

Modèles de Markov cachés pour l'estimation de plusieurs fréquences fondamentales

Francis Bach*, Michael I. Jordan**

* Centre de Morphologie Mathématique
Ecole des Mines de Paris
35, rue Saint Honoré
77305 Fontainebleau, France
francis.bach@mines.org

** Computer Science Division
and Department of Statistics
University of California
Berkeley, CA 94720, USA
jordan@cs.berkeley.edu

1 Introduction

Le suivi de la fréquence fondamentale est un problème important du traitement de la parole et de la musique, et le développement d'algorithmes robustes pour la détermination d'une ou plusieurs fréquences fondamentales est un sujet actif de recherches en traitement du signal acoustique (Gold et Morgan, 1999). La plupart des algorithmes d'extraction de la fréquence fondamentale commencent par construire un ensemble de caractéristiques non linéaires (comme le corrélogramme ou le "cepstrum") qui ont un comportement spécial lorsqu'une voyelle est prononcée. Ensuite, ces algorithmes modélisent ce comportement afin d'extraire la fréquence fondamentale. En présence de plusieurs signaux mixés additivement, il est naturel de vouloir modéliser directement le signal ou une représentation linéaire de ce signal (comme le spectrogramme), afin de préserver l'additivité et de rendre possible l'utilisation de modèles destinés à une seule fréquence fondamentale pour en extraire plusieurs.

L'utilisation directe du spectrogramme nécessite cependant un modèle probabiliste détaillé afin de caractériser la fréquence fondamentale. Dans cet article, nous considérons une variante de modèle de Markov caché et utilisons le cadre des modèles graphiques afin de construire le modèle, apprendre les paramètres à partir de données et développer des algorithmes efficaces d'inférence. En particulier, nous utilisons des développements récents en apprentissage automatique (machine learning) pour caractériser les propriétés adéquates des signaux de parole et de musique; nous utilisons des probabilités *a priori* non-paramétriques afin de caractériser la régularité de l'enveloppe spectrale et nous améliorons la procédure d'apprentissage grâce à l'apprentissage discriminatif du modèle.