

Reconnaissance vocale

Partie expérimentale

Document 1 : reconnaissance vocale et applications



Logo de reconnaissance vocale souvent rencontré

La reconnaissance vocale est une technologie relativement récente, qui séduit énormément les professionnels et le grand public par son aspect pratique, ludique et ergonomique.

Les applications de cette technologie sont nombreuses : navigation sur un serveur vocal au téléphone, apprentissage d'une langue étrangère, commandes vocales dans les voitures, les téléphones ou bien encore dans les salles d'opérations chirurgicales, dictée vocale (écriture en directe d'un texte dicté), identification vocale dans les zones sécurisées ou bien dans le cadre d'une enquête judiciaire, etc.

Le vieux rêve de la communication avec les machines en utilisant simplement la parole n'a jamais été aussi près d'être atteint.

Document 2 : spectrogramme avec Audacity

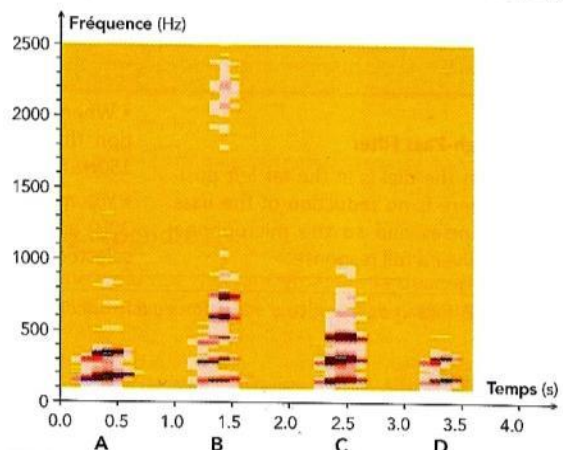
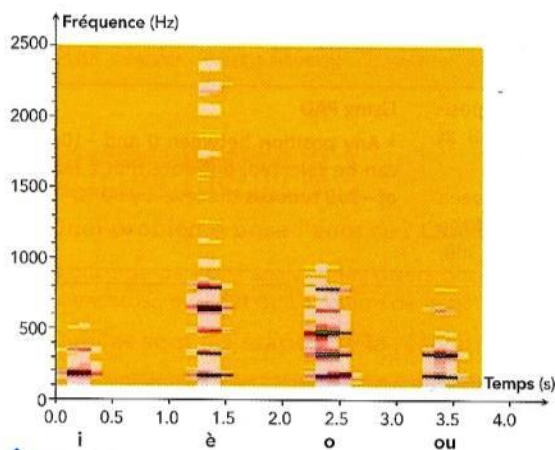
Enregistrer un son sous Audacity ou ouvrir un fichier son déjà existant.

Visualiser le signal sonore.

Visualiser le spectre sonore d'une partie ou de tout l'enregistrement.

Visualiser le spectrogramme du son en utilisant le menu déroulant, à gauche du signal sonore obtenu, en cliquant sur la petite flèche noire. C'est le document le plus complet pour l'analyse d'un son.

- 1) Que représente et comment est construit un spectrogramme ? Pourquoi est-ce le document nécessaire pour la reconnaissance sonore d'une phrase complète (vis-à-vis du spectre sonore) ?
- 2) En vous aidant d'un enregistrement, montrer que les 5 voyelles peuvent être reconnues assez facilement par un ordinateur analysant le spectrogramme d'un enregistrement.
- 3) Attribuer aux lettres A, B C et D du document suivant les mots (ou phonèmes) « lit », « lait », « loup » et « lot » en vous aidant du document a en expliquant.



Etude de documents

Après avoir lu les documents suivants, et en utilisant vos connaissances personnelles et la partie expérimentale, répondre aux questions suivantes

- 1) Quels sont les principes des méthodes employées pour la reconnaissance vocale ? Quels sont les avantages et inconvénients de chacune d'elles ? Pour quelles utilités bien définies va-t-on se servir plutôt de l'une que de l'autre ?

2) Pourquoi la question préliminaire 3) de la partie expérimentale est-elle un cas d'école trop particulier ? Quelles limites rencontre la reconnaissance vocale aujourd'hui au niveau phonétique (« la citerne est pleine d'eau », « la scie terne est pleine d'os ») et au niveau d'identification ?

Document 3 : les deux techniques de reconnaissance vocale

Comment ça marche ?

« Du jour où le téléphone a été inventé, des esprits ingénieux ont relevé que le micro constituait un détecteur de son susceptible d'envoyer un signal électrique au moindre bruit. D'où l'idée d'envoyer ce signal à une machine pour qu'elle obéisse à la voix. [...] Pour que la reconnaissance vocale devienne réalité, il fallait donc un outil capable d'analyser les sons pour séparer le bruit de fond d'un atelier et les ordres du machiniste. [...] Les progrès de la reconnaissance vocale sont donc intimement liés à l'évolution des ordinateurs. En 1971 apparaît le premier dispositif de commande vocale, le *Voice Command System*, fondé sur une calculatrice capable, après un cycle d'apprentissage, de reconnaître 24 ordres. Le procédé va ensuite se scinder en deux grandes familles.

- La première est une simple commande vocale [...]. L'appareil doit être capable de reconnaître l'énoncé d'un nom et, pour cela, il faut répéter plusieurs fois le nom choisi qui sera alors mémorisé sous forme d'une séquence sonore. Une fois que le nom aura été prononcé de trois à cinq fois, l'électronique de l'appareil analyse l'évolution de l'intensité sur plusieurs plages de fréquences (en général huit) pour chacun de ces échantillons vocaux. [...]

Après avoir réalisé une moyenne des huit échantillons, l'appareil en déduit un profil acoustique numérisé du nom et le met en mémoire. En usage normal, pour reconnaître un nom prononcé par l'utilisateur, il compare le nouveau profil à ceux qu'il possède en mémoire. Il attribue alors des notes statistiques de ressemblance et décrète que le nom prononcé est celui qui a obtenu la meilleure note.

- La seconde famille, celle des systèmes de dictée [...], nécessite un ordinateur puissant car leur principe de reconnaissance vocale est infiniment plus complexe : ils doivent être capables de reconnaître des mots prononcés en langage dit naturel, c'est-à-dire au sein de phrases où – notamment en français – les mots s'enchaînent. Qui plus est, ils doivent tenir compte des liaisons, et aussi des homophones qui se créent spontanément par la juxtaposition de mots. [...]

Après amplification et tri par fréquences, grâce à un jeu de filtres électroniques rappelant les « égaliseurs » des chaînes haute-fidélité, un spectrogramme de la phrase est obtenu.

Pour l'ordinateur, la première tâche consiste à séparer chaque phonème. Il le transforme alors en un fichier numérique : un tableau de données.

Ensuite, il compare ces tableaux obtenus à ceux que contient un dictionnaire, stocké sur le disque dur, où sont associés phonèmes ou

groupes de phonèmes et mots réels. Mais, par la présence des liaisons ou par homophonie, de très nombreux mots peuvent correspondre au message parlé. Interviennent alors des logiciels d'analyse contextuelle et sémantique. Souvent, l'ordinateur devra attendre qu'une grande partie de la phrase ait été prononcée pour que ces logiciels puissent commencer leur travail. L'apparition d'un nouveau mot peut bouleverser son sens et donc le choix de coupe des phonèmes. Par déduction statistique, les logiciels retiennent la solution la plus probable qui se révèle, le plus souvent, être la bonne.

Beaucoup de systèmes actuels sont encore monolocuteurs, c'est-à-dire qu'ils ne peuvent « comprendre » qu'une seule personne, et encore après une phase d'apprentissage. On cherche donc maintenant à faire des systèmes multilocuteurs qui « entendent » aussi bien le langage d'un homme à l'accent parisien que celui d'une femme au parler méditerranéen, et cela sans phase d'apprentissage. »

Extrait de G. MARTIN, « La reconnaissance vocale. Comment la parole est venue aux machines ? », *La Recherche*, n° 367, sept. 2003.

Document 4 : les difficultés à l'identification

Identifier un individu par sa voix est une tâche complexe. La voix dépend de paramètres physiologiques, mais aussi de l'humeur ou de l'état de santé de chaque individu. En outre, une voix peut être volontairement déformée, elle peut aussi être imitée.

Les conditions d'enregistrements sont également très importantes (qualité du matériel, bruit ambiant, etc.).

Cependant, l'expérience quotidienne montre qu'une voix (d'un proche, d'une personnalité publique) peut être reconnue.

La comparaison avec des enregistrements de référence permet de donner la probabilité (et non l'assurance) qu'une personne soit l'auteur d'un enregistrement vocal.

La reconnaissance vocale est utilisée dans la police scientifique et technique ou pour le contrôle de l'accès à certaines zones ou matériels.

