

ABSTRACT HDR Monsieur SERIZEL Romain

Nom – Prénom	Serizel Romain
Laboratoire de rattachement	Loria
Intitulé du diplôme HDR	Informatique
Titre de l'HDR	Contributions to speech processing and ambient sound analysis

Abstract (français) – maximum 15 lignes

Nous sommes constamment entourés de sons que nous exploitons pour adapter nos actions aux situations auxquelles nous sommes confrontés. Certains sons comme la parole peuvent avoir une structure particulière à partir de laquelle nous pouvons déduire des informations, explicites ou non. Au cours de la décennie écoulée, des progrès significatifs ont été réalisés dans le domaine du traitement de la parole et du son. Grâce à ces progrès, la parole est devenue un élément central de nombreux outils de communication à distance d'humain à humain ainsi que dans les systèmes de communication humain-machine. Dans les scénarios qui impliquent la présence de perturbations acoustiques telles que du bruit ou de la réverbération les performances peuvent avoir tendance à se dégrader gravement. Dans cette HDR, nous nous concentrons sur le traitement de la parole et de son environnement d'un point de vue audio. Les algorithmes proposés reposent sur une variété de solutions allant des approches basées sur le traitement du signal aux solutions orientées données. Nous proposons des solutions à des problèmes allant de la reconnaissance vocale au rehaussement de la parole ou à l'analyse des sons ambiants. L'objectif est d'offrir un panorama des différents aspects qui pourraient être améliorés par un algorithme de traitement de la parole fonctionnant dans un environnement réel. Nous commençons par décrire la reconnaissance automatique de la parole comme une application finale potentielle et analysons progressivement les limites et les solutions proposées aboutissant à l'analyse plus générale des sons ambiants.

Abstract (anglais) – maximum 15 lignes (pas obligatoire)

We are constantly surrounded by sounds that we continuously exploit to adapt our actions to situations we are facing. Some of the sounds like speech can have a particular structure from which we can infer some information, explicit or not. Within the last decade, there has been significant progress in the domain of speech and audio processing and in particular in the domain of machine learning applied to speech and audio processing. Thanks to these progresses, speech has become a central element in many human to human distant communication tools as well as in human to machine communication systems. These solutions work pretty well on clean speech or under controlled condition. However, in scenarios that involve the presence of acoustic perturbation such as noise or reverberation systems performance tends to degrade severely. In this thesis we focus on processing speech and its environments from an audio perspective. The algorithms proposed here are relying on a variety of solutions from signal processing based approaches to data-driven solutions based on supervised matrix factorization or deep neural networks. We propose solutions to problems ranging from speech recognition, to speech enhancement or ambient sound analysis. The target is to offer a panorama of the different aspects that could improve a speech processing algorithm working in a real environments. We start by describing automatic speech recognition as a potential end application and progressively unravel the limitations and the proposed solutions ending-up to the more general ambient sound analysis.