

Camille-Eva Rakovec

Sous la direction de Mitsuko Aramaki et de Richard Kronland-Martinet
Laboratoire de Mécanique et d'Acoustique, 31 Chemin Joseph Aiguier, 13402 Marseille

Résumé

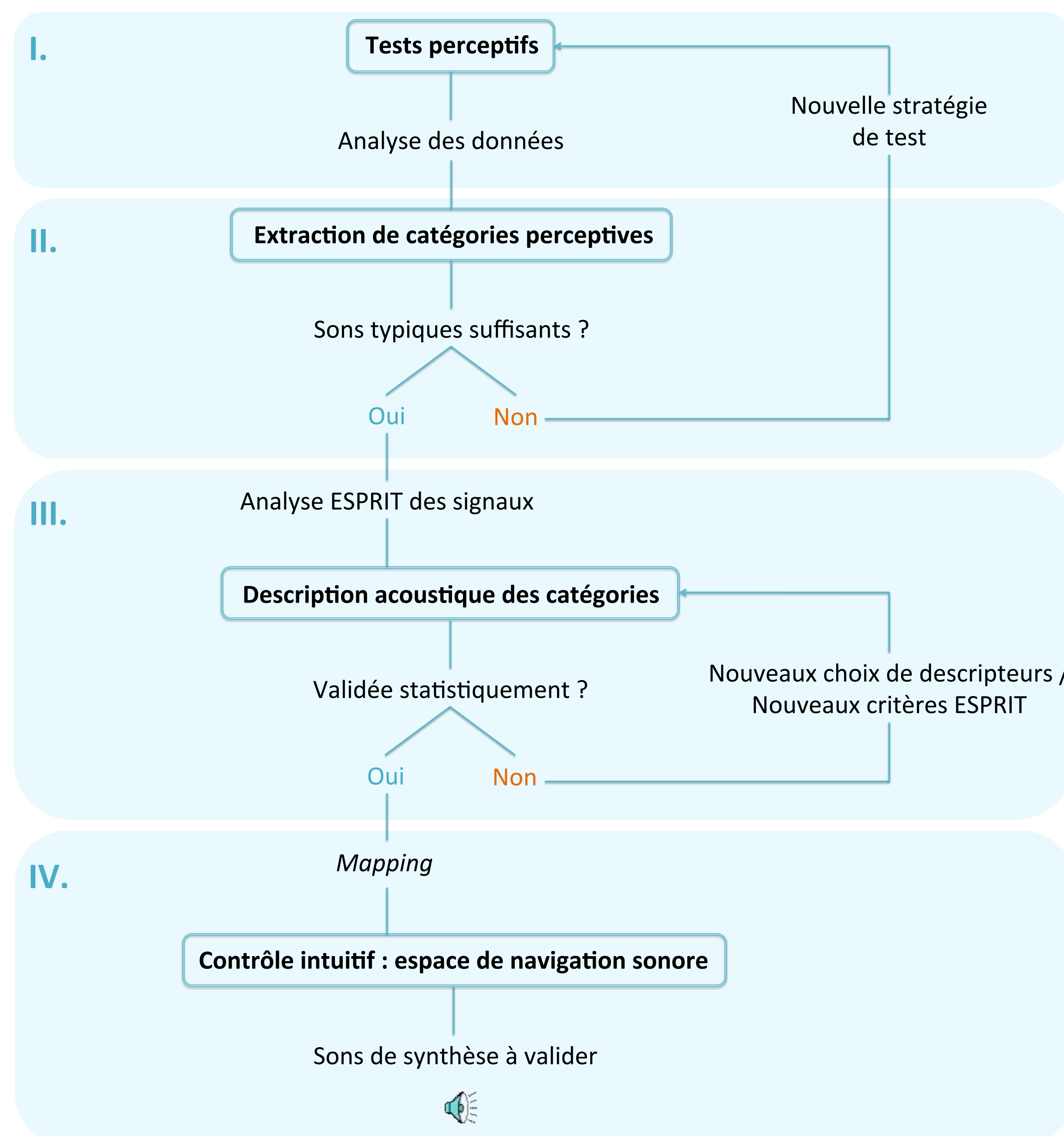
Dans le cadre de la réalité virtuelle, un des enjeux scientifiques actuel réside dans la prise en compte des actions de l'utilisateur dans le processus de synthèse sonore afin d'augmenter la sensation d'immersion et de réalisme. Dans ce contexte, le travail engagé a pour objectif d'aborder les problématiques spécifiques aux sons produits par les objets du quotidien.

Il s'agit tout d'abord de mieux comprendre la perception auditive des caractéristiques physiques de ces sources (forme, taille, présence de cavité...) et de définir des catégories perceptives. En s'appuyant sur des considérations physiques et sur la caractérisation acoustique des signaux produits, il s'agit ensuite de déterminer des "invariants", ou "descripteurs acoustiques", responsables des évocations induites.

En se basant sur ces résultats perceptifs et acoustiques, la finalité du travail consistera à mettre en œuvre des stratégies de contrôle perceptif et interactif, en vue de proposer des outils de synthèse sonore en temps réel pour la réalité virtuelle et augmentée (synthétiseur générique de sons d'environnement ou jeux vidéos par exemple).

Méthodologie générale

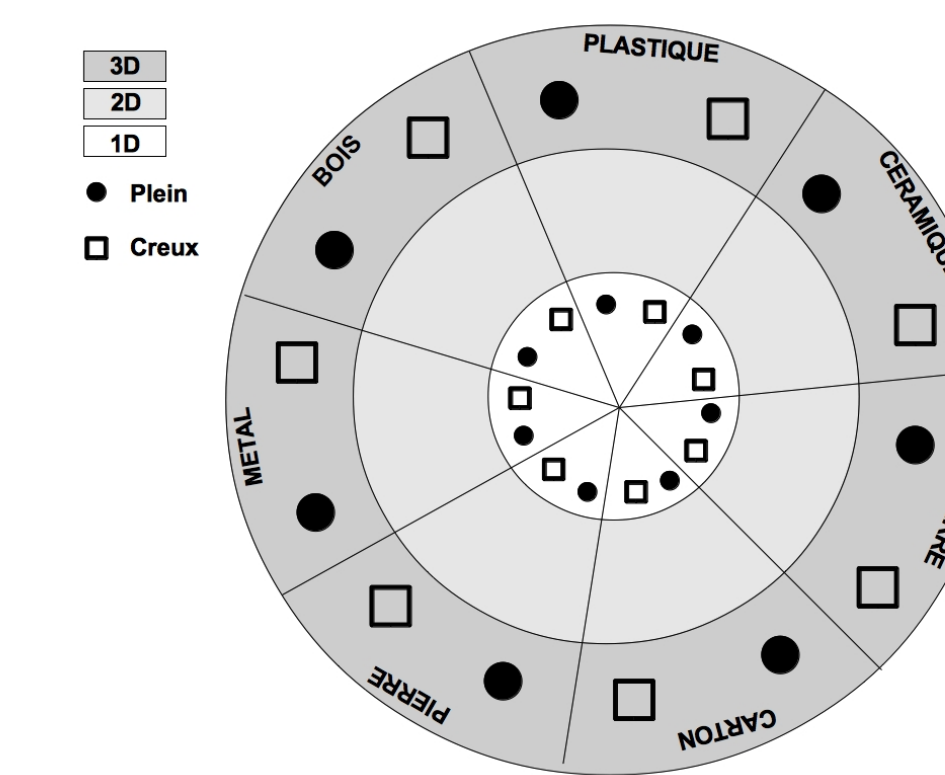
La première question à laquelle nous souhaitons répondre est : **sommes-nous capables de percevoir différentes catégories d'objets ?** Dans cette optique, nous testons perceptivement trois attributs physiques de la source-objet : **son matériau, sa forme et sa cavité (plein/creux)**. Une banque de sons d'impact sur des objets du quotidien a été constituée dans ce sens.



I. Mise en place de tests perceptifs

Banque de sons d'impact

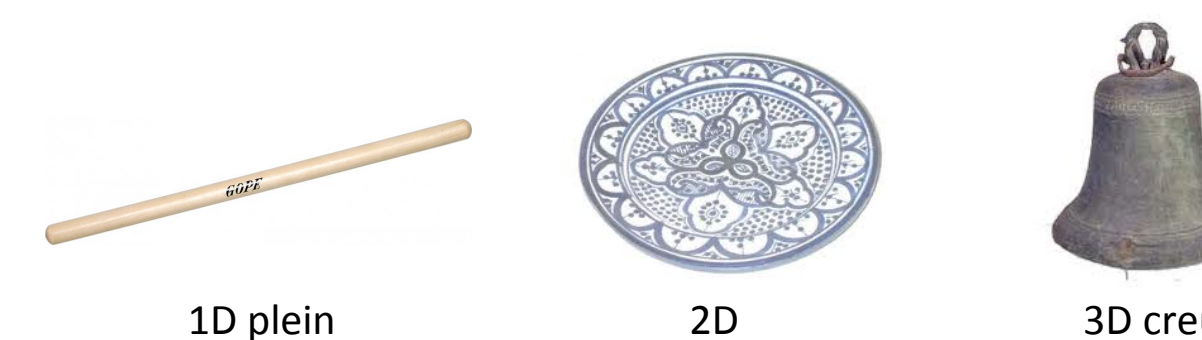
Un son d'impact correspondant à la réponse vibratoire d'un objet en oscillations libres, l'allure, le spectre ou encore l'énergie de sa **réponse impulsionnelle** peuvent nous renseigner sur les caractéristiques de l'objet. Des objets rentrant dans les catégories présentées dans la cible ont donc été collectés et **impactés en salle anéchoïque**. Précisons que la forme "1D" représente des objets longilignes (tube, barre, tiges...), "2D" des objets dans le plan (plaque, assiette, planche...) et "3D" des objets en volume (bloc, boîte, bol...).



Catégories de la banque de sons d'impact

Tests perceptifs

Des **tests perceptifs** sont ensuite élaborés sur ces mêmes catégories. Nous demandons à une vingtaine de sujets d'écouter plusieurs dizaines de sons d'impact et de choisir le **matériau perçu** parmi "métal", "verre", "céramique", "pierre", "plastique", "bois" et "carton", la **forme perçue** parmi "1D", "2D" et "3D" et l'**intérieur perçu** parmi "plein" et "creux". Les sujets ont également la possibilité de spécifier un nom d'objet.



Exemples d'objets impactés

II. Catégories perceptives

Analyse comportementale

Voir les **matrices de confusion** ci-contre. Concernant les matériaux, nous montrons que le métal est le mieux reconnu et la pierre le moins bien. En termes de catégories perçues, l'analyse de confusion nous permet d'établir **deux grands ensembles de matériaux** où la pierre et le bois seraient intermédiaires.

| | Métal | Verre | Céramique | Pierre | Plastique | Bois | Carton |
|-----------|-------|-------|-----------|--------|-----------|-------|--------|
| Métal | 63,55 | 15,22 | 7,27 | 10,03 | 2,42 | 1,38 | 0 |
| Verre | 24,47 | 34,12 | 15,29 | 13,88 | 6,35 | 5,18 | 0,71 |
| Céramique | 14,59 | 19,76 | 39,29 | 15,76 | 4,71 | 5,65 | 0,24 |
| Pierre | 4,20 | 15,76 | 32,77 | 24,16 | 9,66 | 12,61 | 0,21 |
| Plastique | 14,35 | 4,30 | 3,16 | 13,34 | 36,01 | 19,66 | 8,46 |
| Bois | 4,73 | 3,23 | 7,04 | 1,73 | 33,68 | 47,29 | 2,08 |
| Carton | 0,24 | 0 | 0,24 | 3,29 | 31,29 | 15,76 | 48,94 |

Concernant les formes, le 3D est le mieux reconnu, suivie du 2D et du 1D. De plus, les **confusions entre 1D/3D et 1D/2D sont fortes**. La proposition 1D ne semble donc pas pertinente.

| | 1D | 2D | 3D |
|----|-------|-------|-------|
| 1D | 37,10 | 28,73 | 34,16 |
| 2D | 17,04 | 46,38 | 36,58 |
| 3D | 20,44 | 25,86 | 53,70 |

Concernant la présence de cavité ou non, le plein et creux sont tous deux bien identifiés.

| | Plein | Creux |
|-------|-------|-------|
| Plein | 61,16 | 38,84 |
| Creux | 39,18 | 60,82 |

Analyse statistique

Des analyses statistiques effectuées sur les données issues des tests montrent par ailleurs que **l'interaction matériau/forme est assez forte**. En effet, les sujets ont tendance à effectuer des associations caractéristiques des objets de notre environnement de façon instinctive, et non pas à l'image de la banque de sons d'impact. La nature multi-tâche des tests pourrait en être la cause.

III. Description acoustique des catégories

Descripteurs acoustiques

Il s'agit de trouver des descripteurs acoustiques susceptibles de **décrire au mieux les attributs perceptifs** de matériaux, de formes et de cavité des objets impactés. Nous analysons pour cela les **sons typiques de chacune des catégories**, i.e ayant été **catégorisés à plus de 50%** par les sujets. La méthode ESPRIT, décrite ci-dessous, constitue un outil de choix dans la construction de nos descripteurs. A ce jour, nous privilégions :

- MATÉRIAUX

L'amortissement δ , dont la dépendance fréquentielle peut être modélisée par : $\delta = e^{\alpha_g + 2i\pi\alpha_r}$ où α_g désigne l'amortissement global et α_r désigne l'amortissement relatif, fonction de la fréquence.

La rugosité qui reflète la présence de plusieurs composantes spectrales dans une même bande critique.

Le centre de gravité spectral (CGS) qui

Méthode ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) :

Il s'agit d'un algorithme d'analyse-synthèse reposant sur un **modèle de signal**, tel que la réponse impulsionnelle d'un objet en oscillations libres peut-être considérée comme une **somme de sinusoides amorties**. Le signal observé s'exprime alors comme :

$$x(n) = s(n) + w(n)$$

où $s(n) = \sum_{k=0}^{K-1} \alpha_k z_k^n$ est le signal pur, $w(n)$ est un bruit blanc gaussien, $K \in \mathbb{N}$ représente l'ordre de discrétisation, les **complexes** $\alpha_k = a_k e^{i\varphi_k}$ contiennent l'information relative à la phase

Validation des descripteurs

Nous évaluons ensuite leur pertinence par une **analyse prédictive**. Cette analyse repose sur le principe de la régression logistique binaire où la probabilité d'appartenance ou non à une catégorie est donnée par :

$$\pi(x) = P(Y = 1/X = x) = f(L_{cat}(x)) = \frac{e^{L_{cat}(x)}}{1 + e^{L_{cat}(x)}}$$

La loi $L_{cat}(x) = \beta_0 + \sum_{i=1}^n \beta_i x_i$ est une **combinaison linéaire de nos descripteurs** x_i et les β_i , leurs coefficients associés.

S'en suivent une **étape de calibration puis de validation**, qui permettent d'évaluer la performance du modèle prédictif ainsi mis en place.

correspond au centre de gravité du spectre d'énergie du signal et qui est communément associé à la **brillance** du son.

L'étalement spectral (SSP) qui correspond à une largeur de bande et qui est fonction du centre de gravité spectral.

- FORMES

Les MFCC (Mel-Frequency Cepstral Coefficients) des densités de modes et niveaux d'énergie, calculés en effectuant une Transformée en Cosinus Discrète sur les bandes critiques (échelle de Bark) et qui traduisent l'allure formantique des spectres.

Mais aussi **l'amortissement** et le **CGS**.

- PLEIN/CREUX

L'amortissement et les **MFCC** car adaptés à la description formantique du spectre.

$\varphi_k \in [0, \pi]$ et à l'amplitude à l'origine $a_k \in \mathbb{R}_+$, et les **complexes** $z_k = e^{-\delta_k} e^{i2\pi f_k}$, dits **pôles**, contiennent l'information relative à l'amortissement $\delta_k \in \mathbb{R}$ et à la fréquence $f_k \in \mathbb{R}$ de la sinusoïde.

L'objectif de l'analyse est de **trouver l'ensemble de pôles complexes** z_k et d'**amplitudes complexes** α_k qui **décrit au mieux le signal considéré**. Ces paramètres ayant une incidence directe sur le rendu perceptif, une écoute comparative entre le signal d'origine et le signal resynthétisé permet de déterminer les meilleurs critères à utiliser au sein de l'algorithme. Nous construisons alors nos **descripteurs acoustiques** à l'aide des **paramètres** $\{a_k, f_k, \delta_k\}$ estimés.

Le tableau ci-dessous répertorie les descripteurs retenus par notre premier modèle et la performance obtenue par catégorie.

| Descr. Cat. | α_k | α_r | CGS | SSP | Rugo | MFCC des modes | | | MFCC des énergies | | | % correct global |
|-------------|------------|------------|-----|-----|------|----------------|----|----|-------------------|----|----|------------------|
| | | | | | | D1 | D2 | D3 | E1 | E2 | E3 | |
| Carton | x | x | x | | | | | x | | | | 92 |
| Métal | x | x | | x | | | | | | | | 80 |
| Céram | | | x | x | | | | | | | | 77 |
| Pierre | | | x | x | | | | | x | | | 68 |
| Verre | x | x | | | | | | | | | | 68 |
| Bois | x | | x | x | | | | x | | | | 63 |
| Plast | x | x | | | | | | x | | | | 32 |
| 2D | | | x | | | | | | | | | 58 |
| 3D | x | x | | | | | | | x | | | 54 |
| Plein | x | x | | | | | | x | x | | | 74 |
| Creux | x | x | | | x | | | | x | x | | 69 |

Références

M. Mathews. The digital computer as a musical instrument. *Sciences*, 142(11), 1963.
 W.W. Gaver. How do we hear in the world? explorations in ecological acoustics. *Ecological psychology*, 5(4) :285-313, 1993.
 M. Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4) :1-23, 1966.
 M. Aramaki, C. Gondre, R. Kronland-Martinet, T. Voinier, S. Ystad, et al. Thinking the sounds : an intuitive control of an impact sound synthesizer. 2009.
 R Badeau. *Méthodes à haute résolution pour l'estimation et le suivi de sinusoides modulées. Application aux signaux de musique*. PhD thesis, ENST, 2005.

Perspectives

Les **stratégies de mapping** et de **contrôle intuitif de la synthèse sonore** qui constituent le IV. de notre étude dépendront de l'application. La première concernera le **synthétiseur de sons d'impact** actuellement développé au LMA, organisé en trois couches hiérarchiques : labels décrivant l'objet en supérieure, descripteurs acoustiques en intermédiaire et paramètres de synthèse en inférieure. Plus largement, les **applications visées** sont le design sonore, la sonification et le contexte multimodal tel que les jeux vidéos, faisant intervenir la notion d'**interaction avec l'utilisateur**.

Afin d'**affiner les catégories perceptives des formes** et la sémantique associée, nous proposons actuellement un nouveau test perceptif où seule la forme perçue est évaluée et un nom d'objet est demandé.

En termes de descripteurs, il s'agit d'une part d'améliorer l'analyse-synthèse des sons avec ESPRIT **en se basant sur des critères perceptifs**, et d'autre part, de travailler sur le choix et la construction même des descripteurs, comme pour la rugosité par exemple.