

# CFA/VISHNO 2016

## Des imitations vocales à l'identification de sons de l'environnement

G. Lemaître, O. Houix, M. Misdariis et P. Susini

STMS lab (IRCAM, CNRS, UPMC), 1 place Igor Stravinsky, 75004 Paris, France  
guillaumejlemaître@gmail.com



LE MANS

## **CFA2016/416**

### **Des imitations vocales à l'identification de sons de l'environnement**

G. Lemaitre, O. Houix, M. Misdariis et P. Susini  
STMS lab (IRCAM, CNRS, UPMC), 1 place Igor Stravinsky, 75004 Paris, France  
guillaumejlemaitre@gmail.com

Il est plus intuitif et efficace de décrire un son en l'imitant avec la voix (ou des gestes) qu'en le décrivant avec des mots (Lemaitre et Rocchesso 2014). Ces résultats sont le point de départ d'un projet européen dont l'objectif est de permettre aux designers sonores d'utiliser directement la voix et les gestes pour le contrôle de la synthèse sonore, en créant ainsi des esquisses sonores (Rocchesso et al. 2015, skat-vg.eu). Les objectifs sont également d'étudier, à l'aide des imitations, les phénomènes perceptifs et cognitifs impliqués dans l'écoute des sons environnementaux. Dans ce cadre, le travail présenté ici a pour thème l'identification de sons de l'environnement. Il présente les résultats d'une expérience comparant l'identification de 32 sons environnementaux de référence avec des imitations vocales et "sketchs auditifs" de ces sons de référence. Les imitations vocales ont été réalisées par 10 personnes, et les "sketchs auditifs" par une méthode algorithmique adaptée de Suied et al. (2012). La méthode consistait à calculer une représentation temps-fréquence fondée sur un modèle auditif, et d'annuler les coefficients les moins importants, puis d'inverser la représentation. Le choix du nombre de coefficients a permis créer trois niveaux de dégradation: faible, moyen et fort. D'autre part, les sons de référence correspondaient à quatre profils d'énergie: impulsif, répété, stationnaire, complexe. La comparaison entre imitations et sketchs montre que dans la plupart des cas les performances d'identification des imitations vocales sont comparables à des dégradations faibles ou moyennes du signal sonore, et que les imitations sont bien reconnues. L'analyse détaillée des performances d'identification pour les différents profils permet alors de comprendre quelles sont les informations acoustiques utilisées identifier les sons. Ces résultats offrent des perspectives à la fois pour comprendre comment nous identifions les sons, et pour définir l'utilisation pertinente des vocalisations dans les interfaces homme-machine.