applications à la vision par ordinateur, à la reconnaissance de la parole et à la simulation de la pensée abstraite sont discutées.

Contents

1	Introduction	2
2	La perception artificielle	3
2.1	Remarques générales	3
2.2	Le principe de perception corrélative	4

2.3	Le modèle mathématique de la perception corrélative	6
2.4	Résumé de la perception artificielle	9
3	Applications à la reconnaissance de la musique	9
3.1	La séparation des voix	9
3.2	La reconnaissance du rythme et le suivi du tempo	16
3.3	La psychoacoustique et la théorie de la musique	18
4	Autres applications	20
4.1	La reconnaissance de la parole	20
4.2	Le traitement d'images	20
4.3	La modélisation de la pensée abstraite	21
5	La recherche à poursuivre	21
6	Conclusions	22
	Références	22

1 Introduction

La notation automatique de la musique est souhaitable dans la pédagogie musicale, dans l'enregistrement, dans la publication de la musique, dans la composition, dans l'exécution à l'aide d'ordinateur et dans la recherche musicologique.

Selon Chafe et al. (1982), Mont-Reynaud & Goldstein (1985) et Chafe & Jaffe (1986), les problèmes les plus difficiles surgissent dans la séparation des voix et dans le suivi du tempo.

L'état contemporain de la recherche est passé en revue par (Tanguiane, 1990).

Cet article traite de l'élaboration d'un système informatique pour la reconnaissance de la musique. Il est proposé d'implémenter le modèle de la perception corrélative qui est formulé par Tanguiane (1987; 1988a-b; 1990). Ce modèle est basé sur l'autorganisation de données afin d'obtenir leur représentation la moins complexe. De telles représentations coïncident souvent avec les formes de perception, et c'est pourquoi nous appelons notre approche "la perception artificielle".

Dans la section 2, "La perception artificielle", nous exposons les grandes

lignes du modèle de la perception corrélative, qui est utilisée pour la séparation des formes avant leur identification.

Dans la section 3, "Les applications à la reconnaissance de la musique", nous montrons comment notre modèle peut être utilisé dans la séparation des voix, dans la reconnaissance du tempo et du rythme, et comment il explique certains phénomènes psychoacoustiques et certaines règles de la théorie de la musique.

Dans la section 4, "Autres applications", nous discutons les extensions possibles de notre approche à la reconnaissance de la parole, au traitement d'images et à la modélisation de la pensée abstraite.

Dans la section 5, "La recherche à poursuivre", nous énumérons les problèmes théoriques et pratiques qu'il faut considérer.

Dans "Conclusions" nous récapitulons les arguments en faveur de la réalisation du projet.

2 La perception artificielle

2.1 Remarques générales

Nous distinguons deux stades dans la reconnaissance de formes:

- (a) la *séparation d'objets*, i.e. le groupement de données en messages;
- (b) l'*identification d'objets*, i.e. la mise en correspondance des messages déjà séparés avec les concepts connus.

Par exemple, le premier stade correspond à la distinction des lignes, des taches en peinture abstraite, mais l'identification de leur correspondance avec des objets habituels relève du second stade.

Les méthodes utilisées dans la reconnaissance de formes sont basées sur l'identification directe d'objets connus dans des flux de données sans passer par des stades intermédiaires. Nous proposons plutôt une technique spéciale visant à séparer les objets avant leur identification. Le modèle de la perception corrélative est basé sur certains principes généraux et sur les propriétés de la perception humaine plutôt que sur la connaissance particulière des formes à séparer.

Dans un certain sens nous parlons du traitement de données qui précède l'application de méthodes d'intelligence artificielle. Autrement dit, nous parlons du modèle de la perception "non-intelligente". Pour séparer notre

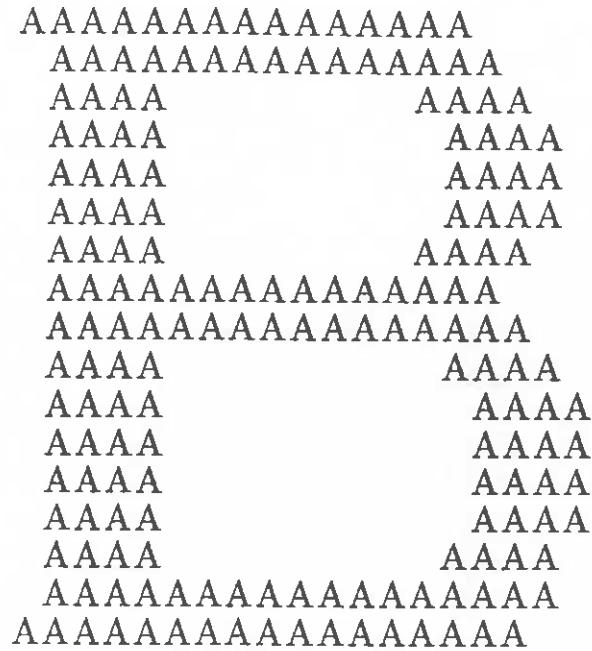


Fig. 1. La forme de haut niveau de *B* composée par les formes de bas niveau de *A*

approche de l'approche d'intelligence artificielle, nous l'appelons *perception artificielle*. Les deux approches ne sont pas juxtaposées, mais sont considérées comme complémentaires.

2.2 Le principe de perception corrélative

Par "*perception corrélative*" nous entendons la capacité de la perception à grouper des stimuli et à construire des formes de relations entre ces groupes. Ces groupes de stimuli et les relations entre eux s'appellent respectivement *formes de bas niveau* et *de haut niveau*.

Par exemple, on peut voir dans la fig. 1 une collection de pixels (stimuli) lesquels forment une série de petites lettres *A* (formes de bas niveau) qui à leur tour forment une grande lettre *B* (forme de haut niveau). Au lieu de retenir tous les pixels, il est plus efficace de retenir la configuration d'une seule lettre *A* puis le contour de *B*.

Une autre propriété d'une telle représentation est qu'il est possible de reconnaître des formes de haut niveau sans avoir reconnu les formes de bas niveau. On peut imaginer qu'au lieu de *A* nous pourrions utiliser un symbole inconnu (non reconnaissable), mais la forme de *B* serait quand même

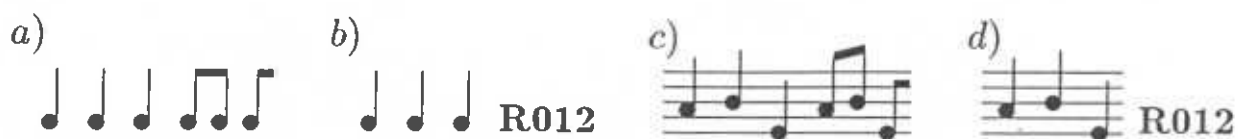


Fig. 2. Quatre représentations du même rythme

Table 1
La complexité de représentation des événements temporels de la Fig. 2

	Représen- tation a)	Représen- tation b)	Représen- tation c)	Représen- tation d)
Complexité de la forme rythmique	6 bytes	3 bytes	12 bytes	6 bytes
Complexité de sa trans- formation	0 bytes	4 bytes	0 bytes	4 bytes
Complexité totale	6 bytes	7 bytes	12 bytes	10 bytes

reconnaissable par les relations entre ces symboles inconnus. Comme nous allons le montrer, la reconnaissance des accords et des intervalles n'a pas besoin de la reconnaissance de la tonalité des notes, ce qui correspond bien à la perception humaine.

Donc, nous cherchons la représentation de données en termes de messages itératifs, ou d'éléments génératifs et de leurs transformations. La raison formelle d'une telle représentation est que, dans certain cas, elle est moins complexe que la sauvegarde des données elle-mêmes. La complexité de données est comprise au sens de Kolmogorov, i.e. comme la totalité de la mémoire requise pour sauvegarder l'algorithme de génération des données (Calude 1988).

Nous allons illustrer l'idée de complexité avec un exemple de perception du rythme. Considérons la forme rythmique de la fig. 2a. Cette succession d'événements pourrait être perçue comme la répétition des trois premiers événements exécutés deux fois plus vite. La représentation cor-

respondante est montrée fig. 2b, où **R012** désigne l'appel de l'algorithme de répétition **R** avec ses paramètres: commencer au temps **0**, répéter **1** fois, exécuter **2** fois plus vite. Cependant, cette succession d'événements est perçue comme une seule forme rythmique longue plutôt qu'une forme courte répétée. Cela signifie que la représentation de la fig. 2b est inadéquate. Si le même rythme est présenté dans le contexte mélodique montré dans la fig. 2c, la sensation de répétition devient beaucoup plus claire. Maintenant, la représentation de la fig. 2d semble adéquate (on peut observer le même effet en perception du thème d'une fugue, qui est bien reconnaissable, même s'il est exécuté en diminution ou en augmentation).

Pour expliquer une telle ambiguïté dans la perception du rythme, estimons la complexité des représentations. Supposons, qu'un octet est nécessaire pour coder une durée, et que deux octets sont nécessaires pour coder une durée avec sa tonalité. Supposons aussi que pour coder l'appel d'algorithme de répétition nous avons besoin de quatre octets. Avec ces conventions, la complexité des représentations est évaluée dans la table 1. On peut voir que dans le cas de rythme pur, sa représentation en extension est moins complexe, tandis que dans le contexte mélodique sa représentation comme répétition est plus efficace.

Nous pouvons dire que le principe de hiérarchie dans la représentation de données s'est pourvu du feed-back, qui guide le processus de recherche de la représentation la moins complexe.

2.3 Le modèle mathématique de la perception corrélative

Puisque nous cherchons une représentation de données en termes de variations de quelques éléments génératifs, ou de sous-messages similaires, il est naturel d'appliquer les méthodes d'analyse de corrélation de données. Pourtant, puisque nous essayons de reconnaître les transformations d'éléments génératifs, nous avons besoin d'appliquer l'analyse de corrélation compte tenu des distorsions de données.

Par exemple, pour reconnaître un objet en mouvement dans une succession de cadres cinématographiques, nous pouvons appliquer l'analyse de corrélation aux images instantanées déformées. Pour mettre un objet en corrélation avec lui-même dans une succession d'images, ces images doivent être déplacées, tournées, etc.

Evidemment, il est impossible de réaliser l'analyse de corrélation en essayant toutes les déformations possibles. Pour exécuter la recherche

directionnelle des déformations qui fournissent un haut degré de corrélation de certains messages, nous proposons la *méthode de la résolution variée*, laquelle est un peu similaire à *scale-space filtering* (Witkin 1983).

L'idée principale de cette méthode est de réduire la résolution de l'image pour rendre négligeable les petites déformations d'éléments génératifs, puis de révéler les messages corrélatifs, et enfin de restituer graduellement la résolution de l'image en ajustant (en déformant) l'image pour ne pas perdre la corrélation obtenue.

On peut suivre le fonctionnement de la méthode avec un exemple de reconnaissance de la quasi-périodicité (élément génératif) dans la succession des événements temporels de la fig. 3.

Considérons la succession suivante des attaques enregistrées avec une unité temporelle de 1/20 sec.:

$$t = 0, 10, 19, 30$$

(fig. 3a). Définissons la fonction "d'événements à une unité près", i.e. pour tous $t = 0, 1, \dots, 30$ posons

$$s(t) = \begin{cases} 1 & \text{pour } t = 0, 10, 19, 30, \\ 0 & \text{ailleurs.} \end{cases}$$

La fonction d'événements $s(t)$ forme une succession binaire (fig. 3b). Nous pouvons essayer de découvrir la période p dans cette succession par les pics de la fonction d'autocorrélation

$$C(p) = \sum_t s(t-p) s(t).$$

Cependant, si nous appliquons l'analyse d'autocorrélation à la succession donnée, nous ne découvrons pas d'autocorrélation et nous ne reconnaissons aucun périodicité (voir table 2). Pour surmonter cette difficulté, nous baissons la résolution de la représentation de la fig. 3b et obtenons la représentation de la fig. 3c. Maintenant il y a des pics d'autocorrélation (voir table 2), et nous pouvons reconnaître une périodicité. Une fois que les événements temporels sont connus, nous déplaçons ces événements pour élever l'autocorrélation. Le résultat de cette procédure est montré fig. 3d. Nous restituons ensuite la résolution, et nous obtenons la succession de la fig. 3e avec la description des fluctuations du tempo correspondant aux déplacements des événements temporels dans notre procédure.

Il faut remarquer, que cette méthode peut être réalisée sur un réseau de neurones avec plusieurs niveaux traités en parallèle.

Nous pouvons dire que le fond mathématique de notre approche est la *reconnaissance de la similitude*. Habituellement, la similitude est seulement mesurée par le degré d'identité; c'est l'idée principale de l'analyse de corrélation. Au contraire, nous mesurons la similitude par le degré de dissimilitude. Pour cela nous considérons la déformation qui est nécessaire pour rendre les objets identiques. Nous utilisons aussi l'analyse de corrélation, mais pour trouver les zones où les objets sont identiques. Puis nous considérons les relations (entre les autres: les différences) entre les objets. Dans un certain sens, nous considérons un concept complémentaire (la différence au lieu de l'identité) sur le domaine complémentaire (là où les objets sont différents, et non là où ils sont identiques). Ceci correspond bien à deux niveaux de représentation de données: au premier niveau nous avons les données, le second niveau est réservé pour la description de leurs déformations.

2.4 Résumé de la perception artificielle

Ainsi, pour séparer des formes de "perception artificielle" il est proposé de:

- (a) représenter les données en termes d'éléments génératifs et de leur transformations;
- (b) trouver les éléments génératifs par la recherche de messages corrélatifs avec déformations de données;
- (c) exécuter la recherche directionnelle des déformations de données par la méthode de la résolution variée;
- (d) surmonter l'ambiguïté dans la représentation des données avec l'aide du critère de la représentation la moins complexe.

3 Applications à la reconnaissance de la musique

3.1 La séparation des voix

La *voix* est associée au groupe de partiels qui bougent en parallèle relativement à l'axe de fréquences pour une échelle logarithmique (fig. 4). Leur trajectoire dynamique s'appelle la *ligne mélodique*, ou la partie polyphonique.

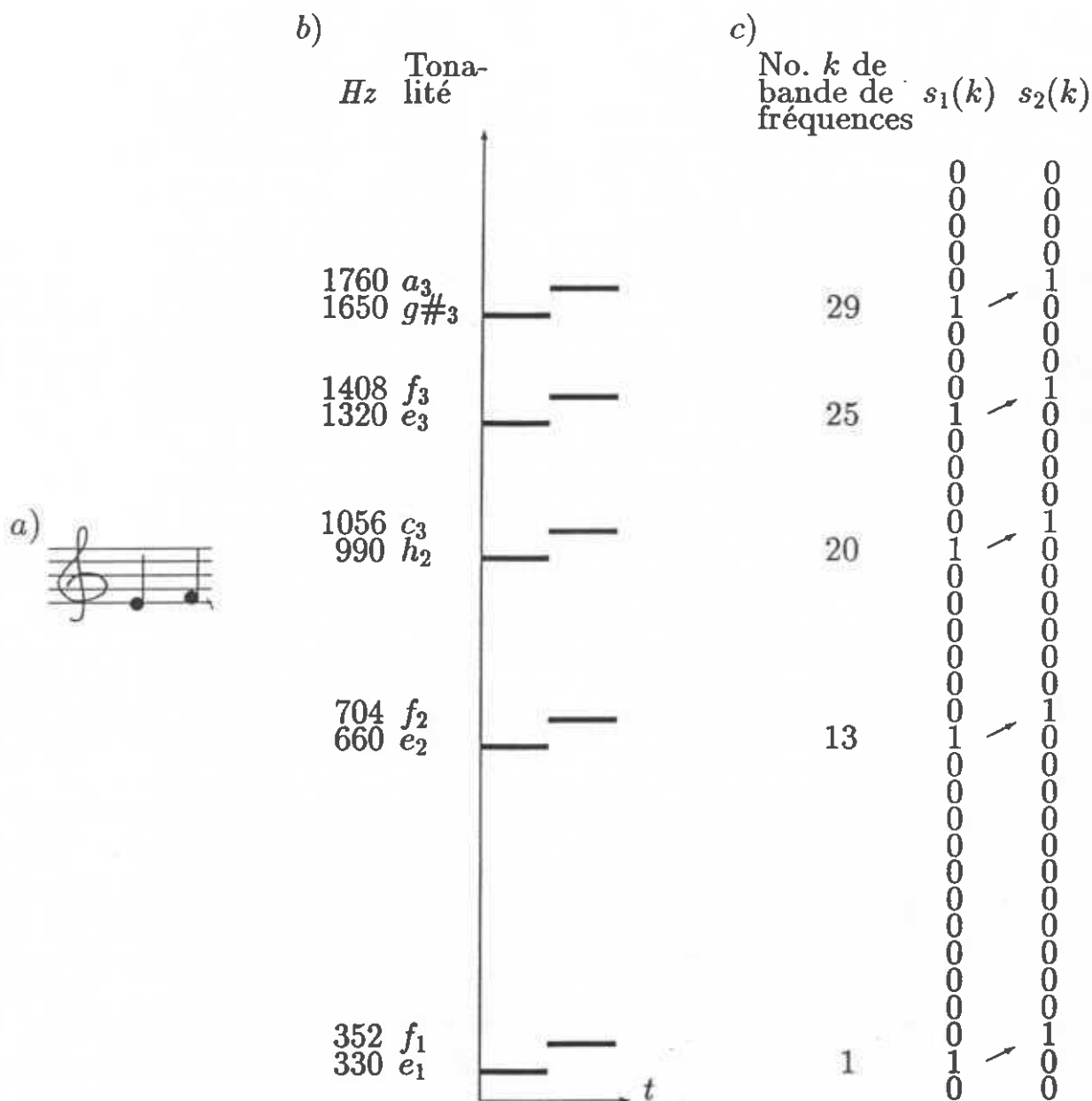


Fig. 4. La représentation spectrale des tons:

- a) les tons e_1 et f_1 en notation habituelle;
- b) leurs spectres auditifs (sur une échelle de fréquences logarithmique) pour les voix harmoniques avec 5 partiels successifs;
- c) les spectres auditifs discrets coupés des tons avec la résolution de fréquences à un demi-ton près — les successions binaires $s_n(k)$; les flèches montrent le mouvement parallèle des partiels.

De notre point de vue, la ligne mélodique est une forme de haut niveau, engendrée par les formes de bas niveau lesquelles sont désignées par les notes. Le contour statique engendré par une répétition d'un groupe de partiels est la forme de haut niveau d'*accord*. Pour séparer les voix (reconnaître des accords) il faut représenter les spectres sonores sur le plan d'éléments génératifs et leur trajectoires (contours).

Il faut remarquer que notre approche de la séparation de voix est basée sur l'idée de similitude entre les voix, par opposition à l'approche de Mont-Reynaud & Mellinger (1989) basée sur l'idée de leur dissemblance, qui proviens, disons, d'un vibrato asynchrone.

Pour illustrer la procédure de la reconnaissance des accords, considérons les deux accords simples montrés dans la fig. 5a.

Représentons les spectres des accords comme dans la fig. 5c. Donc, un accord est identifié avec une succession binaire

$$\{s_n(k)\},$$

où n est le numéro d'accord donné dans la succession des accords considérés (dans notre exemple $n = 1, 2$), k est le numéro de bande de fréquences dans le spectre (dans notre exemple $k = 1, \dots, 34$), et

$$s_n(k) = \begin{cases} 1 & \text{si il y a des partiels dans la } k\text{-ième bande de fréquences} \\ 0 & \text{ailleurs.} \end{cases}$$

Nous allons reconnaître des intervalles mélodiques entre le n -ième accord et l'accord suivant par les pics de la fonction de corrélation

$$C_{n,n+1}(i) = \sum_k s_n(k)s_{n+1}(k+i);$$

$$C_{n,n+1}(-i) = \sum_k s_n(k)s_{n+1}(k-i) = \sum_k s_n(k+i)s_{n+1}(k).$$

Si cette fonction a un pic au point i , nous supposons que les sous-spectres corrélatifs correspondent aux tons similaires et forment donc l'intervalle de i demi-tons. Dans ce cas le subspectre corrélatif s'appelle le *groupe génératif des partiels* de l'intervalle mélodique donné.

L'analyse de corrélation des spectres auditifs de nos deux accords peut être vu dans la fig. 6. Les successions $s_n(k)$ sont montrées avec les déplacements pour rendre visible la corrélation. La valeur de corrélation au déplacement i correspond au nombre de 1 coïncidants dans les successions déplacées. Par exemple, $C_{1,2}(-1) = 0$ parce qu'il y a une coïncidence de 1 dans la colonne $s_1(k+1)$ et la colonne $s_2(k)$.

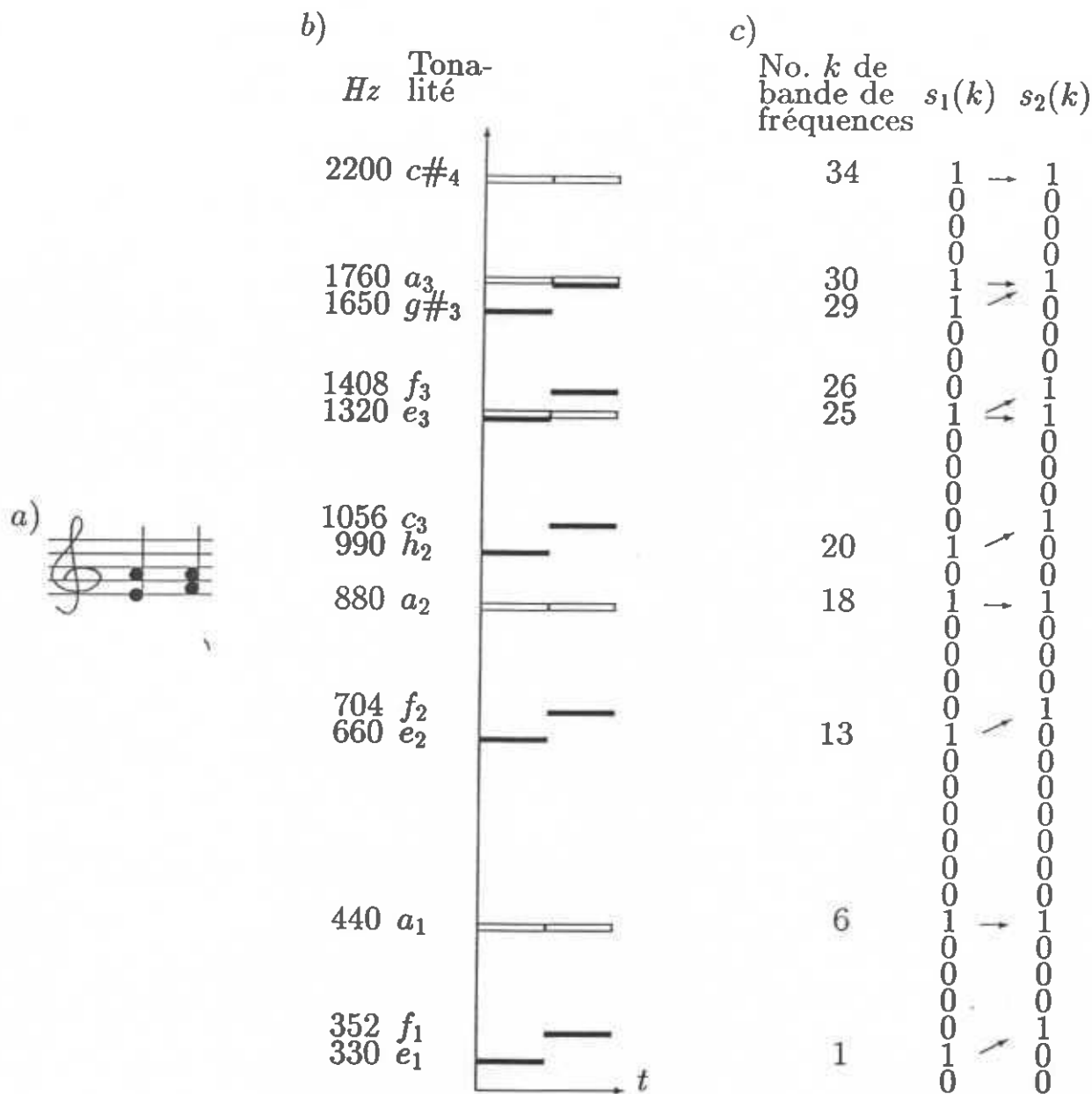


Fig. 5. Les représentations spectrales des accords:

- les accords $(e_1; a_1)$ et $(f_1; a_1)$ en notation habituelle;
- leurs spectres auditifs (sur une échelle de fréquences logarithmique) pour les voix harmoniques avec 5 partiels successifs; (les partiels du ton grave et du ton aigu sont désignés respectivement par les rectangles noirs ou blancs);
- les spectres auditifs discrets coupés des tons avec la résolution de fréquences à un demi-ton près — les successions binaires $s_n(k)$; les flèches montent le mouvement parallèle des partiels.

Tonalité	No. k de bande de fréquences	$s_1(k+i)$							$s_2(k+i)$									
$c\#_4$	34	1							1									
		0	1						0	1								
		0	.	1					0	.	1							
		0	.	.	1				0	.	.	1						
a_3	30	1	.	.	.	1			1	.	.	.	1					
$g\#_3$	29	1	1	.	.	.	1		1	1	.	.	.	1				
		0	1	1	0	1	1			
		0	.	1	0	.	1			
f_3	26	0	.	.	1	1	.	.	0	.	.	1	1	.	.			
e_3	25	0	.	.	1	1	.	.	0	.	.	1	1	.	.			
		1	.	.	.	1	1	.	1	.	.	.	1	1	.			
		0	1	.	.	.	1	1	0	1	.	.	.	1	1			
		0	.	1	.	.	.	1	0	.	1	.	.	.	1			
		0	.	.	1	.	.	1	0	.	.	1	.	.	1			
c_3		0	.	.	.	1	.	.	0	.	.	.	1	.	.			
h_2	20	1	1	.	1	1	.			
		0	1	1	0	1	1			
a_2	18	1	.	1	1	.	1			
		0	1	.	1	.	.	.	0	1	.	1	.	.	.			
		0	.	1	.	1	.	.	0	.	1	.	1	.	.			
		0	.	.	1	1	.	.	0	.	.	1	1	.	.			
f_2		0	.	.	.	1	1	.	0	.	.	.	1	1	.			
e_2	13	1	1	1	1	1	1			
		0	1	1	0	1	1			
		0	.	1	0	.	1			
		0	.	.	1	.	.	.	0	.	.	1	.	.	.			
		0	.	.	.	1	.	.	0	.	.	.	1	.	.			
		0	1	.	0	1	.			
a_1	6	1	1	1	1			
		0	1	0	1			
		0	.	1	0	.	1			
		0	.	.	1	.	.	.	0	.	.	1	.	.	.			
f_1		0	.	.	.	1	.	.	0	.	.	.	1	.	.			
e_1	1	1	1	.	1	1	.			
		0	1	1	0	1	1			
		i	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
		$C_{n,n}(i)$	9	1	1	1	2	5	1	3	9	1	1	1	5	2	1	2
		$C_{1,2}(i)$									5	5	0	1	1	5	1	1
		$C_{1,2}(-i)$	5	0	1	1	6	1	1	2								

Fig. 6. L'analyse de corrélation des spectres des accords (les points sont équivalents à 0).

La table 3 montre les intervalles mélodiques inférieurs à la quinte ($-7 \leq i \leq 7$).

Le groupe de partiels génératif de chaque intervalle mélodique est désigné par les index de bandes de fréquences. Bien que nous n'utilisions pas la tonalité pour la reconnaissance d'intervalles, nous désignons les intervalles reconnus par les notes conventionnelles, correspondant au partiel le plus grave dans le groupe de partiels corrélatif.

De la même façon nous reconnaissons les intervalles harmoniques dans le n -ième accord par les pics de la fonction d'autocorrélation

$$C_{n,n}(i) = \sum_k s_n(k)s_n(k+i).$$

Les tables 4 et 5 montrent les intervalles harmoniques les plus significatifs inférieurs à la quinte ($1 \leq i \leq 7$) dans le premier et le second accord.

Il faut remarquer que le premier intervalle de la table 3 (la tierce descendante $i = -4$) est reconnu par le groupe génératif qui contient 6 partiels. Ce groupe consiste en 5 harmoniques qui forment les tons a_1 du premier accord et f_1 du second et un partiel ajouté par accident. C'est la cinquième harmonique du ton e_1 dans le premier spectre (1650 Hz) et la troisième harmonique du a_1 dans le second spectre (1320 Hz).

Pour des cas plus complexes, quand les accords ont plusieurs notes, la corrélation de spectres auditifs peut être élevée en points qui ne correspondent pas à des intervalles réels. Cela arrive à cause de coïncidence de partiels provenant de tons différents. Heureusement, la structure des groupes de partiels correspondant aux tons réels est répétée régulièrement, ce qui en fait une forme spectrale stable. Au contraire, un groupe de partiels corrélatif qui ne correspond pas aux tons réels a une structure spectrale aléatoire. Par conséquent, pour reconnaître les intervalles véritables, il faut examiner les groupes de partiels génératifs (trouvés par l'analyse de corrélation) et supprimer les groupes accidentels qui n'ont pas de structure stable (qui n'apparaissent qu'une seule fois). Pour reconnaître les formes spectrales stables, il faut appliquer l'analyse de corrélation aux groupes de partiels génératifs différentes entières peuvent être éliminés après l'analyse de formes génératives.

Notre approche a été testée avec la série de 48 expériences simplifiées sur la reconnaissance d'intervalles mélodiques et harmoniques du 130-ième choral à quatre voix de J.S.Bach (fig. 7). Ce choral était considéré comme une succession de 24 accords correspondant aux verticales

Table 3

Les intervalles mélodiques les plus significatifs entre les deux accords de la fig. 5 ($-7 \leq i \leq 7$)

Corrélation $C_{1,2}(i)$	Intervalle i	Groupe de partiels génératif	Nombre de harmoniques successifs	Notation de l'intervalle
6	-4	6 18 25 29 30 34	5	$(a_1; f_1)$
5	0	6 18 25 30 34	5	$(a_1; a_1)$
5	1	1 13 20 25 29	5	$(e_1; f_1)$
5	5	1 13 20 25 29	5	$(e_1; a_1)$
2	-7	13 25	2	$(e_2; a_1)$

Table 4

Les intervalles harmoniques les plus significatifs dans le premier accord de la fig. 5 ($1 \leq i \leq 7$)

Autocorrelation $C_{1,1}(i)$	Intervalle i	Groupe de partiels génératif	Nombre de harmoniques successifs	Notation de l'intervalle
5	5	1 13 20 25 29	5	$(e_1; a_1)$
3	7	6 13 18	2	$(a_1; e_2)$
2	4	25 30	1	$(e_3; g\#_3)$
1	1	29	1	$(g\#_3; a_3)$

Table 5

Les intervalles harmoniques les plus significatifs dans le second accord de la fig. 5 ($1 \leq i \leq 7$)

Autocorrelation $C_{2,2}(i)$	Intervalle i	Groupe de partiels génératif	Nombre de harmoniques successifs	Notation de l'intervalle
5	4	2 14 21 26 30	5	$(f_1; a_1)$
2	5	21 25	2	$(c_3; f_3)$
2	7	14 18	1	$(f_2; c_3)$
1	1	25	1	$(e_3; f_3)$



Fig. 7. Le 130-ième choral de J.S.Bach utilisé pour les expériences sur la reconnaissance des intervalles

harmoniques. Les spectres des accords étaient évalués et analysés avec des hypothèses différentes. Les expériences différaient en type de voix (harmonique, i.e. formée par les partiels avec le rapport de fréquences $1 : 2 : 3 : \dots$, ou inharmonique, i.e. formée par les partiels avec le rapport de fréquences $1 : \sqrt{2} : 2 : \sqrt{2^3} : \dots$), nombre de partiels par voix (5, 10, ou 16), résolution de fréquences (à 1, $1/2$, $1/3$ et $1/6$ demi-tone près), et en restriction sur la grandeur d'intervalles considérés (à 11 demi-tons—pour éviter l'autocorrélation d'octave, ou sans restriction). Le taux de reconnaissance dans nos expériences était toujours supérieur à 98% (Tanguiane 1987; 1988a-b; 1989a-b; 1991a).

3.2 La reconnaissance du rythme et le suivi du tempo

La difficulté dans la reconnaissance du rythme est causée par la relativité de la perception de rythme. Il est vraiment impossible de reconnaître un rythme en reconnaissant les événements temporels séparés sans tenir compte du tempo.

Nous définissons des *formes rythmiques* de bas niveau comme des successions semblables d'événements temporels. Les formes rythmiques peuvent soit être déformées par des déviations du tempo, soit être variées par l'élaboration rythmique, ou les deux. La forme de haut niveau, déterminée par les relations entre les formes rythmiques de bas niveau, est la *courbe du tempo*.

Puisque la similarité de formes rythmiques est comprise relativement à un certain tempo et que, en même temps, les formes rythmiques sont considérées comme les unités de référence pour le suivi du tempo, nous

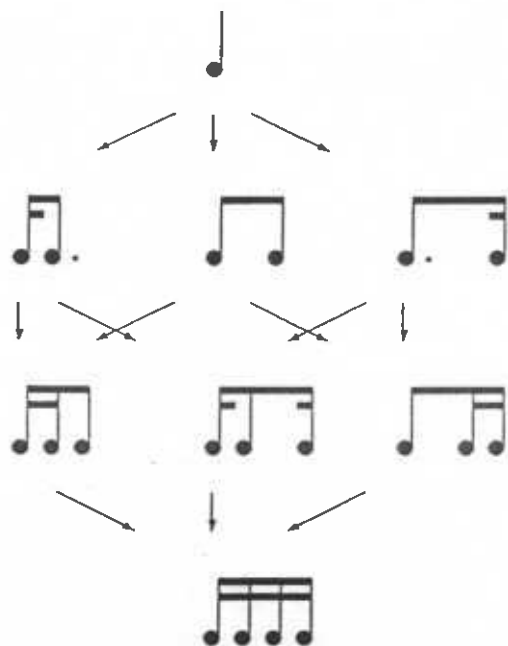


Fig. 8. L'élaboration de la forme rythmique d'une seule noire

appliquons le critère de moindre complexité des données pour surmonter cette interdépendance du rythme et tempo.

Pour la recherche directionnelle de formes rythmiques génératives, nous proposons certaines règles de segmentation et une sorte de grammaire du rythme. La segmentation est supposée être basée sur l'accentuation. Il y a plusieurs cas quand nous supposons l'accentuation (Tanguiane 1991b). En gros, une durée accentuée est celle qui est plus longue que les durées adjacentes.

Selon Mont-Reynaud & Goldstein (1985), nous définissons l'élaboration de la forme rythmique comme une subdivision de leurs durées en conservant la pulsation originale. Le graphe des élaborations successives (subdivisions) d'une seule durée est montré dans la fig. 8. Par segmentation et élaboration, nous décrivons le rythme comme une structure itérative sur plusieurs niveaux. On peut voir un exemple d'une telle description fig. 9 où la partie de caisse claire du *Bolero* de Ravel est présentée. Par les lettres *A, B, C, D* nous désignons les formes génératives de niveaux différents. Par $E(A)$ nous désignons l'élaboration (subdivision) de forme *A*. Ici, chaque forme générative de haut niveau est la somme des formes de bas niveau. Considérant l'élaboration comme la forme générative trans-

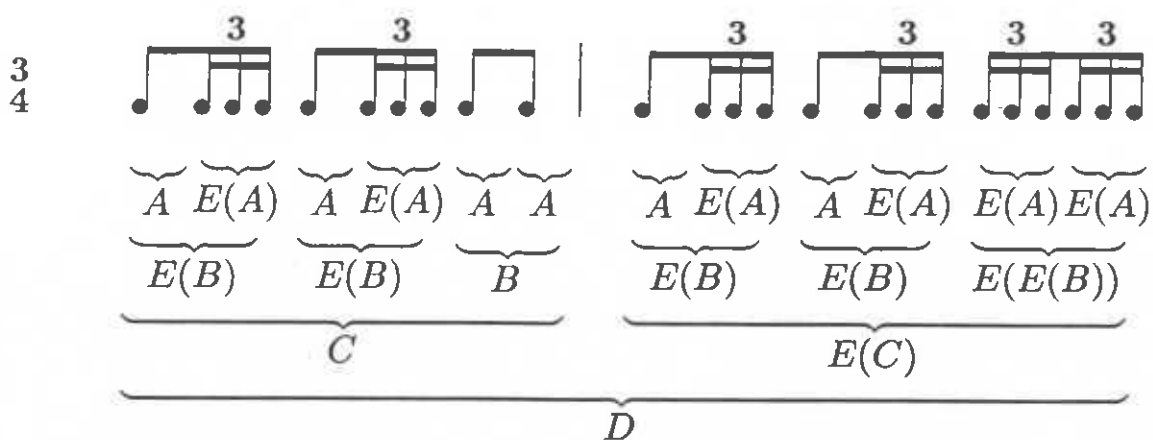


Fig. 9. La structure itérative sur plusieurs niveaux du rythme

formée, nous découvrons la structure itérative sur plusieurs niveaux de ce rythme.

On trouvera plus de détails sur la reconnaissance du rythme à l'aide de notre approche dans Tanguiane (1990, 1991b-c).

3.3 La psychoacoustique et la théorie de la musique

Nous pouvons expliquer certaines règles de la théorie de la musique comme les prescriptions pour aider à la compréhension auditive de la structure de la musique. Dans notre modèle ceci correspond à une simplification de la reconnaissance de la musique.

Par exemple, nous pouvons expliquer l'interdiction des quintes parallèles dans la théorie du contrepoint. Les quintes parallèles suggèrent le mouvement parallèle des partiels qui sont associés aux voix. Cela rend leur séparation plus difficile. En conséquence, l'effet psychoacoustique est plus "timbral" qu'harmonique. Cela peut troubler l'homogénéité de la texture musicale, et c'est pourquoi les quintes parallèles sont évitées (ou sont utilisées intentionnellement).

L'effet timbral de voix parallèles est utilisé dans les orgues quand plusieurs tuyaux accordés sont mis en marche par une seule touche. L'usage de cet effet dans l'orchestration est montré fig. 10. On peut voir que cinq voix parallèles synthétisent une nouvelle voix qui est perceptible comme une seule voix au lieu de cinq.

The image displays a page of a musical score for the Bolero by Maurice Ravel. The score is arranged in a standard orchestral format with multiple staves. The instruments and parts shown are:

- Flutes (Fl. I & II):** Both parts play a melodic line with parallel motion, marked *pp* (pianissimo).
- Clarinets (Cl. I & II):** Both parts play a melodic line with parallel motion, marked *pp*.
- Violins (Viol. I & II):** Both parts play a melodic line with parallel motion, marked *f solo* (forte solo).
- Cello (Cello):** Plays a melodic line with parallel motion, marked *p* (piano).
- Double Bass (Cb.):** Plays a melodic line with parallel motion, marked *p*.
- Woodwinds (Oboe, Bassoon, Clarinet in Bb, Saxophone):** These parts are present but mostly silent or play simple accompaniment.
- Harps (Harpe):** Two harps are present, playing a rhythmic accompaniment.
- String Ensemble (Violins, Violas, Cellos, Double Basses):** The string ensemble provides a harmonic and rhythmic foundation.

The score illustrates the use of parallel motion (voix parallèles) in the woodwinds and strings, which is a characteristic feature of Ravel's Bolero.

Fig. 10. L'usage de voix parallèles dans le *Bolero* de Ravel

Le modèle clarifie aussi la nature et la fonction d'audition d'intervalles. L'audition d'intervalles peut être comprise comme la capacité de reconnaître la distance entre des tons qui sont *similaires en structure spectrale*. Autrement dit, l'audition d'intervalles n'est rien d'autre que la perception corrélative dans le domaine des fréquences. Cette condition est une généralisation plus qu'une restriction. Vraiment, tous les sons musicaux (avec une tonalité claire) satisfont cette condition. D'un autre côté, l'idée de distance est applicable aux sons *similaires* sans tonalité, comme les sons de cloches, ou même les bruits. Par conséquent, nous pouvons éliminer la tonalité du son de la définition d'audition d'intervalles.

Selon notre modèle, la fonction d'audition d'intervalles est la décomposition des flux acoustiques et le suivi des processus acoustiques parallèles (Tanguiane 1987; 1988a; 1991; 1992). C'est extrêmement important pour l'orientation en environnement acoustique.

4 Autres applications

Il faut remarquer que les applications suivantes n'ont pas été testées, mais sont présentées ici pour enrichir la discussion.

4.1 La reconnaissance de la parole

Pour reconnaître les phonèmes, on propose souvent de reconnaître la contribution des parties différentes du système vocal au signal de parole. Puisque les parties du système vocal ont leur propres caractéristiques acoustiques, nous pouvons poser le problème de la séparation des "lignes polyphoniques" du signal de parole, où chaque ligne correspond à une certaine partie du système vocal.

Pour reconnaître un phonème, il faut reconnaître "l'accord" produit par les parties du système vocal activées. Il semble que notre modèle de suivi des voix polyphoniques est adaptable pour cet usage.

4.2 Le traitement d'images

Notre approche peut être généralisée des tableaux de données à une dimension (comme des spectres sonores ou des successions d'événements temporels) à des tableaux à deux dimensions pour les applications au traitement d'images. Elle peut être utilisée pour la séparation d'objets

en dynamique (par analogie avec la séparation des voix) et la reconnaissance des contours (par analogie avec la reconnaissance des accords).

Par exemple, pour isoler un objet en mouvement, il est proposé d'appliquer l'analyse de corrélation aux images instantanées successives. Pour trouver l'objet et sa trajectoire il est nécessaire de trouver les déformations d'images qui fournissent un haut degré de corrélation de ces images. La méthode de la résolution variée est bien adaptée.

4.3 La modélisation de la pensée abstraite

Notre approche est destinée à révéler les relations stables entre les blocs des données. En considérant les relations stables comme des données, on peut révéler les relations stables entre les relations stables, etc. De cette façon nous passons aux concepts abstraits correspondant aux invariants stables de représentation des données.

Les "formes de perception" obtenues de l'analyse des données courantes peuvent concourir avec les formes mémorisées obtenues d'une expérience précédente. La représentation de l'information d'entrée en termes de formes de la mémoire peut être plus efficace qu'en termes de formes de perception courantes. Cela peut expliquer la différence entre la perception des mêmes choses par des personnes pourvues d'expériences différentes.

5 La recherche à poursuivre

Notre approche a montré son efficacité dans des expériences simplifiées sur la séparation des voix, le suivi du tempo et pour l'explication de certains phénomènes psychoacoustiques et des règles de la théorie de la musique.

La continuation de la recherche comprend ce qui suit:

1. Le développement de la théorie mathématique de la perception artificielle. Notamment, il est peut être possible de formuler les conditions où la représentation des données en termes d'éléments génératifs et de leurs transformations est plus efficace que la mémorisation de la totalité des données. Cette théorie peut utiliser certains résultats de l'algorithmique et de la théorie de la complexité.
2. L'implémentation de l'approche proposée de la perception artificielle pour la séparation des objets dans un système musical de traitement

de signal. Ce système peut se pourvoir des moyens d'intelligence artificielle pour l'identification d'objets et il peut être conjugué avec un éditeur de partitions pour la notation automatique de la musique exécutée.

3. Le développement du matériel et du logiciel pour le traitement parallèle des données musicales. Ils peuvent être utilisés pour le développement plus lointain d'une version temps-réel du système mentionné.

6 Conclusions

Ce projet intègre la recherche théorique et l'implémentation. Puisque l'approche de la perception artificielle est déjà testée sur de nombreuses expériences, il est proposé de l'élaborer davantage pour implémenter un système informatique de traitement de signal musical. Les moyens pour la reconnaissance de la musique peuvent être utilisés, premièrement, dans la recherche sur l'informatique musicale, et ensuite dans la pratique musicale générale.

Le développement de la théorie de la perception artificielle et les expériences sur la reconnaissance de la musique peuvent préparer le fond pour les extensions de l'approche proposée à certaines autres branches de l'informatique, comme la reconnaissance de la parole, le traitement d'images et l'intelligence artificielle.

Références

- Calude C. (1988)** *Theories of Computational Complexity*. Amsterdam, etc.: North-Holland.
- Chafe C. & Jaffe D. (1986)** Source Separation and Note Identification in Polyphonic Music. *Proc. of the IEEE-IECEJ'ASJ Int. Conf. on Acoustics, Speech, and Signal Processing, Tokyo, 1986*, vol. 2, 1289-1292.
- Chafe C., Mont-Reynaud B. & Rush L. (1982)** Toward an Intelligent Editor of Digital Audio: Recognition of Musical Constructs. *Computer Music J.* 6(1), 30-41.

- Mont-Reynaud B. & Goldstein M. (1985)** On Finding Rhythmic Patterns in Musical Lines. *Proc. 1985 Int. Computer Music Conf.* San Francisco: Computer Music Association, 391–397.
- Mont-Reynaud B. & Mellinger D. (1989)** Source Separation by Frequency Co-Modulation. *Proc. of the First Int. Conf. on Music Perception and Cognition, Kyoto, Japan, 17–19 October, 1989*, 99–102.
- Tanguiane A.S. (1987)** *Reconnaissance des accords pour notation automatique de la musique polyphonique*. Preprint. Moscou: La commission ethnomusicologique de l'Union de compositeurs de l'URSS, le Centre de calcul de l'Academie des sciences de l'URSS. (En Russe).
- Tanguiane A.S. (1988a)** Reconnaissance des accords, audition d'intervalles et la théorie de la musique. Dans: Alekseev E., Andreeva E., Borođa M., & Tanguiane A. (Eds.) *Les méthodes quantitatives dans l'ethnomusicologie et dans la recherche musicale*. Moscou: Le compositeur sovietique, 155–186. (En Russe).
- Tanguiane A.S. (1988b)** An Algorithm of Recognition of Chords. In: *Proc. of the 14th Int. Computer Music Conference, Cologne, September 20–25, 1988*. Cologne: Feedback-Studio-Verlag, 199–210.
- Tanguiane A.S. (1989a)** La reconnaissance des accords basée sur le modèle de l'audition d'intervalles. *Doklady Akademii Nauk SSSR*, 308(3), 552–556. (En Russe).
- Tanguiane A.S. (1989b)** A Model of Relativity of Perception and Its Applications to Pattern Recognition in Analysis of Performed Music. In: *Proc. of the 1st Int. Conf. on Music Perception and Cognition, Kyoto, 17–19 October, 1989*, 261–266.
- Tanguiane A.S. (1990)** Le principe de la perception corrélative et son application à la reconnaissance de formes. *Matematicheskoye Modelirovaniye*, 2(8), 90–111. (en Russe).
- Tanguiane A.S. (1991a)** Reconnaissance des accords, corrélativité de la perception, et la théorie de la musique. *Musikometrika*, 3, 163–199.
- Tanguiane A.S. (1991b)** Time Determination by Recognising Generative Rhythmic Patterns. *Musikometrika*, 4. (Forthcoming).

- Tanguiane A.S. (1991c)** Criterion of Data Complexity in Rhythm Recognition. Paper submitted to *the 17th Int. Computer Music Conf., Montreal, Quebec, Canada, October 16-20, 1991.*
- Tanguiane A.S. (1992)** *On Defining Interval, Chord, and Melodic Line.* Paper submitted to the Congress of International Musicology Society, Madrid, 3-10 April, 1992.
- Witkin A.P. (1983)** Scale-Space Filtering. *Proc. 8th Int. Joint Conf. on Artificial Intelligence, Karlsruhe, West Germany, vol. 2, 1019-1024.*