La reconnaissance vocale

Les systèmes de reconnaissance vocale se développent et sont améliorés de plus en plus rapidement, et sont même directement intégrés aux ordinateurs. Je trouve ces systèmes fascinants, j'ai donc voulu comprendre leur fonctionnement afin d'en créer un.
De plus, pour distinguer différents phonèmes, le choix de la représentation du signal est important afin de limiter les contraintes extérieures telles que la voix de chaque individu ou encore l'environnement sonore.
Enfin, les algorithmes actuels sont également basés sur des probabilités (comme les chaînes de Markov), ce qui fait intervenir une certaine partie de hasard.

Positionnement thématique (étape 1)

INFORMATIQUE (Informatique pratique), PHYSIQUE (Physique Ondulatoire), MATHEMATIQUES (Autres).

Mots-clés (étape 1)

Mots-Clés (en français)	**Mots-Clés** (en anglais)
Spectre	*Spectrum*
Algorithme	*Algorithm*
Probabilité	*Probability*
Transformée de Fourier	*Fourier transform*
Coefficients cepstraux	*Cepstral coefficients*

Bibliographie commentée

La reconnaissance de la parole est une technique permettant la retranscription de phrases parlées par une machine grâce à des algorithmes adaptés. On observe leur développement de plus en plus rapide depuis les années 1950 [1], jusqu'à faire partie de notre quotidien.

Tout d'abord, depuis la commercialisation du premier logiciel en 1972 qui permettait de transcrire une centaine de mots, puis de la reconnaissance vocale en continue en 1997 par IBM, les progrès techniques et informatiques réalisés permettent aujourd'hui d'avoir des systèmes capables de transcrire des textes entiers avec une marge d'erreur de plus en plus réduite [2]. Cette évolution rapide nécessite cependant de faire des choix concernant la modélisation du signal audio. La représentation du signal utilisée a des conséquences immédiates sur le nombre d'utilisateurs du système, la complexité des algorithmes, le temps de calcul ou encore la mémoire utilisée. En effet, il existe deux modèles principaux : le système mono-locuteur, qui est le plus répandu et qui nécessite une phase d'apprentissage par l'ordinateur, ou multi-locuteur [3]. Dans la suite de ce projet, le modèle étudié sera restreint à un seul utilisateur.

Ensuite, le processus de reconnaissance vocale se déroule en plusieurs étapes. D'abord, il faut recueillir le signal et le numériser. Puis le signal est cadré en trames courtes afin de pouvoir en extraire les caractéristiques principales [4]. Il existe plusieurs méthodes d'extraction, telles que le codage prédictif linéaire (LPC) qui repose sur l'hypothèse selon laquelle la parole peut être

modélisée par un processus linéaire [5], ou encore l'utilisation des coefficients cepstraux (MFCC) [6] qui est la méthode la plus courante et qui se base sur une échelle fréquentielle spécifique mieux adaptée à l'ouïe humaine [7], c'est pourquoi il est nécessaire de choisir et d'adapter la méthode en fonction du résultat souhaité et de l'environnement sonore. Dès lors, nous nous intéresserons à cette dernière méthode par la suite.

Cependant, il est également nécessaire lors de la conception de l'algorithme de choisir entre une technique globale (basée sur des mots entiers, plus facile mais limitée par la mémoire du système), ou analytique (basée sur des phonèmes, plus difficile mais plus efficace) [8]. Ainsi, pour ce projet, la méthode globale a été retenue pour simplifier l'algorithme et puisque la base de données ne contiendra que quelques mots, ce qui ne nécessitera pas de temps de calcul trop important.

Enfin, pour améliorer ce processus de reconnaissance vocale, la plupart des systèmes actuels sont complétés par des algorithmes de prédiction de mots, basés sur des probabilités, notamment des chaînes de Markov [9]. Par conséquent, le dernier aspect de notre démarche consiste à développer un algorithme reposant sur des probabilités, de manière à reproduire à une échelle réduite le processus complet de reconnaissance vocale.

Problématique retenue

Comment élaborer, à partir d'un signal sonore analogique, un processus algorithmique basé sur la méthode d'extraction des coefficients cepstraux (MFCC) permettant d'aboutir à la reconnaissance d'un mot ?

Objectifs du TIPE

L'objectif de ce TIPE est, tout d'abord, de déterminer une représentation adaptée du signal pour distinguer des mots enregistrés. Puis, il s'agit d'élaborer un algorithme, écrit dans le langage de programmation Python, permettant de les différencier et de les reconnaître parmi une banque de mots qui aura été créée au préalable. Enfin, à l'aide de probabilités, le but est de développer un algorithme basé sur des chaînes de Markov afin de prédire le mot suivant en fonction de celui qui aura été reconnu précédemment.

Abstract

Speech recognition is increasingly used with the development of mobiles applications. I took a special interest in an algorithm involving MFC (Mel Frequency Cepstrum). This process is based on Fourier Transform and let a better representation of humans sound. The coefficents given by the PFS is compared to a simple database in order to determine the closest word. Finally, I added a predictive model, based on Markov chains, in order to improve the detection of spoken words.

Références bibliographiques

[1] ? : Wikipédia : *https://fr.wikipedia.org/wiki/Reconnaissance_automatique_de_la_parole , dernière mise à jour le 29 novembre 2016 , consultée le 12 février 2017,*

[2] RODOLPHE BATTAULT : Examen Probatoire pour l'obtention du Diplôme d'Ingénieur du

C.N.A.M : *http://r.battault.free.fr/probatoire/probatoire.html, dernière mise à jour en juin 1998, consultée le 23 janvier 2017,*

[3] LUIZA OROSANU : Thèse soutenue le 11 Décembre 2015 : *Reconnaissance de la parole pour l'aide à la communication pour les sourds et malentendants, Université de Lorraine*

[4] DAN JURAFSKY : web.stanford.edu, Spoken Langage Processing : *,http://web.stanford.edu/class/cs224s/lec/224s.14.lec6.pdf, consultée le 20 septembre 2016*

[5] ABDENOUR HACINE-GHARBI : Thèse soutenue le 09 décembre 2012 : *Sélection de paramètres acoustiques pertinents pour la reconnaissance de la parole, Université d'Orléans*

[6] STEVEN B. DAVIS : Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences : *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-28, NO. 4, AUGUST 1980*

[7] JAMES LYONS : Pratical Cryptography : *http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/, dernière mise à jour en 2012, consultée le 27 septembre 2016*

[8] PHILIPPE FOUCHER : Traitement du signal pour la reconnaissance vocale : *http://www2.univ-paris8.fr/ingenierie-cognition/master-handi/etudiant/cours/informatique/cours_foucher/traitement_signal_05_06/rdf5.pdf, consultée le 17 novembre 2016*

[9] VINCENT ARSIGNY : Modélisation par champ de Markov du signal de parole et application à la reconnaissance vocale : *http://www-sop.inria.fr/asclepios/Publications/Arsigny/ArsignyRapportPolytechnique.pdf, dernière mise à jour en 2000, consultée le 16 décembre 2016.*

[10] VINCENT ARSIGNY : Modélisation par champ de Markov du signal de parole et application à la reconnaissance vocale : *http://www-sop.inria.fr/asclepios/Publications/Arsigny/ArsignyRapportPolytechnique.pdf, dernière mise à jour en 2000, consultée le 16 décembre 2016.*