

**Évaluation de l'utilisabilité de *COSAL / SP2*
depuis les environnements techniques
des équipes scientifiques de l'Ircam**

*Auteur : Pierre-Alexandre PAGES
Équipe Services en Ligne – Ircam
Date : 02/06/2004*

Sommaire

Introduction.....	3
1 – État actuel de la SP2 et de COSAL.....	4
1.1 – Fonctionnalités présentes du système.....	4
1.2 – Avantages et inconvénients.....	7
2 – Analyse des besoins des équipes scientifiques.....	11
2.1 – Équipes concernées.....	11
2.2 – Besoins de l'équipe Analyse / Synthèse.....	12
2.3 – Besoins de l'équipe Systèmes temps-réel.....	14
2.4 – Besoins de l'équipe Représentations Musicales.....	15
2.5 – Besoins des autres équipes.....	16
3 – Solutions proposées et estimations de faisabilité.....	17
3.1 – Solutions ne nécessitant pas ou peu d'adaptations.....	17
3.1.1 – Shared Collection Space.....	17
3.1.2 – Mise en place de la nouvelle nomenclature.....	18
3.1.3 – Ajout de nouveaux formats au convertisseur audio.....	19
3.1.4 – Implémentation de « l'upload massif ».....	19
3.1.5 – Ajout du pitch et autres informations dans le sample.....	20
3.2 – Solutions nécessitant des développements particuliers.....	22
3.2.1 – SPShell Tools	22
3.2.2 – mONSPII.....	23
3.2.3 – Orientation vers l'architecture du MetaData Provider.....	23
3.2.4 – Shared Workspace	24
3.3 – Estimation du travail nécessaire et des délais impartis.....	26
4 – Conclusions.....	28

Introduction

Ce document constitue le rapport de la première partie de mon travail au sein de l'équipe Services en Ligne. Il conclut les phases de prospection et d'étude des besoins préalables aux développements informatiques proposés en partie 3. Les résultats de cette étude sont présentés ainsi que l'état du système *SoundPaletteII*, tel qu'il se trouvait lors de la rédaction de ce document.

1 – État actuel de la *SP2* et de *COSAL*

1.1 – Fonctionnalités présentes du système

La seconde version de l'application *SoundPalette* (*SP2*) est actuellement en cours de finalisation. Outre une refonte importante de l'architecture globale du système, cette nouvelle version apporte de nombreuses fonctionnalités, tant du point de vue de l'utilisateur que de celui des capacités propres à l'application.

Collections et gestion des droits d'accès :

Les banques sonores sont organisées en collections dans la base de données. La *SP2* n'est en aucun cas limitée aux banques de sons actuellement présentes, et c'est entre autres ce point particulier que nous allons tenter de mettre en avant dans cette étude.

Pour le moment, on trouve dans le système l'intégralité de la collection « Studio On Line » (*SOL*), la collection « Sound Ideas » (*SI*), ainsi que les collections personnelles des différentes équipes scientifiques. Le contenu exact des deux premières collections a été décrit dans [1] et [2], respectivement. Le contenu des collections personnelles reste bien évidemment à la discrétion de leurs propriétaires.

Il est possible d'organiser les sons au sein des collections sous forme d'une arborescence de répertoires. Les opérations de gestion et de maintenance ne sont toutefois possibles que si l'utilisateur possède les droits administratifs appropriés. Un système de groupes d'utilisateurs et de droits d'accès permet de restreindre ou d'étendre la visibilité et le champs d'action de chaque personne utilisant l'application.

Capacités de recherche / Query Language :

La *SP2* permet d'effectuer des recherches à l'intérieur de collections particulières ou bien sur l'ensemble des collections pour lesquelles l'utilisateur possède des droits de lecture. Ces recherches sont basées sur différents critères que nous allons détailler et ont pour but de permettre à l'utilisateur d'établir une sélection de sons qui seront téléchargés par la suite, si toute-fois ils correspondent effectivement à ses besoins (concept du panier électronique).

La pré-sélection des échantillons avant leur téléchargement se fait essentiellement à partir de l'interface Web / HTML, via un moteur de recherche textuelle traditionnel, complété par des interfaces permettant la saisie des critères avancés, tels que la hauteur ou la dynamique. Il existe également un applet Java permettant la navigation dans les sons de la base (browsing).

Ces deux moyens d'accès reposent tous deux sur *COSAL* (*CUIDADO Online Service Applicative Layer*), l'API de contrôle de la *SoundPalette*. *COSAL* est une API écrite en Java, qui permet d'utiliser par programmation l'ensemble des fonctionnalités du système *SP2*.

Il est donc tout à fait possible et concevable d'appeler les méthodes de *COSAL* depuis d'autres applications ou environnements, le seul pré requis, dans le cas d'un client lourd (par opposition au « thin client »), étant la disponibilité du runtime Java sur la machine de l'utilisateur.

Le tableau suivant résume les capacités et les critères de recherche de la *SP2*. L'ensemble de ces fonctions est bien entendu accessible grâce à *COSAL*, mais aussi via l'interface Web / HTML.

<i>Type de recherche</i>	<i>Implémenté</i>
Recherche textuelle sur le titre des échantillons	Oui
Recherche sur la hauteur des échantillons (fondamentale, accords...)	Oui
Recherche sur la dynamique des échantillons	Oui
Recherche sur le contenu des descripteurs numériques	En cours
Recherche par similarité	En cours

types de recherches possibles avec la SoundPaletteII

Architecture matérielle :

Le système *SP2* repose actuellement sur un rack de 4 serveurs de production (PIII 1Ghz, moins de 3 ans) supportant

actuellement une charge quasi-nulle. Ces serveurs hébergent toujours la première version de la *SoundPalette*. Leur contenu n'est pas sauvegardé, ce qui est assez étrange pour des serveurs de production...

Ces machines sont connectées à un serveur de fichier *NetApp* (Network Appliance) d'une capacité de stockage potentielle d'un TeraByte, dont seuls 350Go sont installés. L'ensemble est connecté au réseau interne de l'Ircam, ainsi que vers l'extérieur (la *SoundPalette* est accessible depuis Internet sous forme d'un abonnement forum).

Bien que les serveurs *NetApp* soient d'une très grande qualité, leur coût d'entretien est prohibitif. Le contrat de maintenance garantit le déplacement *in situ* d'un technicien de la société, dès les premiers signes de faiblesse d'un des disques durs.

Ce service est bien trop cher pour les moyens restreints dont dispose l'équipe système. Ajouté à cela le fait qu'il existe de nos jours des disques IDE d'une capacité de 200Go, à bien moindre coût, nous nous retrouvons dans une situation assez délicate : mis à part la redondance, le contenu du *NetApp* n'est pas sauvegardé. A la moindre défaillance, les données de la *SP2* peuvent disparaître...

La politique présente de l'équipe système est d'attendre jusqu'à ce que ce moment arrive. La question du remplacement du *NetApp* par des disques IDE ne devra pas être posée trop hâtivement. Il est impératif de prendre en compte dans ce choix, l'ensemble des souhaits et demandes émis par les différentes équipes au cours de l'élaboration de ce document.

Il existe à l'Ircam une demande clairement définie et un réel besoin d'un espace de travail et de stockage partagé. C'est bien évidemment le rôle de l'équipe Services en Ligne, de par sa vocation, de prendre en charge cette demande, en proposant une architecture logique et physique adaptée à ces besoins.

Parties restant à implémenter :

Suite à quelques légers problèmes de validation des résultats lors de l'intégration du module d'extraction des descripteurs audio, la partie concernant la gestion des méta-données ainsi que leur stockage physique n'est pas encore achevée. Il s'agit des composants *MEDEE* (MetaData Server / gestion) et *SONA* (Sound Analysis Server

/ production de métadonnées) dans l'architecture distribuée du système.

Il est cependant prévu que cette partie devienne fonctionnelle avant la seconde moitié de l'année 2004. Il sera alors possible d'effectuer des recherches par similarité ou sur des plages de valeurs numériques de description (ex: recherche des sons dont le centroïde spectral est compris entre deux bornes données).

La partie relative à la classification automatique est en cours de finalisation (composant *ACLA* dans l'architecture) Cette partie était déjà présente dans la première version de la *SoundPalette*. Elle est réimplémentée cette fois avec de nouveaux algorithmes plus performants.

1.2 – Avantages et inconvénients

Les avantages du système *SP2* pour la gestion des échantillons par rapport à un simple « repository » sur disque réseau NFS (tel que celui utilisé par l'équipe Analyse / Synthèse) sont les suivants. Il est toutefois important de noter que certains de ces avantages seront moins prononcés lorsque le serveur *NetApp* sera remplacé par des disques IDE. Le choix d'un système de fichiers « capable » et performant devra alors être fait, avec soins, à ce moment.

- Gestion des groupes / Gestion des droits d'accès : pas d'avantage significatif, mais recommandé (plus souple que les comptes et groupes système).
- Organisation en arborescence de répertoires : pas d'avantage significatif, mais les capacités de recherche de la *SP2* minimisent la taille et la profondeur de l'arbre (ex: on peut se contenter de classer les sons par instruments sans avoir 5 ou 6 niveaux de répertoires).
- Gestion d'un nombre important de fichiers : avantage évident pour la *SP2*, étant donné les problèmes de performance que rencontrent les systèmes de fichiers traditionnels lorsqu'ils sont confrontés à des quantités trop importantes de fichiers dans un répertoire. En pratique l'utilisateur est souvent contraint avec les disques classiques de devoir créer plusieurs arborescences afin d'avoir

différentes représentations des mêmes données (les arborescences supplémentaires ne contenant que des liens symboliques vers les fichiers originaux). Ce genre de pratique est coûteux en temps et en ressources, et est totalement inutile lorsque l'on dispose d'une base de données dotée de capacités de recherche avancées.

- Recherche parmi les échantillons : la recherche se limite uniquement au nom des échantillons dans le cas du disque NFS, ce qui entraîne bien souvent l'emploi de noms exhaustifs, rendant la manipulation des fichiers assez lourde. Les possibilités de recherche de la *SP2* sont bien plus étendues, et l'utilisation des descripteurs numériques permettra de se passer des noms à rallonge.
- Téléchargement des fichiers : pas d'avantage significatif. Le téléchargement se fait à vitesse égale sur un intranet.
- Navigation (browsing) dans les échantillons : avantage ici pour le disque NFS, qui est monté comme un système de fichiers traditionnel, et qui peut donc être parcouru via n'importe quel shell. Le browsing dans la *SP2* est un peu plus laborieux car il requière l'utilisation de l'applet Java.
- Répartition de charge, disponibilité, fiabilité : l'architecture distribuée de la *SP2* permet une répartition de charge totalement impossible avec un disque NFS unique (dans la *SP2* les serveurs de métadonnées et de données sont des entités distinctes). La disponibilité est très grande (serveurs de production et uptime conséquent).
- Sécurité et sûreté des données : en l'absence de politique de backup, pas d'avantage significatif. Ce point est cependant à déplorer.

Ce rapport ne pourrait être objectif si il n'abordait pas aussi les inconvénients du système. La réputation de la *SoundPalette* n'est pas très bonne, en témoigne son passé et les nombreux retours que j'ai pu avoir lors de mon étude. Ce n'est que conscients de ces problèmes que nous pourrions améliorer le système. La *SP2* se veut plus performante et fonctionnelle, mais un certain travail « d'éducation » sera nécessaire, en plus de celui d'intégration, afin de redorer son image et d'inciter le personnel de l'Ircam à l'utiliser.

Les points suivants constituent le résumé de l'ensemble des critiques et remarques que j'ai pu répertorier au cours de cette étude. Certaines concernent juste des détails techniques, d'autres plus des problèmes de fond. Elles sont présentées ici en vrac, sans classement d'importance ni de gravité.

- Qualité médiocre d'une partie des sons de la collection SOL : certains sons comptent plus de 20 secondes de silence, et le rapport signal / bruit est très faible pour les sons joués avec une petite dynamique (les sons ne sont pas normalisés), comme c'est le cas notamment pour les sons de harpe. Le principal problème étant le découpage temporel des échantillons. Un travail de nettoyage semble avoir été tenté mais n'a pas donné de suite. Cependant, l'utilisateur ne saurait se contenter d'une mauvaise impression concernant cette collection particulière pour juger l'ensemble du système... Ces faits sont connus depuis longtemps et ils entâchent la réputation de la *SP2*. Lorsque chaque équipe disposera de sa propre palette, ce problème sera vite oublié. Il est important aussi de noter que ces critiques ne sont valables que pour les utilisateurs ayant un usage « pratique » des échantillons, comme c'est le cas pour les équipes de production, les sons étant tout à fait corrects d'un point de vue purement scientifique.
- Le browsing via l'applet Java est trop lent : frustrant pour l'utilisateur car il n'y a pas de retour. L'ajout d'une barre de progression ou au moins d'une petite animation visuelle permettrait de le faire patienter plus sereinement, sans forcément avoir à modifier plus en profondeur le code.
- Recherche via la page Web : pas de retour non plus lorsque la requête s'exécute, perturbant pour l'utilisateur qui a tendance à relancer la recherche. L'ajout d'un message temporaire serait envisageable, de manière, là aussi, à rendre l'attente plus plaisante.
- L'appui sur la touche « entrée » ne lance pas la recherche (il ne se passe rien), il faut cliquer explicitement sur le bouton. De plus, les recherches renvoient parfois des messages d'erreur.

- Upload en masse des échantillons : impossible (en apparence) pour le moment, or cette fonctionnalité est vraiment nécessaire pour pouvoir disposer d'un réel espace de stockage partagé. L'upload se fait actuellement sample par sample, ce qui est évidemment bien fastidieux.
- Problème des noms de fichiers : les noms des fichiers sont trop longs ce qui rend leur manipulation difficile. Une nouvelle nomenclature, plus compacte, est cependant en cours de définition.
- Problèmes de performance générale : le système est assez lent dans son ensemble, Java n'étant pas connu pour sa vitesse d'exécution foudroyante... Ces problèmes risquent même d'empirer si des études plus poussées confirment la lenteur d'exécution de la bibliothèque d'extraction que laissait deviner nos tests préalables.

2 – Analyse des besoins des équipes scientifiques

2.1 – Équipes concernées

L'ensemble du personnel scientifique de l'Ircam peut utiliser la SoundPaletteII à condition de disposer d'un compte utilisateur. Cependant, nous avons choisi de proposer des services d'intégration personnalisés aux équipes ayant des besoins d'utilisation particuliers du système.

Bien qu'ils ne forment pas à proprement parler une équipe scientifique, les assistants musicaux n'ont pas été oubliés, et leurs besoins sont répertoriés dans la partie « autres équipes ». Il est important de rappeler que les équipes de production apportent de précieux exemples de cas d'utilisation « concrets », et que leurs commentaires sont riches en enseignements sur les cotés pratiques des systèmes que nous développons.

Afin de saisir les attentes de chacun, il est important de bien comprendre les différents types d'utilisation de la SP2. On distingue principalement deux cas : l'utilisation en production (utilisateur final et assistants musicaux) et l'utilisation en expérimentation (équipes scientifiques), bien que ces deux cas ne soient pas opposés, et qu'en pratique les besoins techniques concrets se recoupent.

Les principales attentes des équipes de production et des utilisateurs en abonnement forum concernent la disponibilité et la qualité des échantillons. L'utilisation la plus fréquente étant celle de la banque de sons pour sampleur, permettant de faire des rendus. Les capacités de recherche, de browsing et de téléchargement sont ici importantes.

Les équipes scientifiques souhaiteraient plus une orientation vers un espace de stockage et de travail partagé, permettant l'échange et le travail en commun sur de nombreux fichiers de description et de métadonnées. Il faut noter que de plus en plus, les assistants musicaux travaillent eux-aussi directement avec les fichiers de description (ex: SDIF).

2.2 – Besoins de l'équipe Analyse / Synthèse

Environnement technique :

De par son implication dans le développement de la bibliothèque d'extraction des descripteurs audio, dans le cadre du projet *CUIDADO*, l'équipe Analyse / Synthèse est la première concernée par les travaux d'intégration que nous proposons.

Les environnements techniques utilisés au sein d'AS sont relativement variés : Matlab principalement, sur les trois plateformes (Linux, Windows et MacOSX) mais aussi des applications C/C++, essentiellement sous Linux mais pas exclusivement.

La politique de gestion des échantillons sonores dans l'équipe est un point assez problématique. Lors de cette étude, l'équipe était justement en train de rassembler toutes ses banques de samples sur un repository NFS partagé. Comme nous l'avons exprimé précédemment, c'est typiquement ce que nous essayons d'éviter.

Point assez représentatif de la situation, l'équipe possède sur ce disque NFS une copie locale de la collection *SOL*. Cela montre bien à quel point il est impératif de rendre l'utilisation de la *SP2* plus pratique et intuitive, sans quoi le système sombrera inévitablement dans l'oubli, faute d'une utilisation effective.

Analyse des besoins :

- Intégration la plus transparente possible depuis Matlab (Linux, Windows, MacOSX) [ASN01].
- Intégration éventuelle depuis des programmes C / C++. Nécessite une définition plus précise.
- Recherche sur les descripteurs numériques [ASN02].
- Disposer d'un espace de partage pour l'équipe, dans lequel l'upload et le download seraient faciles. Cet espace doit être une palette particulière, uniquement pour l'équipe [ASN03].
- Avoir un certain contrôle sur le mécanisme d'extraction des descripteurs. Pouvoir lancer des extractions particulières à la main

[ASN04], et pouvoir ajouter des données de description dans le serveur de métadonnées [ASN05]. Le but étant de faciliter l'élaboration et les tests des nouveaux descripteurs développés.

- Pouvoir remplir les champs textuels des métadonnées à la main, afin notamment d'y adjoindre des commentaires [ASN06].
- Pouvoir utiliser les résultats des analyses des autres membres de l'équipe, si celles-ci sont bien faites [ASN07]. Si une analyse était difficile, éviter aux autres de devoir la refaire, et pouvoir ajouter quelques indications en commentaire (il faut donc des associations typées entre les fichiers) [ASN08].
- Pouvoir télécharger les échantillons directement au format MP3 serait un plus. (note: les problèmes évidents de licence concernant ce format particulier d'encodage nous incitent à le mettre de côté, pour favoriser si possible des standards libres) [ASN09].
- Dans la logique de restructuration des espaces de stockage d'échantillons de l'équipe, pouvoir uploader les nouveaux samples « en masse » serait impératif [ASN10].
- L'équipe possède certains échantillons non libres de droits dont il faudrait pouvoir restreindre la visibilité et l'accès [ASN11].

[ASN03] et [ASN11] sont réalisables immédiatement sans modifications par [3.1.1] (voir plus loin). [ASN01] pourra être implémenté par [3.2.2] et aussi par [3.2.1]. [ASN02] est en cours d'intégration et sera disponible avant la seconde moitié de l'année 2004. [ASN04], [ASN05] et [ASN06] pourraient être implémentés par [3.2.3].

[ASN07] et [ASN08] ne pourront être implémentés avec l'architecture actuelle du système. Cependant [3.2.4] pose les bases de cette problématique commune aux différentes équipes de l'Ircam.

[ASN09] pose des problèmes éthiques et juridiques évidents. L'utilisation du format « OGG VORBIS » serait préférable. Ce point est couvert par [3.1.3].

[ASN10] semble avoir été réclamé depuis longtemps. Il est abordé dans la partie [3.1.4].

2.3 – Besoins de l'équipe Systèmes temps-réel

Environnement technique :

L'équipe temps-réel n'avait au début aucun besoin de la *SP2*. Ce n'est qu'après avoir reçu ses objectifs pour le projet *SemanticHiFI* que ses attentes se sont dessinées plus clairement.

Dans le cadre de son futur travail sur l'alignement Text to Speech et le suivi de partition, l'équipe a exprimé un certain nombre de souhaits et d'orientations. La plupart de ces propositions ne sont pas réalisables dans un futur proche, mais toutes seront d'actualité pour *SemanticHiFI*, et entrent dans la logique de l'espace de stockage et de travail partagé.

Les environnements techniques utilisés par l'équipe STR sont principalement *jMax* et *MAX/MSP* sur MacOSX. Une utilisation de la *SP2* directement depuis *MAX* n'est pas souhaitée, les contraintes temps-réel ne pouvant clairement pas être respectées avec un tel système distribué.

Il est important aussi de noter qu'en l'absence de backup, l'équipe refusera catégoriquement de nous confier ses données. Ses membres ont déjà été témoins de plusieurs crashes de disques durs et ne veulent pas que cette situation se reproduise.

Analyse des besoins :

- Disposer d'un groupe particulier pour le suivi de partition [RTN01].
- Ce groupe doit être accessible depuis l'extérieur de l'Ircam, notamment pour l'organisation de la « compétition d'alignement » [RTN02].
- Le travail direct sur les fichiers MPEG7 est peu probable car laborieux. Ces fichiers sont cependant indispensables pour l'indexation et la recherche (production). Pour pouvoir travailler et expérimenter, le système doit être capable de stocker d'autres types de fichiers de métadonnées (SDIF, PSOLA, HMM, MusicXML, etc.) [RTN03] et plusieurs versions de ces fichiers [RTN03b].
- Certains échantillons sont juste des archives, ils devraient être stockés dans le système, mais ne pas être indexés ni apparaître dans les résultats de recherche [RTN04]. Ces sons ne posséderaient comme métadonnées que des informations textuelles globales, qu'il faudrait donc pouvoir renseigner [RTN05].
- La notion d'ensemble de fichiers de travail est très importante. Il

faut différencier plusieurs types d'associations entre les fichiers : Un fichier audio et un fichier de description MPEG7 unique associé (utilisation type production) [RTN06]. Un fichier audio et un ensemble de fichiers de travail associés (SDIF, analyses, partitions...) [RTN07], et éventuellement des commentaires liés à ces fichiers associés [RTN08].

- Il faudrait aussi des associations dans l'autre sens : Une partition et un ensemble d'interprétations associées, ou bien une voix (sous forme de texte) et un ensemble de données d'alignement associées [RTN09]. Dans ce cas la, le fichier central n'est pas forcément un fichier audio.

[RTN01] et [RNT02] sont déjà possibles avec le système actuel [3.1.1].

Tous les autres besoins entrent dans le cadre de *SemanticHiFI* et sont abordés dans [3.2.4].

2.4 – Besoins de l'équipe Représentations Musicales

Environnement technique :

L'équipe RepMus travaille principalement sur le développement du logiciel *OpenMusic*, ainsi que sur le vaste ensemble des problématiques *d'orchestration*

Il existe certes des besoins d'échantillons et de descripteurs audio, comme c'est le cas pour la superposition instrumentale qui consiste à utiliser le son de l'orchestre comme un synthétiseur, de manière à imiter par superposition de samples le spectre d'un son, avec ou sans spécification de l'instrumentarium.

Des études sur la maximisation et la minimisation de la fusion spectrale sont aussi prévues. Comme pour l'imitation de spectre, ces concepts reposeraient sur l'utilisation de descripteurs spectraux particuliers (peaks spectraux ou enveloppe spectrale).

Cependant, toutes ces approches et techniques ne sont pas encore tout à fait d'actualité. Il est trop tôt pour pouvoir dresser la liste des besoins exacts.

Analyse des besoins :

- Pas de besoin immédiat de la *SoundPaletteII*.
- Besoins évidents dans le futur pour les travaux d'orchestration et de fusion spectrale.

2.5 – Besoins des autres équipes

Les souhaits émis par le reste du personnel de l'Ircam, et par les assistants musicaux essentiellement, relèvent plus du domaine des améliorations pratiques et fonctionnelles du système, tant sur le plan ergonomique que sur celui du contenu.

La principale utilisation de la base d'échantillons en production est la réalisation de banques pour sampleur, pour pouvoir faire des rendus dans un séquenceur (ex: ProTools, Cubase...) Cela implique que les sons soient correctement coupés (attaque et relâchement) et qu'ils possèdent une dynamique cohérente, afin de pouvoir mapper facilement les différents modes de jeu.

Pour automatiser tant que possible la réalisation de ces banques, il faut que la nomenclature des noms de fichiers soit suffisamment compacte (la nomenclature actuelle produisant des noms trop longs pour certains logiciels), et que leur hauteur soit facilement accessible (si possible directement dans un chunk du fichier audio) de manière à pouvoir mapper automatiquement un ensemble de sons sur toute l'étendue du clavier. Voir [3.1.2] et [3.1.5].

3 – Solutions proposées et estimations de faisabilité

Ayant répertorié l'ensemble des besoins du personnel de l'Ircam, nous pouvons maintenant envisager un certain nombre de développements et réalisations autour de la *SP2*, afin de rendre son

utilisation la plus simple et la plus attrayante possible.

Bien évidemment, il est peu probable que l'ensemble de ces propositions soit implémenté avant la fin du mois de Septembre 2004. Seules les idées les plus pertinentes seront menées à bien. Cependant, toutes, sans exception, reflètent un ou plusieurs besoins réels des équipes, et devront donc être prises en compte à un moment ou à un autre.

Nous avons distingué deux types de solutions : celles réalisables en relativement peu de temps et ne nécessitant pas ou peu de développements informatiques, comme par exemple l'ajout de « features » au système existant, et celles nécessitant des spécifications plus poussées et/ou des développements informatiques plus importants.

3.1 – Solutions ne nécessitant pas ou peu d'adaptations

3.1.1 – Shared Collection Space

La première étape de notre travail de « démocratisation » de la *SP2* pourrait être de proposer tout d'abord, et dans un futur proche, un espace de stockage partagé pour les échantillons sonores sur nos serveurs.

En utilisant uniquement les fonctionnalités actuelles de la *SP2*, il serait possible de préparer le terrain pour les orientations à venir, et notamment pour l'espace de travail partagé que nous pourrions proposer avec l'avancement de *SemanticHiFI* (voir [3.2.4]).

En transférant tout d'abord les banques de sons des équipes sur nos serveurs, une bonne partie du travail serait faite. Les utilisateurs pourraient commencer à se familiariser avec l'administration de leur espace de stockage partagé :

- Gérer leurs comptes utilisateurs et leurs permissions.
- Créer des collections et des répertoires et organiser les échantillons.
- Lancer des recherches via l'interface Web/HTML ou l'applet Java.

Viendraient alors s'ajouter les nouvelles interfaces pour la recherche et le browsing : scripts shell et interface Matlab (voir plus loin), qui permettront de s'affranchir de l'interface Web et rendront le système plus accessible.

L'architecture du système pourrait alors évoluer peu à peu vers celle de *SemanticHiFI*, en ajoutant progressivement le support d'autres types de fichiers et les liens typés entre les objets.

3.1.2 – Mise en place de la nouvelle nomenclature

Comme l'ont fait remarquer assez récemment les assistants musicaux, le système de nommage actuel des échantillons lors de leur téléchargement pose certains problèmes : les noms générés sont trop longs pour la plupart des sampleurs et logiciels du marché, et la longueur de ces noms rend la manipulation des samples laborieuse.

Une nouvelle nomenclature plus compacte a été proposée [3]. Remplacer la nomenclature actuelle par la nouvelle ne poserait aucun problème technique, il suffirait juste de changer les règles de nommage lors de la conversion et du téléchargement des samples.

Cependant, avant d'entreprendre ce travail, il faudrait être sûr que la nouvelle proposition est adaptée à tous les cas possibles, y compris les modes de jeu « exotiques » que l'on rencontre parfois dans la collection *SOL*. Il ne faut pas oublier que la nomenclature actuelle avait fait l'objet à l'époque d'une étude et d'une spécification claire...

3.1.3 – Ajout de nouveaux formats au convertisseur audio

La logique de conversion des échantillons dans la *SP2* est la suivante : les sons sont stockés dans la base sous leur format d'origine, sans modification. C'est lors du téléchargement qu'une éventuelle conversion est effectuée, si l'utilisateur spécifie un format de fichier différent.

Le composant de conversion est déjà implémenté et mis en place, il ne reste plus qu'à ajouter des nouveaux formats. Suite à la

demande des équipes concernant l'ajout d'un format de fichier compressé, pour notamment réduire le temps de téléchargement, nous avons envisagé d'adopter le format de fichier standard OGG Vorbis (.ogg).

Le format OGG Vorbis possède de nombreux atouts par rapport aux autres formats de compression communément utilisés comme le MP3. Outre une qualité subjective du signal supérieure, ce format offre l'avantage d'être entièrement libre de droits et ouvert (contrairement au MP3 qui nécessite l'achat d'une licence auprès de Thomson / Fraunhofer). A titre d'exemple, la licence MP3 s'élève à US\$ 60 000 pour le décodeur uniquement, comme indiqué sur le site <http://www.mp3licensing.com/>.

La liberté d'OGG assure la pérennité du format mais aussi la disponibilité de codecs et de bibliothèques gratuits, ce qui réduira non seulement les coûts, mais aussi le temps d'implémentation.

3.1.4 – Implémentation de « l'upload massif »

La possibilité de pouvoir ajouter des sons à la base de données « en masse » semble être une fonctionnalité assez demandée. On peut pourtant se questionner sur la légitimité de cette demande :

Dès la première version de la *SoundPalette* il était possible d'uploader un ensemble de fichiers et même l'arborescence des répertoires les contenant, sous la forme d'un unique fichier .zip. Le serveur décompressant automatiquement ce fichier et recréant dans la base la structure des répertoires.

S'agit-il alors d'un manque de communication par rapport à cette fonctionnalité ? Ou bien la création de l'archive .zip est-elle peut-être une tâche trop complexe pour les utilisateurs ? Le temps de compression est-il un facteur limitant ?

Nous avons choisi de ne pas implémenter d'autres

méthodes d'upload à la chaîne au niveau de nos serveurs, la technique du .zip fonctionnant parfaitement. Il est par contre envisageable de mettre en place d'autres procédés, mais cette-fois au niveau du client (dans [3.2.1] et/ou [3.2.2]) comme, par exemple, la technique dite « de batching ».

3.1.5 – Ajout du pitch et autres informations dans le sample

La plupart des formats de fichiers audio actuels permettent, de par leur structure, de stocker un certain nombre d'informations complémentaires en rapport avec l'échantillon (métadonnées). Ces informations vont de l'indispensable (pour pouvoir interpréter le son : SR, bitdepth, encodage), au totalement superflu (données spécifiques à des applications propriétaires).

L'architecture de la *SP2*, et par extension, de *SemanticHiFI*, propose une vision plus dichotomique de l'utilisation des métadonnées : données et données associées sont clairement séparées (dans la *SP2* les métadonnées de description du signal sont stockées dans un fichier MPEG7).

Il serait intéressant de pouvoir adjoindre automatiquement à l'échantillon les métadonnées les plus pertinentes lors du téléchargement (la liste est discutée un peu plus loin). Actuellement, les samples sont stockés sur le serveur de fichier sous leur format d'origine, sans conversion. Les éventuels chunks de métadonnées qu'ils contiennent sont alors préservés. Si toute-fois l'utilisateur télécharge le sample sous un autre format, le module de conversion audio supprime les chunks présents dans tous les cas.

Trois possibilités sont envisageables : la première serait de garder le mécanisme actuel (préservé tout le fichier), mais en indiquant à l'utilisateur le format de fichier d'origine, de manière à ce qu'il puisse être sûr que les métadonnées qu'il contient soient sauvegardées si il télécharge le sample dans ce même format de fichier.

La seconde consisterait à préserver les métadonnées d'origine du fichier dans tous les cas (c.a.d. lors de la conversion), mais cela implique certains problèmes de cohérence si les données

sont intimement liées au format de fichier ou à l'encodage (le time stamping par exemple devient erroné lorsque l'on change la fréquence d'échantillonnage).

La dernière possibilité, plus en accord avec l'esprit de *SemanticHiFI*, serait d'adjoindre systématiquement lors de la récupération des sons, un ensemble de métadonnées extrait du fichier MPEG7 de description associé. Cette solution serait de loin la moins performante, car elle implique de parser le fichier MPEG7 pour récupérer les informations requises, mais elle serait aussi la plus propre, car elle permettrait d'assurer une certaine cohérence en centralisant toutes les métadonnées.

Techniquement, ces informations seraient stockés dans le cas du format RIFF WAVE (Microsoft .wav) dans un chunk 'smp1' (sampleur) ou 'inst' (instrument), ainsi que dans un ensemble de chunks textuels. Pour le format Apple/SGI AIFF, il serait possible d'utiliser les chunks 'INST' (instrument) et les chunks textuels 'COMT' (commentaire), 'NAME' (nom du sample), 'AUTH' (auteur) et '(C)' (copyright).

Le type exact des métadonnées adjointes dans le fichier est à définir, mais il faut que l'ensemble des informations soit sauvegardable dans tous les formats de fichiers proposés (certains formats ne permettent pas de tout stocker).

Les données suivantes ont fait l'objet de demandes de la part des équipes, car elles rendrait l'utilisation en production plus simple (cf: création de banques pour sampleur) Elles sont toutes compatibles avec les formats de fichiers que nous proposons :

- Nom du sample
- Auteur du sample
- Copyright éventuel
- Pitch du sample (MIDI note + tuning)
- Autres informations textuelles

3.2 – Solutions nécessitant des développements particuliers

3.2.1 – SPShell Tools

Déporter *COSAL* sur le client est une approche nouvelle. Jusque là, l'API était exclusivement utilisée coté serveur. Cependant, cette conception fait partie des spécifications de la *SP2*, et constitue donc un cas d'utilisation valable, mais pas encore éprouvé. La solution que nous proposons permettra, en plus de faciliter l'utilisation du système, de valider cette méthode de conception.

Le coeur de l'application SPShell serait constitué d'un exécutable Java fonctionnant sur le modèle du shell (ligne de commande) et permettant d'accéder aux méthodes de *COSAL*. Plus exactement, le principe de fonctionnement se rapprocherait des shells utilisés sur les serveurs XDCC que l'on trouve sur les réseaux IRC (Internet Relay Chat).

SPShell permettra de naviguer dans la base comme dans un filesystem, de simplifier la connexion grâce à un fichier de configuration, d'organiser, de chercher et de rapatrier les échantillons d'une manière simple et efficace.

Grossièrement, l'application reproduirait les méthodes offertes par l'interface Web, sur la machine client, tout en simplifiant certains aspects. SPShell pourra être exécuté depuis n'importe quel système possédant un runtime Java. Un script bash permettra de le lancer depuis la ligne de commande Linux et un script batch (.bat) depuis Windows.

Les spécifications exactes de l'application SPShell seront définies dans un autre document, si toutefois cette approche est adoptée.

3.2.2 – mONSPII

L'étape suivante serait de pouvoir utiliser le SPShell directement depuis Matlab : c'est la vocation de mONSPII. Une

couche ultra fine d'encapsulation sera développée autour du shell, permettant dans un premier temps son utilisation en mode texte dans Matlab.

Dans un second temps, une interface graphique Matlab pourra être développée autour de l'existant, afin de proposer une manière plus intuitive de lancer les recherches (notamment les recherches par similarité).

L'appel natif de méthodes Java depuis Matlab ne pose aucun problème technique, et semble même être très bien documenté.

3.2.3 – Orientation vers l'architecture du MetaData Provider

La partie concernant l'extraction, l'édition, la gestion et le stockage des métadonnées est en cours d'implémentation (composants MEDEE et SONA de l'architecture distribuée de la *SP2*). Bien que les spécifications soient arrêtées depuis quelques temps déjà, nous allons être amenés à les faire évoluer pour nous orienter peu à peu vers le composant *MetaData Provider* de *SemanticHiFI*.

A ce titre, et profitant du fait que l'implémentation dans la *SP2* soit en cours, en parallèle avec la rédaction de ce document, une étude séparée pourrait être menée afin de déterminer ce qu'il sera gardé de l'existant et ce qu'il faudra ajouter.

Cette étude permettrait d'établir et de spécifier un format de plugin pour les modules d'extraction, ainsi qu'un ensemble de méthodes de manipulation des métadonnées et du processus d'extraction (comme par ex. le renseignement manuel des métadonnées textuelles).

Elle permettrait aussi de tirer partie des enseignements et des conseils recueillis lors de l'écriture de ce présent rapport, afin d'offrir aux utilisateurs de notre futur système la souplesse qu'il souhaiteraient avoir vis à vis de la gestion des métadonnées.

3.2.4 – Shared Workspace

Bien que clairement utopique dans le cadre de la *SoundPaletteII*, cette dernière solution est plus une étude précise visant à collecter un ensemble de cas d'utilisation concrets par les équipes de l'Ircam (use cases) d'un système de travail, de stockage et de partage centralisé.

La vision du *Shared Workspace* propose une approche nouvelle du travail sur les projets au sein de l'Ircam. Cet espace de travail commun permettra le stockage sécurisé de l'ensemble des fichiers, données et annotations relatives aux projets en cours. Cette centralisation, associée à des moyens d'interaction simples, favorisera nous l'espérons la coopération entre les différentes équipes.

Si cette étude est menée à bien, nous pourrons disposer d'une quantité importante d'informations pratiques dont nous serons sûrs qu'elles sont bien en adéquation avec les attentes « réelles » de chacun. Ces informations pourraient s'avérer très précieuses pour l'établissement des spécifications du projet *SemanticHiFi*.

Les investigations préalables montrent qu'il existe un besoin clair de pouvoir partager non seulement les échantillons sonores et musicaux, mais aussi un nombre relativement important d'autres types de fichiers, principalement de description numérique ou textuelle. Tous ces fichiers doivent être liés entre eux, de telle sorte à ce que l'accès à l'un puisse donner l'accès aux autres.

Parmi les types de fichiers possibles, voici une liste non exhaustive des principaux formats utilisés :

- Fichiers audio courts
- Fichiers musicaux longs
- Fichiers MIDI
- MusicXML
- Partition dans d'autres formats (.PDF ?)
- Différents fichiers SDIF (différentes représentations)
- Partiels (Additive)
- Harmoniques (Additive)

- HMM
- PSOLA
- Commentaires textuels et annotations
- Données d'alignement text to speech

3.3 – Estimation du travail nécessaire et des délais impartis

Cette partie du document tente de fournir une estimation réaliste du temps nécessaire pour réaliser les différentes solutions proposées. Pour chacune d'entre elles, les facteurs limitants et les prérequis éventuels sont énumérés, et les durées des phases de spécification et d'implémentation sont évalués.

★ [3.1.1] – Shared Collection Space

Prérequis : avant de pouvoir commencer à stocker les banques sonores des équipes, il nous faudra résoudre la question du stockage physique et de la politique de sauvegarde des données, en vue du remplacement du serveur *NetApp*.

Étapes : création des comptes et des collections. Choix de la politique pour les droits et permissions. Upload d'une petite banque de sons afin de valider les cas d'utilisation simples.

Durée : 4 - 5 jours.

★ [3.1.2] – Nouvelle Nomenclature

Prérequis : nomenclature validée pour l'ensemble des sons et modes de jeux, et acceptée par l'ensemble du personnel concerné.

Étapes : mise en place des nouvelles règles. Validation.

Durée : 5(+) jours.

★ [3.1.3] – Nouveaux formats : OGG Vorbis

Prérequis : aucun. Des bibliothèques open source éprouvées sont déjà disponibles.

Étapes : intégration de la bibliothèque. Tests.

Durée : 3 – 5 jours.

★ [3.1.4] – Upload massif : pas de développement prévu

★ [3.1.5] – Chunks d'information

Prérequis : liste exacte des métadonnées pertinentes. Choix de l'approche utilisée (probablement la troisième).

Étapes : implémentation d'un nouveau ServiceProvider Java. Liaison avec MEDEE, récupération des métadonnées. Application au format AIFF, puis RIFF WAVE.

Durée : 1 mois minimum.

★ [3.2.1] – SPShell

Prérequis : spécification exacte de la liste des commandes.

Étapes : gestion des informations de connexion. Fonctions de navigation de base. Fonctions de gestion des données. Fonctions de recherche.

Durée : 2 – 3 mois.

★ [3.2.2] – mONSPII

Prérequis : [3.2.1].

Étapes : couche d'encapsulation fine dans Matlab. Réalisation éventuelle d'une interface graphique.

Durée : 2 – 4 semaines.

★ [3.2.3] – Étude MetaData Provider

Prérequis : aucun.

Étapes : étude des besoins de *SemanticHiFI*. Conception du modèle de plugins. Spécification de l'interface de contrôle du provider. Rédaction du rapport.

Durée : 3 – 4 semaines.

★ [3.2.4] – Étude Shared Workspace

Prérequis : aucun.

Étapes : rassemblement et formalisation des use cases. Étude de faisabilité. Rédaction du rapport.

Durée : 2 mois.

4 – Conclusions

Nous avons dressé dans ce document un état des lieux de la *SoundPaletteII* : ses avantages et ses inconvénients, les fonctionnalités manquantes et celles en cours de développement. Nous avons aussi proposé un certain nombre de solutions permettant de répondre au mieux aux attentes des différentes équipes de l'Ircam.

Plusieurs points intéressants ont été mis en avant : Tout d'abord, il est impératif que les gens fassent bien la distinction entre la collection *StudioOnLine* et le système *SP2*, l'un étant une partie du contenu, l'autre étant le contenant. Malheureusement, cette distinction ne semble pas évidente pour un certain nombre de personnes, qui continuent à critiquer la collection *SOL* sans prendre en compte les avancées du système en lui-même.

Il apparaît aussi que le principe de fonctionnement de la *SP2*, dans laquelle un fichier audio central possède un unique fichier de description *MPEG7* associé, est bien trop limité. La prochaine étape sera donc d'apporter une orientation plus « sémantique » au système, en permettant d'avoir un fichier et son ensemble de fichiers de description variés et de commentaires associés.

Références :

- [1] : « Listing du contenu actuel de la collection SOL » – Max Jacob 12/05/04.
- [2] : « Catalogue Sound Ideas » - Production / Forum Ircam.
- [3] : « Projet Orchestration et Sound Palette (SOL) » – e-mail de Yan Maresz et Manuel Poletti du 26/05/04.