

CFA/VISHNO 2016

Détection de bruit de tirs et évaluation de la gêne à partir d'enregistrements audio en situation réelle

A.-C. Witsel et F. Moïny

UMons, 20, Place du Parc, 7000 Mons, Belgique
anne-catherine.witsel@umons.ac.be



LE MANS

Cet article présente une étude comparative de différents algorithmes de détection et classification des bruits de tirs à partir d'enregistrements audio réalisés en extérieur à proximité d'un club de tir. Cette recherche s'inscrit dans le cadre de la gestion du bruit généré par les installations extérieures de tir sportif et récréatif. Elle a pour objectif de faciliter le travail d'évaluation du niveau de gêne et de servir de base à un système de monitoring. Dans ce contexte, les algorithmes ont été comparés en termes de fiabilité mais aussi de rapidité de calcul et de ressources informatiques. Par ailleurs, les signaux de bruits de tirs ont été analysés afin de déterminer les caractéristiques temporelles et spectrales les plus adéquates pour les distinguer des autres bruits environnementaux.

1 Introduction

Les bruits de tir sont principalement causés par l'expansion rapide des gaz de combustion lorsque le projectile sort de la bouche du canon de l'arme (détonation à la bouche), ce qui provoque un bruit très impulsif et de grande amplitude. Par ailleurs, lorsque le projectile se déplace à une vitesse supersonique, une onde de choc acoustique est également produite. Ces caractéristiques acoustiques rendent les bruits de tir particulièrement gênants et audibles à grande distance.

Les activités de tir pratiquées en extérieur, comme le ball-trap, constituent donc une importante source de pollution sonore. Les clubs de tir sont soumis à des normes visant à limiter la charge acoustique due à toutes les détonations relevées durant une période de temps écoulée (généralement une heure). Si le bruit à la source est constant au cours du temps (combinaisons armes/munitions similaires pour tous les tirs), le niveau de bruit dans le voisinage peut donc être maîtrisé en limitant le nombre de tirs par heure.

Selon la norme internationale ISO 1996 [7, 8], il est recommandé d'utiliser le niveau d'exposition L_{AE} comme descripteur d'un bruit isolé et le niveau continu équivalent L_{Req, T_n} sur la période d'analyse T_n comme niveau d'évaluation. Pour les bruits de tirs, un facteur correctif de 12 dB doit être ajouté afin de tenir compte de leur caractère impulsif.

Cependant, lors de mesures réalisées en un même endroit mais dans des conditions de vent très différentes [11], nous avons mis en évidence des variations du niveau d'exposition moyen allant jusqu'à 19 dB, les niveaux à la source étant comparables. Le nombre de tirs autorisés par heure pour ne pas dépasser les limites fixées par les autorités locales varie donc fortement en fonction des conditions météorologiques.

Afin d'aider les clubs de tir à combiner une gestion efficace de leurs installations avec les normes acoustiques, nous développons un système de monitoring capable d'estimer le niveau d'évaluation de la gêne avec une précision comparable à celle obtenue avec une analyse manuelle. Ce système a pour objectif de fournir très régulièrement le nombre de tirs par heure maximal autorisé mais aussi d'avoir un suivi sur le long terme du niveau de bruit chez les riverains, les contrôles occasionnels de la police de l'environnement pouvant ne pas être représentatifs de la situation générale. Le principe de ce système de monitoring est schématisé sur la Figure 1.

Dans ce contexte, cet article se concentre sur le module d'identification des tirs et présente une étude comparative de différents algorithmes de détection et classification des bruits de tirs à partir d'enregistrements audio réalisés en extérieur à proximité d'un club de tir. La difficulté principale de cette détection automatique est la présence de bruits environnementaux dont le niveau peut être comparable aux bruits de tirs.

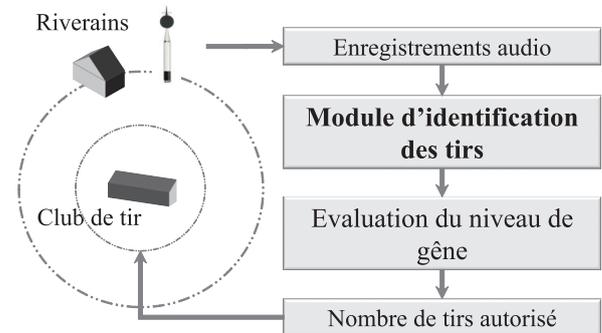


FIGURE 1 – Architecture du système de monitoring

2 Le module d'identification des tirs

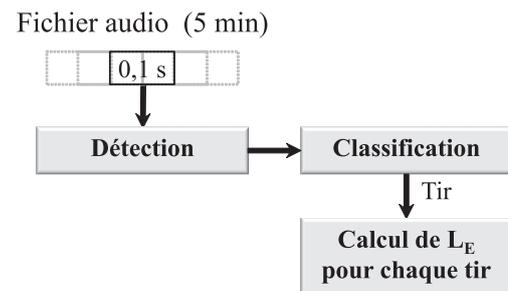


FIGURE 2 – Module d'identification des bruits impulsifs

Comme montré sur la Figure 2, les fichiers audio sont traités par tranche de 5 min. Le signal audio est d'abord divisé en segments de 0,1 s (avec un chevauchement de 50%). Cette durée a été choisie compte tenu du délai minimal entre deux détonations successives qui est de l'ordre de 0,2 s. Lorsque deux tireurs tirent en même temps, ce délai peut être inférieur mais dans ce cas il est difficile de les distinguer à l'oreille et les deux détonations sont comptées comme une seule. Réduire d'avantage la longueur des segments ne présente donc pas d'avantage.

Afin de limiter la charge de calcul, l'identification des tirs est réalisée en deux étapes : le module de détection détermine les segments correspondant à un bruit impulsif et ceux-ci sont ensuite classés comme tir ou comme autre bruit. Le début et la fin de chaque détonation sont ensuite définis plus précisément et leur niveau d'exposition est calculé. Toutes les cinq minutes, le système de monitoring peut donc estimer le niveau d'évaluation de la gêne et actualiser le nombre de tirs autorisés.

Les mesures étant réalisées à plusieurs centaines de mètres de la source de bruit, les niveaux relevés peuvent être comparables à certains bruits environnants et sont noyés dans le bruit de fond. Pour le calcul du niveau de gêne, un filtre de pondération A doit être appliqué au signal et le bruit de fond est ainsi fortement réduit.

2.1 Détection des bruits impulsifs

Cette étape de détection a une grande influence sur la performance globale du système d'identification des bruits de tirs, puisque les détonations non détectées lors de cette étape sont complètement perdues. On veillera cependant à limiter le nombre de faux positifs afin de faciliter leur élimination par l'étape de classification.

Il existe de nombreuses méthodes de détection des bruits impulsifs [4, 3, 10]. Dans cette étude, nous avons comparé deux méthodes : la première basée sur l'évolution temporelle de l'énergie quadratique moyenne, la seconde sur celle du flux spectral (distance entre les spectres de deux segments successifs).

Pour chaque segment de 0,1 s, l'énergie quadratique moyenne (e_k) ou le flux spectral (f_k) est calculé. Afin de faire ressortir d'avantage les pics, un filtre médian et un redressement en simple alternance sont successivement appliqués [1]. La fonction de détection p_k d'un signal v_k (e_k ou f_k) est déterminée comme suit :

$$m_k = \text{median}((g_k)) \quad (1)$$

$$\text{avec } g_k = \{v_{k-i}, \dots, v_k, \dots, v_{k+i}\} \quad (2)$$

$$\hat{p}_k = v_k - m_k \quad (3)$$

$$p_k = \begin{cases} \hat{p}_k & \text{si } \hat{p}_k > 0 \\ 0 & \text{dans les autres cas} \end{cases} \quad (4)$$

Pour assurer une bonne détection, la longueur du filtre médian ($2i + 1$) doit être supérieure à la largeur des pics. Dans notre cas, nous l'avons fixée à 0,5 s. La position des maxima locaux de la fonction de détection est enfin déterminée. Seuls les maxima présentant une différence d'amplitude avec les minima locaux précédent et suivant supérieure à un seuil sont retenus. La détermination de ce seuil est discutée au paragraphe 2.1.1.

La Figure 3 montre un exemple de portion de signal (pondéré A), l'évolution temporelle de son énergie quadratique moyenne et sa fonction de détection (les points verts correspondant aux bruits détectés).

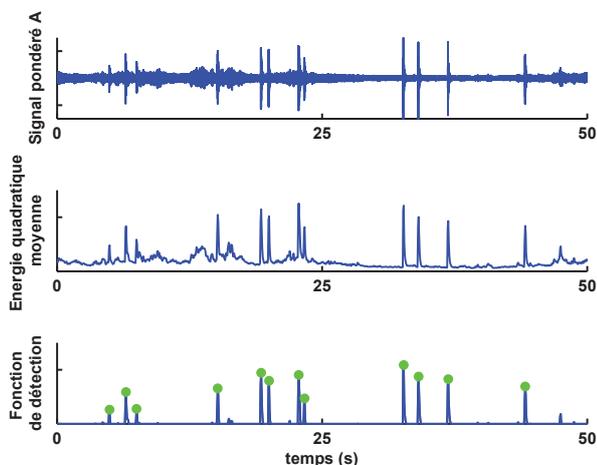


FIGURE 3 – Fonction de détection d'un signal

La détermination de l'énergie quadratique moyenne, du flux spectral et la recherche des maxima locaux ont été effectués sous Matlab à l'aide de la MIRtoolbox [9].

2.1.1 Détermination du seuil

Pour déterminer la valeur optimale du seuil de détection des maxima de la fonction de détection, nous avons calculé, pour différentes valeurs de seuil, le taux de vraies détections positives (TPR) et le taux de fausses détections positives (FPR) :

$$TPR = \frac{\text{Nombre de tirs correctement détectés}}{\text{Nombre total de segments avec tirs}} \quad (5)$$

$$FPR = \frac{\text{Nombre de fausses détections de tirs}}{\text{Nombre total de segments sans tirs}} \quad (6)$$

Le seuil optimal est choisi comme celui qui minimise la distance euclidienne entre la paire correspondante (TPR,FPR) et un détecteur parfait (TPR=1,FPR=0). La valeur retenue est donc celle qui maximise le nombre de tirs détectés en minimisant la détection d'autres bruits. Le seuil est optimisé séparément pour chaque méthode de détection en utilisant les données d'apprentissage (Données A).

2.2 Classification

2.3 Sélection des paramètres

Pour chaque segment détecté comme un bruit impulsif, un certain nombre de paramètres couramment utilisés en traitement du signal audio [6] sont calculés :

- le centre de gravité spectral (*Centroid (C)*) ;
- la fréquence délimitant 85 % de la puissance spectrale (*Roll-off point (RF)*) ;
- les 13 premiers *MFCC* ;
- le niveau continu équivalent par bande de tiers d'octave entre 100 et 5000 Hz (18 bandes) (L_{eqf}) ;

La valeur médiane de la corrélation (Corr) maximale entre un segment et chaque segment correspondant à un tir dans les données d'apprentissage a également été calculée [3].

Cela donne un total de 34 paramètres pour chaque segment. Afin de réduire la charge de calcul et maximiser l'efficacité du processus de classification, il est préférable de limiter le nombre de paramètres.

Une première sélection a été réalisée en calculant le *Fisher's discriminant ratio*. Le FDR (7) donne une indication de la capacité de séparation de chaque paramètre [4]. Dans un premier temps, seuls les paramètres avec un FDR supérieur à 1 ont été sélectionnés. Cette liste a ensuite été affinée en fonction de l'algorithme de classification choisi.

$$FDR = \frac{\sum_{i=1}^k \sum_{j=1}^k (\bar{x}_i - \bar{x}_j)^2}{\sum_{i=1}^k \sigma^2(x_i)}, \quad (7)$$

k étant le nombre de classes.

2.4 Algorithmes de classification

Deux algorithmes de classification ont été testés. Dans les deux cas, l'étape d'apprentissage a été réalisée à partir des bruits détectés dans les Données A avec la méthode utilisant l'énergie quadratique moyenne. Cela correspond à 218 tirs et 51 autres bruits (cf. paragraphe 4.1).

Le premier algorithme utilise une méthode naïve bayésienne et la liste de paramètres a été optimisée à partir des résultats de classification (*wrapper method*). Dans cette méthode, un algorithme recherche séquentiellement la combinaison de paramètres qui minimise le taux de mauvaises classifications. Les paramètres ainsi retenus sont la corrélation et les niveaux continus équivalents dans les bandes 13 et 14. Dans l'étape de prédiction, la probabilité a posteriori qu'un segment appartienne à chaque classe (tir ou autre bruit) est calculée et le segment est classifié en fonction de la plus grande probabilité a posteriori.

Le second algorithme est un arbre de classification [2] basé sur le niveau continu équivalent dans la bande 11 (L_{eqf11}), les premier et quatrième *MFCC* et le point de "roll-off" (*RF*).

Ces deux modèles ont été établis avec la Statistics toolbox de Matlab.

3 Données expérimentales

La détection et la classification des bruits de tirs a été réalisée à partir de fichiers audio enregistrés par un sonomètre de classe 1 avec une fréquence d'échantillonnage de 48 kHz et codés sur 24 bits. Les données proviennent de deux campagnes de mesures différentes menées à proximité d'un club de tir. Toutes les mesures ont été réalisées au même endroit, situé approximativement face aux tireurs et à environ 700 m. La directivité des armes à feu étant maximale dans la direction du tir, cette position correspond à la situation la plus défavorable du point de vue de la gêne acoustique. Dans les deux cas, les conditions de vent étaient légèrement défavorables à la propagation du bruit.

Les données d'une des deux campagnes ont été divisées en deux ensembles de 30 minutes, l'un servant de données d'apprentissage (Données A), le deuxième de données de validation dans des conditions de mesure similaires (Données B). Le troisième ensemble (Données C) a permis de réaliser une validation dans des conditions de mesure différentes. Sa durée est de 20 minutes. La détection étant réalisée sur des fichiers audio de 5 minutes, les Données A et B comportent chacune 6 fichiers audio et les Données C, 4.

Parallèlement aux enregistrements sonores, les niveaux continus équivalents sur une période de 50 ms ($L_{Aeq,50ms}$) ont été mesurés. A partir de ces niveaux, une analyse manuelle a été réalisée pour chaque ensemble de données afin de localiser les détonations et calculer le niveau d'évaluation de la gêne. Les résultats de cette analyse sont utilisés comme référence lors des étapes de validation.

4 Résultats expérimentaux

4.1 Comparaison des deux méthodes de détection

Les deux méthodes de détection ont été appliquées à l'ensemble des données avec comme valeurs de seuils celles optimisées pour les données d'apprentissage. Les résultats sont donnés dans le Tableau 1.

TABLEAU 1 – Comparaison des méthodes de détection

| Données | Nombres de tirs réels | Vrais positifs | Faux positifs |
|-----------------------------|-----------------------|----------------|---------------|
| Energie quadratique moyenne | | | |
| A | 218 | 218 | 51 |
| B | 240 | 238 | 76 |
| C | 435 | 418 | 59 |
| Flux spectral | | | |
| A | 218 | 217 | 152 |
| B | 240 | 236 | 110 |
| C | 435 | 415 | 60 |

En terme de détection des bruits de tirs, les deux méthodes donnent de très bons résultats pour toutes les données, le taux de détections positives étant supérieur à 95%.

Par contre, avec les Données A et B, on obtient un nombre plus élevés de faux positifs avec la méthode utilisant le flux spectral. Ceux-ci proviennent essentiellement de fichiers audio ne comportant pas ou peu de détonations (Tableau 2). L'algorithme de localisation des pics de la fonction de détection normalise le signal par rapport à sa valeur quadratique moyenne. Dans le cas du flux spectral, cela a pour conséquence de faire ressortir d'avantage les bruits environnementaux en l'absence de bruits de tir. Un nombre significatif de faux positifs est également détecté dans ces fichiers avec l'autre méthode mais dans une moindre mesure.

En pratique, cet inconvénient peut être contourné en ne prenant pas en compte les résultats correspondant aux fichiers où aucun tir n'a été émis, cette information étant facilement accessible en plaçant un compteur de tirs à proximité des tireurs.

4.2 Classification

Les bruits impulsifs détectés par la méthode utilisant l'énergie quadratique moyenne ont été classifiés à l'aide des deux algorithmes de classification. Pour les données d'apprentissage (Données A), nous avons utilisé une validation croisée "leave-one-out", l'apprentissage se faisant

TABLEAU 2 – Comparaison des méthodes de détection

| Données | Nombres de tirs réels | Vrais positifs | Faux positifs |
|-----------------------------|-----------------------|----------------|---------------|
| Energie quadratique moyenne | | | |
| A (fichier5) | 5 | 5 | 8 |
| A (fichier6) | 0 | 0 | 19 |
| B (fichier1) | 0 | 0 | 39 |
| Flux spectral | | | |
| A (fichier5) | 5 | 5 | 23 |
| A (fichier6) | 0 | 0 | 84 |
| B (fichier1) | 0 | 0 | 72 |

sur $(n - 1)$ observations et la validation sur la n^{eme} . Les résultats sont présentés dans le Tableau 3.

On constate que les deux méthodes donnent des résultats similaires avec toutes les données de validation, avec une légère supériorité de l'arbre de classification. Une large majorité des faux positifs issus de la phase de détection sont ainsi éliminés. On remarque cependant qu'avec les Données C, un nombre plus important de tirs correctement détectés sont mal classifiés. Les conditions de mesure de ces données étant légèrement différentes de celles des deux autres ensembles, des données d'apprentissage plus représentatives des différentes conditions météorologiques rencontrées devraient conduire à une diminution du nombre de faux négatifs.

Le Tableau 3 donne également le taux de vraies détections positives (TPR) et le taux de fausses détections (proportion de fausses détections parmi les détections) du système global de détection/classification des tirs.

4.3 Estimation du niveau d'évaluation de la gêne

Afin d'analyser l'influence des performances du système d'identification de tirs sur le niveau d'évaluation de la gêne, celui-ci a été calculé à partir des résultats de classification présentés ci-dessus et comparé aux valeurs obtenues par une analyse manuelle (Tableau 4). On constate que les résultats obtenus avec l'arbre de classification donnent des niveaux très proches des valeurs de référence même avec les Données C dont le taux de vraies détections positives est pourtant plus faible. Cela est dû au fait que les niveaux d'exposition des tirs manquants sont en moyenne plus faibles que ceux de tirs correctement identifiés.

5 Conclusion

Dans cet article, deux méthodes de détection et de classification de bruits de tirs à partir d'enregistrements

TABLEAU 3 – Comparaison des méthodes de classification

| | Vrais positifs | Faux positifs | Taux global vraies détections positives | Taux global fausses détections |
|----------------------------|----------------|---------------|---|--------------------------------|
| Données A - 218 tirs réels | | | | |
| Détection | 218 | 51 | - | - |
| Bayésien | 217 | 3 | 99,5% | 1,4% |
| Arbre | 217 | 1 | 99,5% | 0,5% |
| Données B - 240 tirs réels | | | | |
| Détection | 238 | 76 | - | - |
| Bayésienne | 229 | 10 | 95,4% | 4,2% |
| Arbre | 236 | 6 | 98,3% | 2,5% |
| Données C - 435 tirs réels | | | | |
| Détection | 418 | 59 | - | - |
| Bayésienne | 367 | 5 | 84,4% | 1,3% |
| Arbre | 370 | 7 | 85,1% | 1,9% |

audio ont été comparées dans le contexte d'un système de monitoring du niveau d'évaluation de la gêne autour d'un club de tir. Les meilleurs résultats ont été obtenus avec un module de détection des potentiels bruits de tirs basé sur l'évolution temporelle de l'énergie quadratique moyenne et un arbre de classification utilisant les paramètres L_{eqf11} , $MFCC_1$, $MFCC_4$ et RF .

Le système d'identification de tirs proposé a été validé sur trois ensembles de données récoltées en un même endroit lors de deux campagnes de mesure différentes. Les résultats obtenus avec les données d'apprentissage (validation "leave-one-out") et les données de test enregistrées dans les mêmes conditions donnent d'excellents résultats. En revanche, le troisième ensemble de données (conditions météorologiques différentes) conduit à un taux de tirs mal identifiés plus important. Une amélioration de la sensibilité du système nécessiterait l'utilisation de données d'apprentissage représentatives des différentes conditions météorologiques rencontrées. Cependant, les tirs non identifiés présentant un niveau d'exposition moyen légèrement plus faible que les autres, le niveau de gêne obtenu est proche de la référence.

Dans un travail futur, le système sera validé avec des données provenant d'autres campagnes de mesure afin de confirmer ces résultats prometteurs.

TABLEAU 4 – Niveaux d'évaluation de la gêne

| Données | Classification | | Référence |
|---------|----------------|----------|-----------|
| | Bayésienne | Arbre | |
| A | 48,4 dBA | 48,4 dBA | 48,3 dBA |
| B | 45,4 dBA | 47,2 dBA | 47.4 dBA |
| C | 46,4 dBA | 46,7 dBA | 47,3 dBA |

Références

- [1] M. Alonso, G. Richard et B. David, Tempo And Beat Estimation Of Musical Signals, *Proc. ISMIR 2004*.
- [2] L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, *Classification and regression trees*, CRC press (1984).
- [3] A. Chacon-Rodriguez, P. Julian, L. Castro et P. Alvarado, Evaluation of gunshot detection algorithms, *IEE Transactions on circuits and systems*, Vol. 58 no. 2, 363-373 (2011).
- [4] A. Dufaux, Detection and recognition of impulsive sound signals. *Phd Thesis*, University of Neuchatel, Switzerland, 2001.
- [5] I. L. Freire et J. A. Apolinario, Gunshot detection in noisy environments, *Proc. The 7th International Telecommunications Symposium* (2010).
- [6] L. Gerosa, G. Valenzise, M. Tagliasacchi, F. Antonacci et A. Sarti, Scream and gunshot detection in noisy environments, *Proc. EUSIPCO 2007*.
- [7] ISO1996-1, *Acoustics - description, measurement and assessment of environmental noise - part 1 : Basic quantities and assessment procedures* (2003).
- [8] ISO1996-2, *Acoustics - description, measurement and assessment of environmental noise - part 2 : Determination of environmental noise levels* (2007).
- [9] O. Lartillot et P. Toivainen, A Matlab Toolbox for Musical Feature Extraction From Audio, *International Conference on Digital Audio Effects* (2007).
- [10] R. Maher et J. Studniarz, Automatic search and classification of sound sources in long-term surveillance recordings, *AES 46th International Conference* (2012).
- [11] A.-C. Witsel, Propagation, assessment and management of shooting noise, *Rapport interne* (sept. 2015).