



## **Identification de paramètres physiques du signal pour l'évaluation de la gêne en bureaux ouverts**

J. Rouch et P. Chevret

INRS - Dept I.E.T., Rue du Morvan CS 60027, 54519 Vandoeuvre-Lès-Nancy, France  
jeremy.rouch@inrs.fr

Le projet de recherche GABO (Gène Acoustique en Bureaux Ouverts) financé par l'ANSES a pour objectif de quantifier la gêne sonore dans les espaces de travail ouverts du secteur tertiaire. Le travail présenté ici constitue l'une des étapes de ce projet en s'attachant à l'analyse physique des signaux sonores élémentaires qui composent l'ambiance bruyante d'un bureau paysager. Une série d'enquêtes préalables, réalisées auprès de salariés du secteur tertiaire, a déjà permis d'identifier les principales sources de bruit gênantes dans un bureau ouvert. Suite à ces résultats, il a été choisi de mettre en œuvre un modèle de classification automatique de signaux sonores suivant leur source pour, à terme, détecter la présence de ces sources dans l'ambiance acoustique d'un bureau ouvert.

Après avoir présenté la méthode de classification retenue, nous décrirons l'échantillon de signaux sonores servant de base d'apprentissage et les paramètres physiques initialement choisis. Nous présenterons ensuite les deux méthodes que nous avons utilisées successivement pour la sélection des paramètres les plus pertinents. Les résultats de ces deux méthodes seront alors exposés et discutés.

## 1 Introduction

Le bruit en bureaux ouverts est souvent cité par les employés travaillant dans ce type d'espace comme un facteur important de gêne. Cette gêne est souvent reliée, par les personnes concernées, aux différentes sources sonores à l'origine de l'ambiance acoustique du lieu de travail. Partant de ce constat, le projet GABO (Gène Acoustique en Bureaux Ouverts) a été orienté vers la mise au point d'une méthode d'évaluation de la gêne acoustique à partir d'une part, de l'identification des différentes sources responsables de cette gêne et de la détermination du pouvoir de nuisance de chacune d'entre elles, et d'autre part, de la possibilité de détecter ces sources dans les ambiances sonores de bureaux ouverts. L'idée générale de la méthode d'évaluation de la gêne sonore en bureau paysager qui pourra être développée au terme de ce projet consistera donc, à partir d'un enregistrement de l'ambiance acoustique d'un bureau ouvert, à déterminer les différentes sources présentes et leurs occurrences, puis à affecter en fonction de ces données et du pouvoir de nuisance de chaque source une valeur quantitative de la gêne acoustique.

Une étude préliminaire conduite en 2012 par M. Pierrette [1] a permis de mettre en avant 4 types de sources gênantes : les sonneries de téléphones, les paroles intelligibles, les paroles non-intelligibles et les bruits de machines de bureau (imprimantes, photocopieurs). La gêne sonore induite par ces sources est actuellement étudiée en laboratoire à l'INSA de Lyon [2]. La méthode de détection automatique de ces sources est menée en parallèle au sein de l'INRS. Pour accomplir cette tâche, dans un premier temps nous souhaitons nous intéresser à la discrimination de chacune de ces quatre sources. Cela revient à se poser la question de savoir comment déterminer automatiquement si un signal sonore est une sonnerie de téléphone, de la parole intelligible, de la parole non-intelligible ou un bruit produit par une machine.

Pour aborder ce problème nous avons choisi de mettre au point une méthode de classification automatique à partir de données physiques issues du signal. La méthode retenue est la régression logistique multinomiale (RLM). Il s'agit d'une méthode relativement peu restrictive car semi-paramétrique, et qui, au-delà de la simple classification, aboutit à des modèles facilement interprétables car basés sur une approche linéaire [3]. Cette méthode s'avère avoir déjà été utilisée pour des problèmes similaires, notamment en classification automatique de la musique [4] ou en évaluation de la gêne en acoustique environnementale [5].

Avant d'établir le modèle final de classification, il est nécessaire de définir et de sélectionner les variables d'entrée de ce modèle ; c'est à dire les descripteurs extraits

d'une analyse physique du signal qui permettront d'obtenir un modèle efficace et le plus simple possible. Le travail présenté ici consiste en la description des méthodes et des résultats de cette étape de sélection. Après avoir rappelé le principe de la RLM, nous présenterons le panel de départ des descripteurs physiques du signal choisi. Nous exposerons ensuite les deux méthodes utilisées successivement pour déterminer le panel final de ces descripteurs.

## 2 La régression logistique multinomiale (RLM)

Dans cette partie nous allons brièvement décrire le principe de la RLM et la méthode de détermination des paramètres d'un modèle basé sur cette approche. Cette partie, assez générale, est reprise de plusieurs ouvrages, en particulier [3] [6] [7].

### 2.1 Principe et formulation générale

La RLM est une méthode statistique qui, utilisée dans un but d'inférence (prédiction), permet d'estimer une variable expliquée catégorielle à partir d'un jeu de variables explicatives. Considérons un échantillon d'apprentissage dont chaque individu est composé de  $J$  données explicatives  $X_j$  (ou variables exogènes) et d'une variable expliquée (ou variable endogène) catégorielle appartenant à un ensemble de  $K$  classes. La RLM établit  $K-1$  modèles du type

$$C_{k \neq ref} = a_{k,0} + \sum_{j=1}^J a_{k,j} X_j \quad (1)$$

Où  $a_{k,j}$  sont les coefficients du modèle associé à la classe  $k$  par rapport à la classe de référence (qui est l'une de  $K$  classes de la variable expliquée). Les  $C_k$  (ou *Logit*) sont des fonctions logarithmes du rapport des probabilités d'appartenance d'un individu à la catégorie  $k$  par rapport à la catégorie de référence<sup>1</sup> :

$$C_k = \ln \left( \frac{\pi_k}{\pi_{ref}} \right) \quad (2)$$

Le choix de la classe de référence peut s'effectuer au hasard. Certes les coefficients du modèle dépendent de ce choix, mais il n'influence pas les valeurs finales des probabilités d'appartenance d'un individu à la catégorie  $k$ .

<sup>1</sup> On notera que  $C_{k=ref} = 0$ .

Ces probabilités sont données par :

$$\pi_k = \frac{\exp(C_k)}{1 + \sum_{k=1}^{K-1} \exp(C_k)} \quad \forall k \in K \quad (3)$$

Pour affecter une classe à un individu, il suffit de suivre une règle d'affectation. La plus utilisée et la plus logique étant d'affecter un individu à la classe pour laquelle sa probabilité d'appartenance est maximale. Soit :

$$Y = y_k \Leftrightarrow y_k = \arg \max_k (\pi_k) \quad (4)$$

## 2.2 Recherche des paramètres d'un modèle

Compte-tenu de l'aspect catégoriel de la variable expliquée, la détermination des paramètres du modèle ne peut s'effectuer à partir de la méthode des moindres carrés. Elle s'effectue à partir de la recherche du maximum de vraisemblance ou du maximum du logarithme de la vraisemblance. La fonction de log-vraisemblance appliquée à la RLM est une double somme sur les  $\omega$  individus ayant servi à la construction d'un modèle et sur les  $K$  classes :

$$\text{Log}L = \sum_{\omega} \sum_k y_k(\omega) \cdot \ln(\pi_k(\omega)) \quad (5)$$

Les termes  $y_k(\omega)$  sont des termes de sélection ; ils valent 1 si l'individu  $\omega$  appartient effectivement à la classe  $k$  et 0 s'il n'y appartient pas.

D'un point de vue pratique, la détermination des estimateurs des paramètres maximisant la log-vraisemblance peut se faire à l'aide de la méthode de recherche d'extremum de Newton-Raphson. Un premier jeu de paramètres est fixé, puis par évaluations successives des dérivés partielles premières et secondes des modèles de RLM, ce jeu de paramètres est actualisé jusqu'à ce qu'il corresponde au modèle maximisant la vraisemblance (à une valeur d'erreur près).

## 3 Échantillon et descripteurs de signaux sonores rencontrés en bureaux ouverts

Le problème que nous souhaitons traiter à partir de la RLM est la classification automatique de signaux sonores suivant leur origine. Il nous faut donc constituer un panel (ou échantillon) d'apprentissage composé de signaux sonores. Les variables expliquées sont les sources sonores d'origine ; les variables explicatives sont un jeu de descripteurs calculés sur ces signaux.

Dans cette partie nous allons dans un premier temps détailler les signaux que nous avons sélectionnés et utilisés. Dans un second temps, nous décrirons les différents descripteurs initialement choisis.

### 3.1 Échantillon des signaux sonores

Nous nous intéressons à quatre types de sons : les sonneries de téléphone, ceux d'imprimantes et de photocopieurs, ceux de conversations intelligibles, et ceux de conversations non-intelligibles. Chaque signal est composé du son isolé produit par l'une de ces quatre

sources. Ils sont tous au format WAV, échantillonnés à 44100 Hz, et durent 12 s. Tous ont été calibrés de sorte à avoir un Leq12s à 55 dB(A). 30 signaux (ou individus) par type de sources ont été sélectionnés, soit 120 pour l'ensemble de notre échantillon.

Les signaux de sonnerie téléphoniques sont d'origine numérique, tels que rencontrés aujourd'hui dans la grande majorité des bureaux. Les durées et les intervalles entre deux sonneries pouvant varier selon l'enregistrement, ce signaux contiennent 3, 4, 5 ou plus de sonneries individuelles.

Les signaux d'imprimantes et photocopieurs correspondent à l'enregistrement d'une ou de plusieurs phases d'impression (préchauffage, impression, mise à l'arrêt, etc.) sans manipulations artificielles ; il s'agit simplement d'une portion de 12 s d'un enregistrement réalisé lors d'une impression ou d'une photocopie. Les signaux d'imprimantes peuvent être issus d'imprimantes laser ou à jet d'encre.

Les signaux de paroles intelligibles, correspondent à des enregistrements d'une personne à la fois, lors d'une interview radiophonique. Cette origine permet un enregistrement de qualité tout en assurant un débit, une diction et un phrasé spontanés, semblable à ce qui peut être observé dans un environnement de travail classique.

Les signaux de parole non intelligibles sont constitués d'une superposition de 5 signaux de paroles intelligibles issus de ceux décrits précédemment. Le caractère non-intelligible a été contrôlé *a posteriori* par écoutes.

### 3.2 Choix initial de descripteurs du signal

La sélection des variables explicatives d'un modèle de RLM pour la classification automatique des sons de bureau est initiée dès le choix initial de descripteurs audio. Ces descripteurs sont des quantités obtenues sur chaque signal à partir de méthodes de traitement du signal. Le choix d'un descripteur peut avoir une origine différente suivant le but recherché (RLM inférentielle ou descriptive, rapidité d'extraction, etc.) Ce choix peut aussi être guidé par une connaissance du domaine qui mettrait alors en avant des descripteurs *a priori* discriminants. Pour la situation propre aux travaux présentés ici, les descripteurs initiaux ont été choisis selon trois axes : ceux *a priori* discriminants, ceux qui sont déjà utilisés dans un cadre proche de notre étude, à savoir l'analyse automatique de la musique, ceux liés à des critères psycho-acoustiques ou à des indicateurs déjà utilisés en acoustique pour une évaluation de la gêne.

Deux paramètres *a priori* discriminants ont été choisis : l'indice de modulation à 4 Hz (Mod4Hz) du signal que l'on sait particulièrement élevé dans la parole intelligible (la fréquence 4 Hz correspondant à la moyenne du débit syllabique), et le maximum de la fonction d'autocorrélation normalisée en valeur absolue (RxxMax) que l'on sait important pour les sonneries de téléphone.

Six paramètres physiques du signal utilisés dans le cadre de la classification automatique de la musique ont été choisis [8]. Le centre de gravité spectral (SC), le « Roll-Off » et la brillance sont des descripteurs liés à la répartition spectrale de l'énergie. Le « Roll-Off » correspond à la fréquence du spectre en dessous de laquelle se concentre 85% de l'énergie du signal. La brillance correspond au pourcentage d'énergie du spectre au-dessus d'une fréquence fixe (1500 Hz). Le centre de gravité spectral correspond au barycentre du spectre. Le Flux

spectral moyen (FS\_Moy) est un indicateur de variabilité du spectre d'un signal en fonction du temps. Il s'agit de la valeur moyenne de l'écart (au sens des moindres carrés) entre les spectres de deux fenêtrages successifs du signal temporel. Le paramètre « Noisiness » mesure l'absence de caractère tonal d'un signal en faisant le rapport de l'énergie de bruit du spectre (énergie totale – énergie des composantes tonales) sur l'énergie totale du spectre. Enfin, nous avons choisi d'inclure la moyenne des hauteurs de la fréquence fondamentale (F0) calculée sur les spectres de fenêtrages successifs du signal (lorsqu'aucune fréquence fondamentale n'est détectée, la valeur de ce paramètre est mise à 0 Hz). Pour le « Noisiness » et la hauteur de la fréquence fondamentale, l'extraction de la partie tonale du signal a été réalisée à l'aide de l'algorithme de détection d'émergences fréquentielles de *Terhardt* [9].

Quatre paramètres issus de l'analyse acoustique environnementale et de la psycho-acoustique ont été choisis. La différence de niveau fractile est une manière de quantifier les émergences énergétiques temporelles ( $\Delta L_{Aeq} = L_{Aeq50\%} - L_{Aeq5\%}$ ). Un indicateur semblable à celui présenté ci-dessus a été construit à partir de la sonie fractile des sons dynamiques telle que celle modélisée par *Zwicker* [10] ( $\Delta N = N_{5\%} - N_{50\%}$ ). Nous avons aussi considéré une valeur d'acuité médiane ( $Ac_{50\%}$ ) obtenue par un algorithme basé sur la modélisation de ce paramètre telle que proposée par *Von Bismarck* [11]. Enfin, nous avons inclus dans le panel initial de descripteurs, la tonalité moyenne (T\_Moy) calculée à partir du modèle proposé par *Aures* [12].

## 4 Sélection de variables explicatives par diminution de la colinéarité

### 4.1 Principe

La RLM étant une méthode de régression linéaire, elle demande une considération particulière des éventuelles relations linéaires entre les variables explicatives (colinéarité). La présence de telles relations entraîne nécessairement une augmentation importante de l'incertitude sur les estimations des paramètres du modèle. Ce dernier est alors instable car très sensible à une très petite variation de la taille de l'échantillon d'apprentissage. De plus, ces paramètres peuvent prendre des valeurs ou des signes contradictoires avec les connaissances du domaine abordé, faussant alors toute éventuelle démarche descriptive. Des problèmes méthodologiques dans la construction d'un modèle peuvent aussi apparaître avec la présence de fortes colinéarités entre variables explicatives. Ces modèles sont construits à partir de l'inversion de la matrice ( $X^T X$ ) où  $X$  est la matrice des observations des variables explicatives. Or, en cas de fortes colinéarités celle-ci peut apparaître singulière ou mal conditionnée, ce qui entraîne une impossibilité d'inversion ou des erreurs numériques d'inversion. Enfin, la colinéarité entre deux variables explicatives entraîne une redondance d'informations dans le modèle, allant alors l'encontre du principe de parcimonie si celui-ci est voulu ; ce qui est notre cas car nous souhaitons aboutir à un modèle simple en ne conservant qu'un nombre minimum de variables explicatives.

### 4.2 Méthode

Afin, de détecter la colinéarité entre variables explicatives nous avons choisi de nous baser sur une quantité souvent citée dans la littérature et qui présente l'avantage d'être directement liée à l'incertitude sur l'estimateur du modèle. Cette quantité, appelée facteur d'inflation de la variance ou VIF (pour Variance Inflation Factor) existe pour chaque variable explicative [13]. La relation entre la colinéarité et le VIF est évidente lorsque celui-ci est écrit en fonction du coefficient de détermination  $R_j^2$  de la régression linéaire de variable explicative  $j$  par rapport aux autres variables explicatives :

$$VIF_j = \frac{1}{1 - R_j^2} \quad (6)$$

La relation entre la variance  $V(a_j)$  de l'estimateur d'un paramètre  $a_j$  et le VIF <sub>$j$</sub>  de la variable explicative  $j$  est donnée par :

$$V(a_j) = \frac{\sigma_\varepsilon^2}{n\sigma_j^2} VIF_j \quad (7)$$

où,  $\sigma_\varepsilon^2$  est la variance des résidus,  $\sigma_j^2$  celle des valeurs de la variable explicative  $j$  et  $n$  le nombre d'individus dans l'échantillon d'apprentissage. Cette formulation fait clairement apparaître la relation de proportionnalité entre la variance d'un estimateur et le VIF, et permet donc de fixer une limite objective de tolérance concernant la valeur du VIF. Couramment, celle-ci est fixée à  $VIF < 4$ , car des valeurs supérieures de VIF correspondent à plus qu'un doublement de l'écart-type de l'estimateur.

Afin d'éliminer les variables présentant une forte colinéarité nous proposons une méthode récursive qui consiste à calculer le VIF pour chaque variable puis à éliminer l'une des variables qui présenterait une colinéarité importante ( $VIF > 4$ ). Le VIF est alors à nouveau calculé sur les variables restantes puis, en fonction de valeurs nouvellement prises par cet indicateur, une seconde variable est éliminée. Ces deux opérations (calcul de VIF et élimination d'une variable) sont répétées jusqu'à ce que le VIF de chacune des variables restantes soit inférieur à 4. La sélection de la variable à éliminer à chaque itération peut être envisagée à partir du VIF maximum. Cependant, lorsque plusieurs variables ont des VIF importants et du même ordre de grandeur, il peut être raisonnablement supposé que les relations de colinéarité importantes mises ainsi en évidence ont lieu entre ces variables. Si tel est le cas, la conséquence de la suppression de n'importe laquelle de ces variables n'est alors que peu dépendante de la variable éliminée ; les valeurs de VIF à l'itération suivante seront à peu près identiques. Ainsi, plutôt que d'effectuer une suppression automatique de variable par détection de VIF maximum, il peut être avantageux de laisser ce choix à l'utilisateur qui peut alors faire un compromis entre valeur du VIF et d'autres dimensions du problème ; dimensions liées par exemple à des considérations pratiques comme une complexité de calcul de la variable explicative.

### 4.3 Résultats

Afin d'appliquer cette méthode de sélection de variable pour traiter la colinéarité entre descripteurs initiaux, nous avons considéré l'ensemble de l'échantillon des 120 signaux sonores. Pour chacun, les douze descripteurs

présentés à la partie 3.2 (variables explicatives) ont été calculés. Les VIF de ces douze variables ont été calculés. Certains descripteurs ont été éliminés au fur et à mesure en fonction du VIF correspondant si celui-ci était supérieur à 4, ainsi qu'en fonction de la complexité de leur calcul. Par exemple, le calcul de la sonie ou de la tonalité s'avère relativement complexe, alors que le calcul du  $\Delta L_{Aeq}$  ou celui du « Noisiness » s'avère plus simple.

Les résultats successifs de cette sélection de variables sont présentés dans le tableau 1. On peut y lire que pour un ensemble de départ de douze critères explicatifs, nous avons abouti à un ensemble de six critères relativement indépendants (au sens de « peu colinéaires » ou  $VIF < 4$ ). Sur les trois descripteurs psychoacoustiques, nous avons pu en éliminer deux, ne reste que la différence de sonie fractile qui est le plus simple à calculer. L'indicateur d'émergence  $\Delta L_{Aeq}$  a été conservé. Les deux descripteurs *a priori* (Indice de modulation à 4 Hz et maximum de d'autocorrélation) discriminants ont été conservés. Concernant les autres descripteurs, ceux en lien avec la répartition spectrale de l'énergie ont tous été supprimés (« Roll-Off », centre de gravité spectral et Brillance) ainsi que la fréquence fondamentale moyenne ; il ne reste que le flux spectral et le « Noisiness ».

Le tableau 1 illustre aussi l'explication faite précédemment concernant la suppression de variables à VIF forts mais comparables. En effet, si l'on s'intéresse à la première colonne du tableau (It\_0), nous pouvons constater que la variable au plus fort VIF est le centre de gravité spectral (SC). Cependant nous avons choisi de supprimer le « Roll-Off » qui a un VIF un peu moins important mais du même ordre de grandeur. À l'itération suivante (It\_1), ce choix a fait fortement baisser le VIF du centre de gravité spectral, sans pour autant diminuer les VIF des autres variables de manière aussi importante. Ainsi, les valeurs initiales très importantes de VIF du centre de gravité spectral et du « Roll-Off » indiquent bien que ces deux descripteurs sont surtout corrélés entre eux et beaucoup moins avec les autres. Supprimer l'un ou l'autre est nécessaire pour la stabilité d'un modèle de RLM, mais un choix entre les deux, motivé uniquement par une valeur de VIF maximum, n'aurait été que peu pertinent.

Tableau 1 - VIF associé à chaque variable explicative (descripteur) en fonction de la passe de sélection (It).

	It_0	It_1	It_2	It_3	It_4	It_5	It_6
Mod_4Hz	2,4	2,3	2,3	2,3	2,3	2,3	2,3
RxxMax	7,4	7,4	7,3	7,3	5,0	4,9	3,6
SC	87,9	17,0	10,6	8,6			
Roll-Off	75,4						
Brillance	11,4	11,3	9,6				
FS_Moy	2,0	2,0	1,7	1,7	1,7	1,6	1,6
Noisiness	6,0	5,8	5,7	5,7	5,7	3,8	3,7
F0	6,7	6,7	6,3	5,0	2,9	2,9	
$\Delta L_{Aeq}$	4,2	4,1	4,0	3,7	3,7	3,6	3,6
$\Delta N$	3,9	3,9	3,8	3,8	3,7	3,7	3,7
Ac <sub>50%</sub>	17,1	16,1					
T_Moy	2,3	2,3	2,2	2,2	2,0		

## 5 Sélection de variables explicatives par RLM récursives

### 5.1 Principe

La méthode de sélection de variables explicatives présentée à la partie 4 nous a permis de ne retenir que des variables qui interfèrent peu entre elles. Cependant, parmi toutes ces variables, certaines ne sont pas utiles à ce modèle. Afin de ne retenir que le nombre minimum de variables explicatives nécessaires nous proposons d'effectuer une seconde étape de sélection de variables. Celle-ci repose sur les constructions successives de modèles incluant une ou plusieurs des variables explicatives et un test de performance. Les variables à finalement retenir étant celles qui apparaissent dans le modèle réussissant le mieux le test de performance. Ce test de performance consiste en l'application du modèle à  $n$  individus dont la classe est préalablement connue, puis à dénombrer les  $n_{faux}$  individus mal classés par le modèle. Le critère de performance final choisi est un taux d'échecs en prévision défini par :

$$\tau_{err} = \frac{n_{faux}}{n} \quad (8)$$

### 5.2 Méthode

Afin d'éviter de biaiser les résultats d'évaluation de performance en testant les différents modèles construits sur des données ayant aussi servi à leur apprentissage, nous avons choisi de mettre en place une procédure de validation croisée. L'échantillon initial est divisé en deux sous échantillons, un pour les données d'apprentissage, un pour les données de validation. Il est alors possible, pour un même jeu de critères explicatifs retenu de construire plusieurs modèles reposant chacun sur un échantillon d'apprentissage différent. Si l'échantillon de test de deux modèles avec un même jeu de paramètres est composé d'individus strictement différents on parle alors de validation croisée à plis. Cette approche permet d'assurer aux différents modèles ainsi testés d'être les plus différents possibles car construits sur des échantillons d'apprentissage les plus différents possibles. Cependant, même si une telle stratégie est mise en place, pour un jeu de paramètres donné nous ne pouvons obtenir que peu de modèles à tester par rapport au nombre de modèles possible. Dans notre cas, si la partition de l'échantillon initial est de 80% d'individus pour l'apprentissage et 20% pour le test (soit par classe, 24 signaux pour l'apprentissage et 6 signaux tests), il est possible de faire une validation croisée à 5 plis. Ce qui équivaut à construire 5 modèles différents, alors que le nombre de modèles possibles s'élève à  $\binom{24}{30}^4 \approx 1,2 \cdot 10^{23}$ .

Nous avons choisi d'effectuer 200 validations croisées à 5 plis en faisant varier les individus sélectionnés dans les échantillons d'apprentissage et de test. Ainsi, la pertinence d'un jeu de variables explicatives sera évaluée à partir de 1000 modèles différents.

Compte tenu du fait que nous disposons d'un jeu initial réduit à six paramètres et de capacités de calcul suffisantes, il nous a été possible de tester toutes les combinaisons de ceux-ci (soit  $2^6 - 1 = 63$ ) plutôt que devoir mettre en œuvre une approche par sélection rétrograde ou progressive qui consisterait à diminuer ou augmenter petit à petit le nombre de variables.

Pour chaque jeu de paramètres possible, un nombre important de modèles est donc testé. Pour chacun de ces modèles, le taux d'erreur en prévision est donné. Pour sélectionner le jeu de paramètres optimal, il peut alors être envisagé de comparer un descripteur statistique de ce taux d'erreur comme sa moyenne. Cependant, il convient au préalable de vérifier la pertinence de l'utilisation de ce descripteur, par exemple en observant la densité de probabilité de ce taux d'erreur.

### 5.3 Résultats

L'application de la méthode de sélection de variables présentée précédemment a été réalisée sur notre échantillon initial en considérant une séparation de celui-ci comme proposée dans le paragraphe précédent (80%-20%). La figure 1 représente les densités de probabilité expérimentales des taux d'erreurs obtenues pour trois jeux de paramètres à taux d'erreur moyen différents. Selon le jeu de paramètres utilisé, il peut apparaître plusieurs maximums locaux dans ces distributions, cependant un maximum global reste toujours clairement identifiable. L'utilisation d'une valeur moyenne de taux d'erreur pour qualifier la qualité d'un modèle reste donc à peu près pertinente. On observe aussi une dispersion de ces distributions qui peut être importante. Cela traduit le fait que pour un même jeu de paramètres, les 1000 modèles construits ne sont pas très stables en prévision. Ceci pourrait probablement être dû au trop faible effectif d'apprentissage. Il conviendrait donc, avant de construire un modèle définitif de RLM, d'augmenter le nombre d'individus par classe de notre échantillon d'apprentissage.

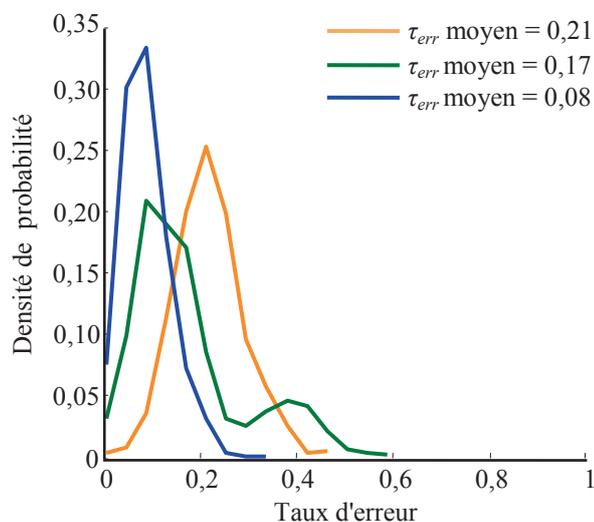


Figure 1 - Densité de probabilité empirique du taux d'erreur de classement de trois jeux de paramètres différents.

Le tableau 2 fait apparaître les moyennes des taux d'erreurs observées pour 6 jeux de paramètres testés (moyennes effectuées sur les 1000 modèles), ainsi que le classement de ces jeux de paramètres relativement à tous les autres en fonction de cette valeur moyenne. Il y apparaît aussi, pour comparaison, le taux d'erreur du modèle trivial. On peut y lire que quel que soit le jeu de paramètres testé, celui-ci n'incluant même qu'un seul paramètre, le taux d'erreur moyen est inférieur à celui du modèle trivial. Ainsi, tous les descripteurs utilisés semblent avoir individuellement un effet sur la classification.

Nous pouvons aussi lire dans le tableau 2 que le jeu de paramètres ayant le plus mauvais pouvoir moyen de prévision (taux d'erreur moyen de 0,53) n'est composé que d'un seul paramètre. Ce résultat n'est pas surprenant puisqu'il est logique que plus un modèle est restreint, moins il a de potentiel explicatif. Il ne faut pas pour autant considérer cette réflexion comme une généralité. En effet, il apparaît que le jeu de paramètre classé 52<sup>ème</sup> parmi les 63 possibles par rapport à leur taux d'erreur, comporte plus de paramètres que les deux jeux de paramètres classés 2<sup>ème</sup> et 3<sup>ème</sup>. Ce comportement trouve probablement une partie de son explication dans le fait que même si la colinéarité entre les variables explicatives a été potentiellement réduite, certaines variables peuvent toujours interférer entre elles.

Parmi les quatre jeux de paramètres produisant le taux d'erreur moyen le plus faible, celui incluant l'ensemble des six paramètres sélectionnés par traitement de la colinéarité n'est classé que quatrième. Cependant, il faut noter que ce classement ne doit être pris qu'à titre indicatif. D'une part la question de l'efficacité relative d'un jeu de paramètres ne peut, en toute rigueur, être abordée qu'à partir d'une analyse statistique de significativité de différences pas encore conduite à ce stade de nos travaux et, d'autre part la sélection d'un jeu de paramètres doit prendre en compte d'autres critères que sa seule efficacité, par exemple la facilité de calcul des paramètres. Si nous mettons les quatre premiers jeux de paramètres sur un pied d'égalité vis-à-vis de la qualité de prévision qui ne diffère en moyenne que de 1%, nous pouvons constater que l'on peut se contenter de trois variables explicatives pour affecter un signal à sa classe d'appartenance avec une moyenne de 8% à 9% d'erreur. Ces trois variables sont soit le flux spectral moyen, le « Noisiness » et la différence de sonie fractile, soit le flux spectral moyen, le « Noisiness » et l'indice modulation à 4 Hz. Nous retiendrons ce dernier ensemble de paramètres car la modulation à 4 Hz est de calcul plus simple que la sonie. De plus, comme déjà signalé dans la sous-partie 3.2, la modulation à 4 Hz peut être plus facilement mise en relation avec l'une des classes d'appartenance de nos signaux, à savoir la parole intelligible, facilitant ainsi l'interprétation d'un modèle basé sur ce paramètre.

Tableau 2 - Résultats de la sélection de variables par RLM successives.

Mod_4Hz	X	X		X			
RxxMax	X			X	X		
FS_Moy	X	X	X	X	X		
Noisiness	X	X	X	X	X		
$\Delta L_{Aeq}$				X	X		
$\Delta N$	X		X	X	X	X	
Taux d'erreur Moyen	0,08	0,08	0,09	0,09	0,35	0,53	0,75
Classement	1	2	3	4	52	63	modèle trivial

## 6 Conclusion

Les travaux présentés ici se sont intéressés au problème de la sélection de descripteurs acoustiques pertinents pour

l'établissement d'un modèle de classification automatique de signaux sonores rencontrés en bureaux ouverts. Après avoir présenté la méthode de classification retenue, à savoir la régression logistique multinomiale, nous avons décrit l'échantillon des signaux sonores utilisé pour effectuer cette sélection. Sur 120 signaux au total, 30 correspondent à des sonneries de téléphone, 30 à des bruits de machine, 30 à de la parole intelligible, et 30 à de la parole non-intelligible. 12 descripteurs acoustiques initiaux ont ensuite été présentés et décrits. Enfin, deux méthodes successives de sélection de variables ont été appliquées à ces descripteurs.

Afin d'éviter la redondance d'information utiles pour, d'une part tendre vers la simplicité d'un futur modèle de classification, et d'autre part permettre sa stabilité, la première méthode de sélection repose sur l'évaluation de la colinéarité entre les descripteurs. De 12 descripteurs initiaux, nous en avons ainsi retenu 6 jugés linéairement indépendants. Pour ne retenir que les plus pertinents parmi ces 6 descripteurs, nous avons utilisé une approche par validations croisées récursives de modèles de régression logistique. Il est alors apparu que potentiellement 3 descripteurs relativement simples de calcul pouvaient s'avérer suffisants : le flux spectral moyen, le « Noisiness » et l'indice de modulation à 4 Hz. Cependant, il s'est aussi avéré que ces conclusions devaient être plus amplement confirmées du fait d'une taille d'échantillon par classe de signaux tout juste suffisante. Ainsi, avant la création d'un modèle définitif de classification automatique, il conviendrait soit d'utiliser une base de données de signaux plus importante, soit éventuellement de diminuer le nombre de classes. Cette seconde option sera à envisager en fonction des tests psycho-acoustiques sur la nuisance individuelle de chaque source menée à l'INSA de Lyon. Cependant, si elle est envisagée, il conviendrait au préalable de réappliquer à ces nouvelles classes les méthodes de sélection de descripteurs acoustiques présentées ici.

## Remerciements

Ces travaux ont pour cadre le projet de recherche GABO (Gêne acoustique dans les bureaux ouverts) effectué avec le concours apporté par le Programme Environnement-Santé-Travail de l'ANSES avec le soutien des ministères chargés de l'écologie et du travail.

## Références

- [1] M. Pierrette, E. Parizet, et P. Chevret, Perception and evaluation of noise sources in open plan office, *21<sup>st</sup> International Congress on Acoustics* (2013).
- [2] L. Brocoloni et al., Influence du type de source sonore sur l'évaluation de la gêne et de la performance dans les bureaux ouverts. *12<sup>ème</sup> Congrès Français d'Acoustique. Poitiers* (2014).
- [3] T. Hastie, R. Tibshirani, J. Friedman, et J. Franklin, *The elements of statistical learning: data mining, inference and prediction*, Springer (2005).
- [4] C. Weihs, U. Ligges, F. Mörchen, et D. Müllensiefen, Classification in music research, *Advances in Data Analysis and Classification*, **1**, 255–291 (2007).
- [5] D. Botteldooren, L. Dekoninck, et D. Gillis, The influence of traffic noise on appreciation of the living quality of a neighborhood, *International journal of environmental research and public health*, **8**, 777–798 (2011).
- [6] D. W. Hosmer Jr et S. Lemeshow, *Applied logistic regression*. John Wiley & Sons (2004).
- [7] R. Rakotomalala, *Pratique de la Régression Logistique*. Notes de cours, Université Lyon 2 (2011). [http://eric.univ-lyon2.fr/~ricco/cours/cours/pratique\\_regression\\_logistique.pdf](http://eric.univ-lyon2.fr/~ricco/cours/cours/pratique_regression_logistique.pdf)
- [8] A. Lerch, *An introduction to audio content analysis: Applications in signal processing and music informatics*. John Wiley & Sons (2012).
- [9] E. Terhardt, G. Stoll, et M. Seewann, Algorithm for extraction of pitch and pitch salience from complex tonal signals, *The Journal of the Acoustical Society of America*, **71**, 679–688 (1982).
- [10] E. Zwicker, Procedure for calculating loudness of temporally variable sounds, *The Journal of the Acoustical Society of America*, **62**, 675–682, (1977).
- [11] G. von Bismarck, Sharpness as an attribute of the timbre of steady sounds, *Acta Acustica united with Acustica*, **30**, 159–172, (1974).
- [12] W. Aures, Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale, *Acta Acustica united with Acustica*, **59**, 130-141 (1985).
- [13] G. Saporta, *Probabilités, analyse des données et statistique*. Editions Technip (2011).