Le groupe Ecoute Interactive de la Musique de la Bibliothèque Nationale de France

Philippe Aigrain et Philippe Lepain
Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier, 118, route de Narbonne
F-31062 Toulouse Cedex, France
e-mail: aigrain@irit.fr, lepain@irit.fr
URL: http://www.irit.fr/ACTIVITES/EQ_AMI/ami.html

Abstract: The Bibliothèque Nationale de France has asked the authors to chair a working group on interactive listening to musical recordings, its perspectives and its tools. The group members are ethno-musicologists, composers, musical analysts, performers, music teachers and critics, specialists in psycho-acoustics or curators of musical archives and libraries. In the group activity, the general investigation of the praxis of listening to recorded music has been combined with the practical investigation of interactive listening using the discrete manipulation listening software SATIE developed at IRIT by the authors, and comparison with other existing software. This paper presents the group, its aims and work methods, the sofware tools it uses for experimentation, and the first conclusions of its activity. More generally, the paper argues for the usefulness of a continued research effort targeted at building listening tools.

Résumé : La Bibliothèque Nationale de France a confié aux auteurs la mission d'animer un groupe de réflexion et d'expérimentation sur l'écoute interactive de la musique enregistrée et ses outils. Les membres de ce groupe sont ethno-musicologues, compositeurs, analystes, interprètes, enseignants, critiques, psycho-acousticiens ou responsables d'archives et de bibliothèques musicales. L'activité du groupe a associé la réflexion de fond sur les mécanismes et les outils d'écoute de la musique enregistrée avec l'expérimentation pratique du logiciel d'écoute par manipulation discrète SATIE développé à l'IRIT. Cette communication présente le groupe, les objectifs qu'il s'est fixé, sa méthode de travail, les logiciels à partir desquels il expérimente, les premières conclusions de son activité, et plus généralement effectue un plaidoyer pour la poursuite d'une activité de recherche en matières d'outils pour l'écoute.

1 Le contexte institutionnel

Parmi d'autres collections, le Département de l'Audiovisuel et du Multimédia de la Bibliothèque Nationale de France (B.N.F.) gère l'une des plus importantes archives sonores musicales mondiales : ses collections comportent plus d'un million de phonogrammes, représentant plus de 300 000 heures d'enregistrements, dont plus de 90 % d'enregistrements musicaux. La plus grande part de ces collections est constituée par le dépôt légal des phonogrammes édités, qui constitue une vue d'ensemble des enregistrements sonores édités ou importés en France. Le reste est constitué d'archives sonores collectées ou ayant fait l'objets de donations, en particulier dans les domaines du jazz, des musiques traditionnelles et des enregistrements de parole. Comme la plupart des archives sonores de bibliothèque, cette collection était, dans ses locaux du square de Louvois, assez peu utilisée à travers l'écoute, la majorité des usages portant sur des recherches discographiques. La mise en place de la nouvelle Bibliothèque Nationale de France va profondément renouveler les conditions d'accès à ces collections, et, depuis l'origine du projet, une réflexion a été engagée sur le développement de nouveaux systèmes et de nouvelles formes d'écoute, en particulier pour les chercheurs et artistes pour qui l'écoute constitue une étape d'un travail. La B.N.F. possède également, au Département de la Musique, l'une des plus importantes collections de partititions et de manuscrits musicaux (et d'autres documents relatifs à la musique).

L'équipe de recherche *Analyse des Médias et Interaction* de l'IRIT[1] a été associée par la B.N.F. à la conception d'une partie de ses systèmes de consultation audiovisuels, en particulier en ce qui concerne la

[1] Institut de Recherche en Informatique de Toulouse.

consultation de documents numériques (ou numérisés) par des chercheurs effectuant un travail prolongé et analytique sur ces documents. Ce type de consultation reçoit à la B.N.F. le nom de consultation "savante". C'est dans le cadre de cette collaboration de longue haleine que la B.N.F. nous a confié début 1995 la mission d'animer un groupe d'usagers potentiels des nouveaux outils d'écoute interactive, et de donner à l'activité de ce groupe l'indispensable support concret de l'expérimentation grâce à la mise à disposition de ce groupe des logiciels d'écoute développés par notre équipe. Ces logiciels étant en cours de développement et même, pour partie, de spécification, c'est une expérience unique de conception centrée sur les usages qui a été rendue possible. Les logiciels développés par notre équipe pour l'écoute interactive visent un cadre beaucoup plus large que la consultation d'archives : nous estimons que l'écoute interactive et les manipulations qu'elle permet ont leur place dans la pratique du musicien et de l'amateur de musique au foyer comme au studio. L'écoute par des chercheurs de documents d'archives représente cependant pour nous une situation de référence et le support d'une exigence intellectuelle dans la conception des outils.

Le principe même du groupe reprend celui des groupes similaires animés par Bernard Stiegler sur la lecture informatique à la B.N.F. (1990-1993) et sur la consultation du dépôt légal de la télévision et de la radio à l'Institut National de l'Audiovisuel (depuis 1994). Nous avions participé à ces groupes[2] et avions pu tirer de leur expérience divers enseignements et critiques (voir plus bas). Grâce au prestige de la B.N.F. et au caractère original de la démarche, il a été possible, malgré l'absence de moyens financiers spécifiques, de réunir dans le groupe un ensemble exceptionnel d'intervenants (cf section 3.1). Y sont représentés, à un degré ou à un autre, aussi bien des disciplines scientifiques comme la psycho-acoustique et l'ethno-musicologie, que des pratiques artistiques (composition, interprétation, direction), techniques (ingénieurs du son), pédagogiques et critiques, ainsi que les conservateurs de collections et les éditeurs (discographiques et multimédias).

L'activité du groupe se situe dans un contexte où l'accès aux enregistrements sonores (notamment à travers des réseaux) motive d'importants projets : projet européen JUKE-BOX, projet international IDEAMA, projet de serveur du patrimoine musical contemporain de l'IRCAM, divers projets industriels de "musique à la demande", par exemple. A l'origine, ces projets portaient presque uniquement sur l'accessibilité des documents, mais une prise de conscience de l'importance des interfaces d'écoute et des modes de représentation des contenus musicaux se développe.

2 De l'existant au possible

2.1 Déplacer les limites de l'écoute analytique de la musique

Un des objectifs du groupe Ecoute Interactive de la B.N.F. consiste à étudier les conditions d'un rapprochement entre deux mondes qui évoluent parallèlement : celui des personnes confrontées à l'écoute analytique, raisonnée et argumentée au sein de leur travail quotidien, et celui des acteurs du développement des techniques de production et de traitement des enregistrements sonores. L'un n'existerait pas sans l'autre, mais les liens qui les unissent restent souvent ponctuels et mal définis. L'une des raisons du faible niveau d'échanges entre ces deux mondes a longtemps résidé dans le cloisonnement entre un marché grand public et un marché de spécialistes. Depuis l'avènement de l'enregistrement et du traitement numérique du son, les conditions de cet équilibre sont profondément bouleversées. Un ethno-musicologue travaillait hier sur un magnétophone REVOX de prix inaccessible au commun des mortels pour pouvoir monter ses extraits ou les ralentir ; aujourd'hui, un simple micro-ordinateur permet les mêmes manipulations. Si la technologie numérique permet des opérations d'un intérêt évident comme l'accès direct aux plages d'un disque compact, elle en propose d'autres qui peuvent se révéler pour le moins surprenantes comme la lecture des plages dans un ordre aléatoire ! Il convient donc de séparer le bon grain de l'ivraie, et surtout de définir de nouvelles fonctionnalités d'accès, de manipulation et d'annotation des enregistrements sonores. C'est précisément le coeur des réflexions du groupe, car en l'absence de ces fonctionnalités, l'écoute analytique rencontre des limites majeures :

- difficultés de repérage dans des enregistrements de longue durée,
- difficultés d'analyse et de comparaison liées aux limites de la mémoire auditive, ou à la perception de phénomènes très brefs ou ténus,
- difficultés à qualifier un phénomène sonore perçu, à faire la preuve de ce qu'on avance à son sujet.

[2] Aux deux pour Philippe Aigrain, à celui de l'INA pour Philippe Lepain.

Nous avons proposé au groupe comme hypothèse de départ la nécessité de créer une représentation visuelle pour permettre l'accès, la localisation, l'indexation, l'écoute et l'annotation du document sonore. Ce choix se situe dans la continuité de la tradition de la transcription descriptive en ethno-musicologie ou pour les musiques électro-acoustiques. Les ethno-musicologues ont été, pour des raisons évidentes, les pionniers de la transcription descriptive [Ell92], de la production d'une représentation visuelle de la musique telle qu'elle a été interprétée. Dès que fut battu en brèche le principe de l'application de la notation prescriptive occidentale à l'ensemble des musiques, ils se tournèrent vers des outils de transcription automatique ou assistée, dont le melograph de Charles Seeger fut l'exemple le plus réellement utilisé. Au début des années 1980, Robert Cogan [Cog84], généralisa cette approche à l'ensemble des types de musiques et au cadre informatique. La transcription des musiques électro-acoustiques donna lieu plus récemment, en particulier au G.R.M. de l'I.N.A. [Bes91] à des travaux qui introduirent non seulement la représentation visuelle mais aussi l'écoute interactive comme moyen de la produire. La visualisation du son était donc un choix naturel. Mais vers quel type de représentation s'orienter et surtout quelle forme d'interaction y associer?

Les outils informatiques actuels comme les éditeurs sonores ou certains programmes de composition utilisent des représentations classiques de forme d'onde ou sonographique (évolution spectrale du son au cours du temps). Ces représentations graphiques pèchent malheureusement par les possibilités d'indexation temporelle très réduites, et la difficulté à y maîtriser les phénomènes d'échelles temporelles très variables qui font l'objet de l'écoute analytique. Elles n'en constituent pas moins des éléments de référence auxquelles l'effort de définir de nouvelles formes d'écoute interactive à support visuel doit sans arrêt se confronter : faisons-nous vraiment mieux qu'un spectrographe interactif? Charles Seeger, dans un article fameux [See91], a posé le cadre fondamental pour les réflexions sur les formes de représentation du contenu sonore, et en particulier le choix entre le caractère linéaire des représentations (facilitant la mise en correspondance temporelle entre deux enregistrements et les changements d'échelle de vision) et leur caractère symbolique (facilitant l'abstraction d'éléments pertinents pour une écoute donnée et simplifiant le repérage). Les travaux du groupe se situent, on le verra, dans la recherche d'un compromis renouvelé entre ces deux dimensions.

La musique organise des éléments sonores en un flux temporel continu. L'utilisateur qui désire se repérer dans un document musical a besoin de désigner tel ou tel élément musical, et de le situer par rapport aux autres. L'informatique rend possible des mécanismes de manipulation d'unités discrètes, depuis longtemps appliqués au texte et dans une moindre mesure à la vidéo, et dont nous avons proposé un modèle général pour l'ensemble des médias dans [AJLL95]. Les outils actuels de visualisuation d'enregistrements sonores ne proposent que le repérage "classique" en heure-minute-seconde-fraction de seconde. Ce codage temporel linéaire ne correspond pas aux besoins de l'écoute interactive. Nicolas Meeùs a souligné à quel point un outil permettant l'annotation et la segmentation du document musical peut être utile pour les analystes [Mee94]. De nombreux chercheurs ont décrit des méthodes de segmentation des documents musicaux ou cherché à modéliser la segmentation implicite des auditeurs humains [LJ83, Del92, Tho85, Imb87, Bre90, Haw93, Tod95]. Ces travaux peuvent servir de point d'appui pour la conception de méthodes de segmentation automatique d'enregistrements musicaux dont nous verrons plus loin l'intérêt pour la production d'une représentation morphologique du document sonore, servant de base à sa manipulation interactive.

2.2 Nouvelles approches de la perception assistée

Au démarrage du travail du groupe Ecoute Interactive, nous lui avons proposé un programme [Aig95, Lep95], un ensemble d'objectifs à plus ou moins long terme. L'intelligibilité des travaux du groupe suppose de retranscrire les éléments essentiels de ce programme.

2.2.1 Fournir une aide à l'appéhension du matériau musical

On sait aujourd'hui qu'il n'est sans doute pas possible de créer une représentation objective du contenu musical d'un enregistrement, mais ne peut-on pas se fixer une tâche moins ambitieuse, mais aussi utile : fournir une représentation qui serve de médiation avec ce contenu musical, qui soit avec lui dans un rapport d'analogie maîtrisable pour l'auditeur. Qu'il soit possible de localiser un passage musical sans passer obligatoirement par une phase d'écoute (identification claire sur la représentation des phénomènes de répétition et de régularité par exemple). Que l'accès au passage ainsi localisé ne demande que des manipulations évidentes et ergonomiques de désignation sur la reprśentation. Que les données sous-jacentes

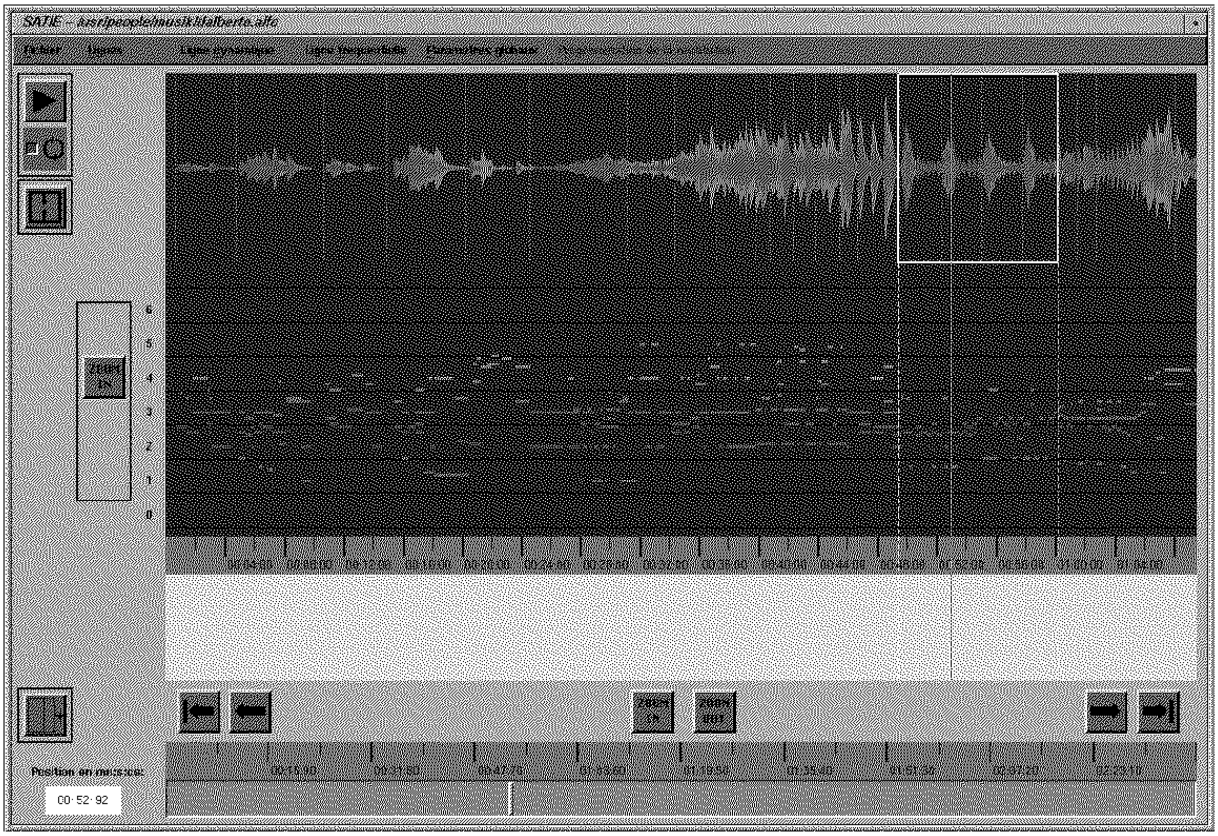

FIG. 1 - : Utilisation du logiciel SATIE en sélection d'extraits dans la ligne dynamique (enregistrement de Michel Dalberto jouant la sonate op. 42, D. 845 en La Mineur de Franz Schubert) - CD Denon CO-73787

à la représentation soient adaptées à la recherche d'un objet par son contenu par exemple mélodique (saisi à l'aide d'un clavier MIDI). Que des éléments perceptivement importants soient lisibles sur la représentation même lorsqu'ils se situent à des échelles temporelles très différentes (attaques d'instruments, vibrato, tremolo, par exemple).

Au-delà de la localisation et de la désignation se pose le problème du contrôle de la restitution. Plusieurs modes d'écoute présentent un intérêt potentiel. La consultation simplifiée simule l'action des touches d'un magnétophone. La consultation rapide balaye le document en écoutant le début de chaque objet. La consultation détaillée utilise la représentation visuelle et en contrôle finement les paramètres. La programmation de la restitution joue sur l'enchaînement temporel de différents extraits. Enfin, toujours dans le domaine de la restitution, des outils de perception par la synthèse peuvent être utiles (mixage du document avec une pulsation percussive, re-synthèse d'une mélodie à partir des résultats d'une procédure de détection des hauteurs perçues).

Un des buts essentiels de la perception assistée est de permettre la comparaison de deux documents. Cela implique de les mettre en correspondance temporelle et perceptive et à tout le moins d'avoir une représentation cohérente d'un document à l'autre (choix d'un mode de visualisation absolu et non relatif à un document).

Assister l'usager dans son travail consiste également à lui permettre de définir ses propres objets temporels (limites et types) et d'annoter le document, c'est à dire de positionner des notations textuelles et graphiques sur la représentation du document. Le résultat de son travail doit pouvoir être reporté dans un carnet de notes, qu'il s'agisse d'extraits (dans les limites du statut juridique et économique des documents) ou de représentations du document. Enfin, la nécessité de sauvegarder des historiques et versions du travail effectué pose également des problèmes complexes.

2.2.2 La proposition d'utiliser des unités discrètes et des lignes multiples de représentation

Au-delà de ce programme fonctionnel, notre proposition initiale au groupe s'effectuait à travers le logiciel SATIE développé dans notre équipe depuis plusieurs années (cf figure 1). L'espace de cet article ne permettant pas une présentation complète de ce logiciel, nous décrivons ci-dessous ses principes essentiels :

- la projection du contenu du document sur plusieurs "lignes" qui correspondent chacune à l'évolution d'un paramètre perceptif dans le temps. On distingue les lignes dynamique (variations d'amplitude), fréquentielle (hauteur perçue), de timbre (entrée et sortie de classes de timbres), métrique (éléments rythmiques et pulsation) et spatiale (analyse de spatialisation et effets sonores). Cette approche nous permet d'éviter une représentation surchargée qui aurait l'ambition de projeter toutes les dimensions perceptives sur un seul espace (par exemple 3D). De plus, les différents angles de vue obtenus sur un document peuvent correspondre à des écoutes culturellement différentes. Vincent Dehoux, l'un des membres du groupe, nous a fait remarquer que l'identification d'une séquence jouée par un instrument percussif s'effectuera certainement par des critères métriques pour un occidental, mais par des critères de timbre ou de mélodie pour un pygmée d'Afrique centrale. Dans l'état actuel de notre logiciel, les différentes lignes sont à des niveaux très variables d'achèvement ou de maturité : la ligne dynamique est complètement définie, la ligne fréquentielle a fait l'objet de nombreuses explorations, les lignes de timbres et métriques en sont au stade des recherches préliminaires, la ligne spatiale relève de la simple imagination,
- le repérage dans le document est facilité par la présence de deux lignes supplémentaires : la ligne de temps locale qui indique la position relativement à la page visible du document et la ligne de temps globale qui situe l'endroit où l'on se trouve dans le document lui-même. L alocalisation est renforcée par une organisation de chacune des lignes en objets discrets multi-échelles qui peuvent être sélectionnés de manière ergonomique par désignation (*clic* souris) et glissement (*drag*) avec gravité. Les objets sont fournis par une procédure de segmentation à base "intrinsèque" dans la terminologie de [Mee94] : définie sur chacune des lignes, elle met en relief les éléments principaux d'un paramètre perceptif et non une combinaison de l'ensemble des paramètres perceptifs. Elle gagne ainsi en neutralité par rapport au matériau analysé, en indépendance par rapport à une école de pensée en matière de segmentation. Ces segmentations n'ont pas prétention à constituer des analyses de l'oeuvre, mais seulement des briques de manipulation pouvant servir à l'écoute et à la construction de segments d'analyses par l'usager,
- la production automatique des segmentations propres à une ligne, qui peuvent ensuite faire l'objet d'un ajustement par l'utilisateur. Dans nos travaux les différentes segmentations sont issues d'une première phase de discrétisation du flux auditif en unités relativement microscopiques que nous appelons *trames acoustiques*, à l'aide d'une méthode de micro-segmentation [LAO95]. Trois niveaux de segmentation sont ensuite définis :
  - le premier niveau est constitué d'*éléments* typés par rapport aux paramètres perceptifs essentiels de la musique (dynamique, fréquence, timbre, rythme). Un élément peut être composé de plusieurs trames : par exemple un accord joué par un orchestre symphonique réunit les trames correspondant aux transitoires d'attaque, aux phases plus ou moins stationnaires et aux phases d'extinction des harmoniques présentes. Un élément n'est qu'une subdivision d'un motif et n'a pas de sens musical. Dans l'exemple ci-dessus, l'accord est un élément perceptif mais dénué de tout sens musical en dehors de son contexte,
  - le deuxième niveau est composé d'objets sélectionnables appelés *motifs*. Ils sont caractérisés par un ensemble d'éléments définis à partir de lois de détection de ruptures ou de groupements. Ces lois pourront être aussi bien exprimées sous forme de règles du type de celles proposées par Lerdahl et Jackendoff [LJ83] que sous la forme de détection de ruptures par détection de contours multi-échelle comme proposé par Neil Todd pour la dynamique [Tod95]. Dans tous les cas on se limitera aux règles et méthodes qui peuvent capturer les éléments perceptifs assez indépendants d'un contexte culturel (un sous-ensemble des règles locales dans le cas des règles de Lerdahl et Jackendoff),
  - le troisième niveau correspond aux grandes divisions appelées sections qui structurent une pièce de musique. Par exemple, si un morceau dure 3 minutes et qu'on observe deux silences de 3 secondes chacun assez espacés alors que la densité moyenne des événements sonores est d'une note par demi-seconde, on inférera l'existence de trois sections dynamiques. Comme les éléments et les motifs, les sections possèdent un type, sont sélectionnables et peuvent être jouées.

3 Un exemple de conception centrée sur les usages

La conception de systèmes interactifs en y associant étroitement des usagers potentiels (cf par exemple [Nor86, NL95]) connait une vogue importante ... dans le discours. En pratique, il s'agit d'un exercice difficile, du fait de l'existence de contraintes contradictoires :

- l'activité critique des usagers et leur capacité de proposition ne peut vraiment s'exercer que face à une proposition déjà opérationnelle qui, en particulier s'il s'agit d'une réalisation innovante dont il n'existe pas encore d'équivalent, a déjà nécessité un effort de développement important. La nécessité de disposer d'un véritable système interactif (et non de maquettes d'apparence) est particulièrement forte pour des domaines comme l'écoute interactive, qui sont fortement multi-modaux (mécanismes d'interaction associant l'audition, la vision, et le geste avec une granularité temporelle fine),
- il y a toute chance que les réactions et suggestions des usagers conduisent à refondre sur des points importants le système en cours de conception, nécessitant un nouvel investissement en développement (et l'acceptation de revenir sur des choix de façon douloureuse pour l'ego des concepteurs),
- l'ensemble du processus se déroule dans un contexte temporel serré qu'il s'agisse de compétition scientifique ou industrielle ou des contraintes de calendrier d'ouverture d'un organisme, de complétion d'une thèse, etc.

Nous nous sommes efforcés de trouver un compromis satisfaisant entre ces contraintes en nous situant dans des environnements de développement suffisamment souples (nos logiciels sont développés sous X/Motif sur des stations Silicon Graphics Indy), et en adoptant une conception extrêmement modulaire.

3.1 Composition et méthode de travail du groupe

La composition du groupe “Ecoute Interactive de la Musique” a été la suivante :

- animateurs : Philippe Aigrain et Philippe Lepain, informaticiens, IRIT,
- participants à l'ensemble de la démarche :
  - pour la Bibliothèque Nationale de France: Gérald Grunberg, Directeur du Département de l'Audivisuel et du Multimédia ; Catherine Massip, Directeur du Département de la Musique ; Dominique Chrismann puis Philippe Jedrecy (chefs de projet “Système Audiovisuel”) ; Pierre Echegaray, informaticien, secrétaire du groupe,
  - Simha Arom et Vincent Dehoux, ethno-musicologues, LACITO-CNRS,
  - Steve McAdams et Daniel Pressnitzer, psycho-acousticiens, Université René Descartes et IRCAM,
  - Gérard Condé, critique musical,
  - François Delalande, compositeur, INA-GRM.
- participants occasionnels ou invités pour des séances de travail spécifiques :
  - Chrisophe Coin, interprète (violoncelle et viole de gambe), professeur d'interprétation et directeur d'ensemble,
  - Philippe Krumm, interprète et directeur de collection d'édition discographique (musiques traditionnelles),
  - Frank Jaffres, ingénieur du son (prise de son, montage, édition discographique),
  - Greg Faris, ingénieur du son (restauration),
  - plusieurs responsables de l'Association Française de Pédagogie Musicale préparant un CD-ROM sur Nocturnes de Debussy.

Le groupe s'est réuni une journée par mois pendant un an. Les réunions ont été consacrées à 3 types de travaux :

- des échanges de vues généraux sur les principes et les effets de l'écoute interactive,

– des exposés décrivant de façon détaillée des pratiques d'écoute (y compris non interactives), exposés alimentés notamment par un questionnaire sur les pratiques d'écoute conçu par les auteurs,

– l'expérimentation en commun de divers logiciels d'écoute interactive, sur des documents appartenant à un corpus défini par les membres du groupe (voir ci-dessous). Outre le logiciel SATIE, l'expérimentation a porté sur le logiciel sonographique temps réel SONOSCOPE[3], également développé dans notre équipe, le logiciel de montage sonore SoundEditor de Silicon Graphics, et divers logiciels de domaine public pour le traitement sonore.

En dehors des réunions certains membres du groupe ont utilisé les logiciels de façon plus approfondie sur des documents les intéressant particulièrement. Chaque réunion a donné lieu à un compte-rendu détaillé, et l'activité du groupe se conclut provisoirement par la production d'un rapport réunissant les contributions de chacun des membres (cf section 5).

3.2 Corpus et problématiques d'expérimentation

Chacun des membres du groupe a choisi dans les collections du Département de l'Audiovisuel et du Multimédia, un ou quelques documents sonores, en indiquant les motivations qui le poussaient à entamer un travail à leur sujet. Voici quelques exemples de ces documents et des problématiques associées :

– comparaison de plusieurs interprétations du monologue de Didon dans Didon et Enée de Purcell,

– comparaisone de la lisibilité polyphonique de plusieurs orchestrations des canons de l'Offrande Musicale de J.S. Bach,

– mise en relief de différentes analyses d'enregistrements de musique avec percussions et voix chuchotées et chantées des pygmées du Gabon et centrafricains,

– analyse cherchant à reconstruire le mode de production des Variations pour une Porte et un Soupir de Pierre Henry,

– comparaison de différentes éditions discographiques d'enregistrements de Caruso.

L'ensemble des documents ainsi sélectionnés ont été réunis pour former un corpus commun d'expérimentation accessible à l'ensemble du groupe.

4 L'apport du groupe et l'évolution fonctionnelle du logiciel d'écoute

Nous donnons ci-dessous les principaux points de débat du groupe et les enseignements que nous nous sommes efforcés d'en tirer pour la conception de logiciels d'écoute interactive.

4.1 Le principe même de l'écoute interactive avec représentation visuelle

Un débat récurrent s'est développé, en particulier à l'initiative de Gérard Condé, sur le principe même d'utiliser une représentation visuelle comme support de l'écoute, et sur la tendance au découpage de l'écoute (navigation, répétitions, retours en arrière) induite par son caractère interactif. Ce débat a permis de clarifier le rôle des représentations visuelles : celles-ci sont des médiations et, comme telles, sont nécessairement biaisées, le problème étant alors celui de la maîtrise de ce biais dans l'usage. La possibilité de re-synthétiser le contenu sonore correspondant aux représentations, permettant de mesurer ainsi la distance qui sépare la représentation du contenu brut serait évidemment l'outil idéal de cette maîtrise. Celle-ci n'étant pas encore présente dans notre logiciel, la possibilité de paramétrage et de modes de représentations multiples[4] permet le même type d'appropriation des spécificités et des biais de divers types de représentations visuelles.

[3] SONOSCOPE est un logiciel sonographique temps réel avec calcul de pages (une page de spectrogramme est toujours calculée en avance, ce qui permet l'écoute avec un curseur se déplaçant sur la représentation). SONOSCOPE ne permet pas la définition et l'indexation d'extraits, ni d'opérations de traitement ou de re-synthèse.

[4] allant par exemple pour la ligne fréquentielle d'une représentation presque sonographique à des tentatives de suivi de voix en passant par la représentation en rouleaux de piano.

De même, l'appropriation par l'utilisateur des outils d'écoute interactive se traduit d'abord par une tendance au morcellement de l'écoute, qui est évidemment nuisible à la perception du contenu musical. Mais passé ce stade de découverte de l'outil, un équilibre s'installe entre l'indispensable écoute linéaire, respectueuse du temps musical, et l'écoute exploratoire et analytique de segments.

4.2 Manipulation discrète et segmentations : quelles segmentations faut-il automatiser ?

La désignation et la manipulation rapide d'extraits musicaux qui est rendue possible par l'existence de segmentations sous-jacentes représente une révolution dans l'ergonomie de l'écoute dont il est difficile de mesurer l'importance sans usage concret. La simple possibilité de sélection par glissement avec gravité à l'échelle des segments sous-jacents à une ligne donnée permet de définir un extrait avec un gain de temps de l'ordre du facteur 3 par rapport au positionnement sans gravité. Or ce type de manipulation se reproduit des dizaines de fois dans une séance d'écoute interactive. Mais deux conditions doivent être remplies pour que les bénéfices de la manipulation discrète se réalisent : la représentation visuelle doit être suffisamment intelligible pour que l'usager comprenne ce qu'il est en train de sélectionner, et les segments sous-jacents doivent inclure parmi leurs limites, les limites de l'extrait qu'il veut définir. C'est ici que l'approche multi-lignes joue tout son rôle :

- dans la ligne de temps, la gravité est à l'échelle des éléments dynamiques, c'est à dire des notes et masses sonores. Cette gravité fine permet une sélection qui est déjà beaucoup plus rapide que la sélection sans gravité, mais cependant moins performante qu'une sélection à l'échelle des motifs (sauf bien sûr si l'on est en visualisation très détaillée sur le plan temporel). Ce mode de sélection est indispensable car il est le seul qui fournisse une garantie de finesse suffisante pour satisfaire tous les besoins de définition d'extraits pour l'écoute. Le très grand nombre de limites à définir pour les éléments (densité moyenne de 4 par seconde) impose absolument que cette segmentation soit effectuée par un processus automatique,
- dans les autres lignes, la sélection à l'échelle des motifs, illustrée dans la figure 1, satisfait au niveau de performance maximale une proportion importante des besoins de définition d'extraits. Mais elle n'est efficace que si la qualité des segments sous-jacents est excellente. Nous avons sur ce point été amenés à relativiser partiellement l'importance d'une production automatique fiable de ces segments :
  - en raison de la difficulté évidente de cette production,
  - mais surtout grâce à l'implantation de mécanismes ergonomiques originaux pour la définition ou le réglage par l'utilisateur des limites de ces segments. Le plus important de ces mécanismes concerne la rétroaction auditive pendant le positionnement des limites. Celui-ci, basé sur l'écoute en boucle de segments de durée variable, permet aussi bien le balayage accéléré du document (y compris en arrière) que le réglage fin des limites d'un segment avec une très bonne intelligibilité. Ce mécanisme est particulièrement efficace associé à la gravité à l'échelle des éléments.

  Compte tenu de cette possibilité, l'effort nécessaire pour un utilisateur pour positionner des limites correspondant à des segments d'une à cinq ou dix secondes sur tout un document est parfaitement réaliste dans l'optique d'un travail approfondi sur le document. Nous n'en poursuivrons pas moins nos travaux de recherche sur l'amélioration des techniques de segmentation automatique à l'échelle des motifs, mais il est rassurant que l'usage de la manipulation discrète ne soit pas subordonné à leur débouché.

4.3 La traversée des échelles temporelles

Le logiciel SATIE dans sa version actuelle ne propose pas d'échelle temporelle plus fine que un pixel pour 32 ms (1p/32ms). Cette échelle ne permet pas aux utilisateurs d'observer en detail les phénomènes sonores comme avec le logiciel SONOSCOPE. Ce choix de départ, justifié par la volonté de dissocier la notion de vue globale sur le document et de vue de détail d'une portion du document est maintenant remis en question par le groupe d'usagers. L'utilisateur souhaite aussi bien appréhender le document dans son ensemble que focaliser son attention sur un évènement sonore très court (analyser la qualité des attaques d'un orchestre par exemple). Les informations perceptives utiles n'étant pas les mêmes aux différentes échelles, il convient donc d'adapter les représentations à celles-ci et de définir des politiques de déformation de la représentation lors du passage d'une échelle à l'autre. La visualisation sous forme de plusieurs lignes simplifie ce problème car elle le reporte sur chaque paramètre perceptif individuellement. Dans la ligne dynamique, la traversée des échelles (y compris très fines) est relativement aisée car elle repose sur la transformation d'une courbe d'évolution des amplitudes. Le problème est bien plus délicat en ce qui concerne la ligne fréquentielle : une vision sonographique assez bien adaptée aux échelles allant de 1p/1ms a 1p/30ms devient vite brouillonne à de plus grandes échelles ; une vision de type rouleau de piano (projection des harmoniques sur des hauteurs perçues en 1/2 tons) assez bien adaptée aux échelles allant de 1p/30ms a 1p/100ms devient également confuse à l'échelle du document. Il faut donc modifier (de manière parfois radicale) le mode de représentation de certaines lignes lors de la traversée des échelles temporelles.

4.4 L'importance des fonctionnalités de traitement sonore

Afin de comparer plusieurs documents sonores, deux fonctions essentielles ont été demandées par les membres du groupe : le pitch-shifting et le time-stretching. La première concerne surtout la mise au même diapason de deux objets différents. La seconde, très utile pour l'étude de rythmes complexes (ralentissement) permet également l'alignement temporel de deux passages musicaux. Une autre classe de traitements sonores concerne l'écoute analytique de détail. Aux échelles de visualisation les plus fines, nous désirons accentuer ou extraire par re-synthèse une portion de la représentation. Enfin, jouer sur des paramètres comme la brillance ou simuler un environnement acoustique particulier constitue pour l'usager un moyen de comprendre plus finement les relations qui unissent représentation et écoute.

4.5 Tout reste à inventer pour l'annotation et la production d'exemples sonores

L'une des premières questions des usagers membres du groupe, et en particulier de ceux qui doivent effectuer une communication publique des travaux motivant l'écoute d'enregistrements (pédagogues, critiques, analystes) a été de savoir quels résultats de leur travail ils pourraient réutiliser par la suite dans d'autres lieux : pouvaient-ils imprimer les représentations et les annoter manuellement, pouvait-on "enregistrer" un exemple sonore (boucle de comparaison de deux extraits, par exemple) avec la représentation visuelle correspondante, pouvait-on importer une portion d'image de partition numérisée pour l'intégrer à la représentation ? Or sur ce plan, tout reste à faire, à la fois sur le plan juridique ou d'accords contractuels (délimitation d'un droit de citation sonore dans le respect des droits de propriété intellectuelle), et sur le plan technique. Sur ce dernier plan, de nombreuses questions doivent être explorées :

- quelle part de l'annotation est-il plus pertinent de rendre possible à l'intérieur même du logiciel d'écoute et quelle part organiser par exportation dans d'autres logiciels d'extraits de document ou de représentations ?
- pour l'annotation interne au logiciel d'écoute, la gestion de la portée temporelle des annotations, de leur disposition spatiale en regard des représentations, de leur expressivité graphique méritent des travaux qui sont à peine abordés aujourd'hui,
- il existe encore de grandes incertitudes sur les normes et formats de codage de ces documents multimédias très particuliers que sont des exemples sonores visuellement annotés. Or seule l'existence de formats d'échange peut rendre possible l'exportation des résultats du travail du chercheur et sa communication.

5 Perspectives : les pragmatiques de l'écoute

S'il est bien trop tôt pour savoir quelle forme prendra demain l'écoute interactive de la musique, ou pour savoir si elle restera un outil de spécialiste ou se diffusera beaucoup plus largement au sein des amateurs et praticiens, l'activité du groupe a donné dès à présent à chacun de ses membres une certitude : le continent de l'écoute musicale mérite d'être exploré plus avant, et de l'être au niveau du concret de ses pratiques, de son appareillage, de sa temporalité, de son contexte humain. Les travaux déjà cités de psycho-acousticiens ou de sociologues de la réception comme Emmanuel Pedler peuvent être complétés par l'attention aux situations techniques de l'écoute (par exemple dans les métiers techniques du son), les outils techniques eux-mêmes nous fournissant de nouveaux moyens d'observation. La démarche de Pierre Schaeffer [Sch66] peut trouver de nouvelles dimensions, comblant, selon la formule de François Delalande, une partie du fossé que Pierre Schaeffer décrivait entre le support physique du son et la réalité subjective de l'écoute. L'écoute musicale peut devenir consciente de son concret sans s'appauvrir, et ceux qui la mettent en oeuvre peuvent guider le développement d'outils qui, autrement, n'obéiront qu'à des logiques de facilité technique. C'est à cette fin que nous avons proposé à la Bibliothèque Nationale de France l'organisation d'un colloque au printemps 1997 sur les pratiques et les outils de l'écoute.

Références

[Aig95] Philippe Aigrain. Pour l'écoute interactive de la musique enregistrée. IRIT - Bibliothèque Nationale de France, Janvier 1995. Rapport introductif du groupe de travail Ecoute interactive savante de la musique enregistrée.

[AJLL95] Philippe Aigrain, Philippe Joly, Philippe Lepain, and Véronique Longueville. Representation-based user interfaces for the audiovisual library of year 2000. In *Proc. IS&T/SPIE'95 Multimedia Computing and Networking*, volume 2417 of *SPIE Conference Proceedings*, pages 35–45, San Jose, February 1995.

[Bes91] Dominique Besson. La transcription des musiques électroacoustiques : que noter, comment et pourquoi? *Analyse Musicale*, 24:37–41, Juin 1991.

[Bre90] Albert S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT Press, 1990.

[Cog84] Robert Cogan. *New Images of Musical Sounds.* Harvard University Press, Cambridge, 1984.

[Del92] Irène Deliège. Perception et analyse de l'oeuvre musicale : points de rencontre. *Analyse Musicale*, 26:7–13, Février 1992.

[Ell92] Ter Ellington. Transcription. In Helen Myers, editor, *Ethno-Musicology: An Introduction*, Norton-Grove Handbooks in Music, pages 110–152. W. W. Norton, London, 1992.

[Haw93] Mike Hawley. *Structure Out of Sound.* Phd dissertation, MIT Media Laboratory, Cambridge, Mass, USA, 1993.

[Imb87] Michel Imberty. De la perception du temps musical à sa signification psychologique : à propos de la Cathédrale Engloutie de C. Debussy. *Analyse Musicale*, 6:28–37, Janvier 1987.

[LAO95] Ph. Lepain and R. André-Obrecht. Micro-segmentation d'enregistrements musicaux. In *Actes des Journées d'Informatique Musicale*, volume 95-13, pages 81–90, LAFORIA, Université Paris 7, Avril 1995.

[Lep95] Philippe Lepain. Conception d'un système interactif d'écoute et de consultation de documents musicaux : axes de réflexion. IRIT - Bibliothèque Nationale de France, Mars 1995. Note du groupe de travail Ecoute interactive savante de la musique enregistrée.

[LJ83] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music.* MIT Press, 1983.

[Mee94] Nicolas Meeús. De la forme musicale et de sa segmentation. *Musurgia*, 1(1):7–23, 1994.

[NL95] William M. Newman and Michael G. Lamming. *Interactive Systems Design.* Addison-Wesley, 1995.

[Nor86] D. A. Norman. Cognitive engineering. In D. A. Norman and S. W. Draper, editors, *User-Centered System Design*, pages 31–65. Laurence Erlbaum, Hillsdale, NJ, USA, 1986.

[Sch66] Pierre Schaeffer. *Traité des objects musicaux.* Seuil, 1966.

[See91] Charles Seeger. Notation prescriptive et notation descriptive (traduction). *Analyse Musicale*, 24:6–12, Juin 1991.

[Tho85] Lasse Thoresen. Un modèle d'analyse auditive. *Analyse Musicale*, 1:44–60, 1985.

[Tod95] Neil P. McAngus Todd. The auditory "primal sketch": A multiscale model of rythmic grouping. *Journal of New Music Research*, 23:25–70, 1995.