

LA RECONNAISSANCE VOCALE EST MORTE : PET À SON ÂME

LE 10 NOVEMBRE 2010 DR GOULU

La reconnaissance vocale était très prometteuse à la fin des années 90 mais le taux d'erreur de mots plafonne à 10% depuis 2001, soit environ le triple du taux d'erreur d'un être humain.

D'après "2001 l'Odyssée de l'Espace", nos ordinateurs devraient comprendre notre voix depuis 9 ans. Depuis 1997, on trouve des logiciels de reconnaissance vocale pour PC, et depuis peu nos téléphones disposent de cette fonction. Mais on ne l'utilise pas. Je ne connais personne qui dicte ses e-mails, et vous ?

Comme tous les geeks j'ai essayé de temps en temps, parfois passé une heure à lire des textes d'apprentissage de la voix la plus monotone possible à la nouvelle version d'un soft, et puis abandonné devant ses piètres performances. Ça ne marche pas, ou pas assez bien.

Robert Portner analyse cet échec dans "**Rest in Peas: The Unrecognized Death of Speech Recognition**", titre subtilement traduit en français dans le présent article.

Le problème, c'est qu'après une phase de progrès rapides à la fin du siècle passé, le **taux d'erreur de mots** plafonne à 10% depuis 2001, soit environ le triple du taux d'erreur d'un être humain. Et encore, c'est pour l'anglais "standard". Le taux d'erreur est bien plus élevé pour d'autres langues, et catastrophique pour une conversation entre supporters de foot à la sortie du match. (Voir **l'évolution du taux d'erreur de mots du National Institute of Standards and Technology**).



Pourtant dans les années 1990, des systèmes très fiables avaient été mis au point pour

distinguer quelques mots bien choisis dans des cockpits d'avion ou des chiffres au téléphone, et on s'était légitimement attendus à ce que la Loi de Moore permette de traiter rapidement le langage naturel.

Et effectivement, aujourd'hui on sait bien reconnaître des mots isolés. On sait à peu près éliminer les absurdités non conformes à la grammaire dans des phrases simples comme "le chat ment je la sous rit." Mais pour distinguer entre "le chat mange la souris", "le chat mange là, sous l'riz" et "le chaman gela, sourit" et , il faut comprendre le sens de la phrase, voire le contexte dans lequel elle est prononcée...

Si l'ordinateur doit connaître la différence entre un quadrupède carnivore et le trou d'une aiguille pour traiter une phrase triviale, on imagine que ce n'est pas demain qu'on dictera des contrats* ou des rapports à une machine. De gros projets ont été lancés par des poids lourds de l'informatique pour tenter de modéliser la connaissance humaine.

Par exemple le projet MindNet de Microsoft¹ a analysé des millions de pages de textes existants pour construire un graphe sémantique gigantesque, duquel il ressort effectivement que dans une phrase comportant "chat" et "souris", **le plus probable est que le chat chasse la souris**. Un tel graphe peut certainement être utile en **traduction automatique** car on dispose d'un texte de départ, mais pour la reconnaissance vocale il faudrait étendre le graphe à la structure des phrases utilisées en conversation courante, qui peut être bien distincte du langage écrit. Et pour faire ça automatiquement, il faudrait la reconnaissance vocale...

Comme le note Portner, on pensait au début que la reconnaissance vocale était un premiers pas vers l'intelligence artificielle. Aujourd'hui de nombreux chercheurs estiment que l'intelligence artificielle est indispensable pour atteindre une reconnaissance vocale de qualité acceptable². Les gros projets de recherche ont été abandonnés les uns après les autres, bloqués devant le mur si bien décrit par les **Perlisismes** sur l'intelligence artificielle comme :



"Une année de travail sur l'intelligence artificielle est suffisante pour vous faire croire en Dieu"



Le nombre de recherches sur "reconnaissance vocale" ou "**Dragon Naturally Speaking**" sur Google baisse régulièrement depuis 2001. Comme aucune idée fondamentalement nouvelle ne vient relancer la recherche, la reconnaissance vocale est morte, en toute discrétion.

Note : Ça me rappelle l'histoire de la secrétaire d'un célèbre ingénieur de la génération dictaphone qui avait commandé "310 mètres d'isolation entre 2 étages" au lieu de "3 centimètres". Ça c'est avec les 2% d'erreurs de transcription humaines...

>> Article publié initialement sur le blog de **Dr Goulu**

>> Illustrations Flickr CC : **TruShu** et Loguy en CC pour OWNsciences

1. Microsoft Research : **MindNet** [↔]

2. Janet M. Baker et al. "**Research Developments and Directions in Speech Recognition and Understanding**", IEEE Signal Processing Magazine [75] MAY 2009 [↔]

REDGUFF

le 11 novembre 2010 - 2:09 • SIGNALER UN ABUS - PERMALINK



Bonjour.

*En espéranto, le taux d'erreur est de 0% : c'est une langue phonétique !
C'est en plus très facile à apprendre.*

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

MACROPHONE



le 11 novembre 2010 - 14:20 • SIGNALER UN ABUS - PERMALINK



*A (re)lire : le Guérisseur de cathédrales, Philip K. Dick.
Ecrit en 1969, dans ce roman de SF, des personnages s'adonnent au "jeu" de retrouver le sens original de phrases passées plusieurs fois dans des traducteurs automatiques.*

VOUS AIMEZ



VOUS N'AIMEZ PAS



LUI RÉPONDRE

DR. GOULU

le 12 novembre 2010 - 10:38 • SIGNALER UN ABUS - PERMALINK



@macrophone : avec les outils de traduction de Google on peut guérir des cathédrales soi même... Exemple : <http://drgoulu.com/2009/05/14/tels-que-le-corbeau-le-renard/>

@redguff : j'ai trouvé http://www.formation-radio.com/form/flash_info_e2.htm cité sous http://fr.wikipedia.org/wiki/Reconnaissance_vocale#Performance qui montre un exemple de reconnaissance ou l'espéranto permet 0% d'erreur, effectivement. Apprendre une nouvelle langue pour mieux parler aux machines ? Pourquoi pas... J'avais un ami qui essayait de "parler modem" à l'époque des coupleurs acoustiques...

"Dans la symbiose homme-machine, c'est l'homme qui doit s'adapter parce que la machine ne peut pas." (Alan Perlis)

VOUS AIMEZ



VOUS N'AIMEZ PAS



LUI RÉPONDRE

JEAN-PAUL ROUZÉ

le 26 novembre 2010 - 9:19 • SIGNALER UN ABUS - PERMALINK



J'ai testé personnellement ou professionnellement depuis pas mal d'années de nombreux softs de reconnaissance vocale. J'en ai même recommandé il y a quelques mois l'achat dans le cas de l'aménagement du poste de travail d'une personne handicapée.

De ces expérimentations me viennent quelques réflexions :

** Pour le principal obstacle à l'utilisation vient du mode vocale qui... manque de confidentialité :-). En effet vous n'avez pas obligatoirement envie que vos voisins entendent : "j'ai une réunion à 11H00 avec ce Connard de X, merci de m'appeler à 11H10 pour une conférence café urgente". De plus imaginez la cacophonie si tous le monde dictait...*

** C'est Dragon, le plus courant et presque le seul, que j'ai approfondi un maximum. Dans l'utilisation de ce logiciel au moins deux points à observer 1°) parler de façon naturelle en liant les mots surtout pas : "Je-ne-suis-pas-un-robot" 2°) être patient ce n'est qu'au bout de quelque temps, après avoir suffisamment "appris" de votre dictée que le logiciel se montre sous son meilleur jour (pensez à bien corriger vos dictées au début, c'est fondamental).*

A noter que je suis en train d'utiliser un clavier, pourquoi ?

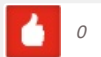
** Je suis sous Linux*

** Si je trouvais une version Linux, les bons softs coutent beaucoup trop cher pour moi.*

** Je suis dans une chambre de clinique et mon voisin de chambre...*

Je regrette beaucoup ces obstacles car en tant que dyspraxique j'ai mon interface digitale (mes doigts) qui ont du mal à se coordonner avec mon CPU (la masse de gélatine contenue dans ma boîte crânienne) Par contre par de problème avec le vocal :-)

VOUS AIMEZ



VOUS N'AIMEZ PAS



LUI RÉPONDRE

JICE - SITES SCIENTIFIQUES

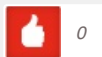
le 29 novembre 2010 - 14:32 • SIGNALER UN ABUS - PERMALINK



@JP Rouzé : il est vrai que je me suis aussi un peu penché vers ce genre de problématique (sensibilisé moi même par les patients de ma mère orthophoniste). Le commun des mortels ne pensent pas à la difficultés que peuvent rencontrer des mal voyants ou des gens paralysés des mains par exemple à accéder à Internet. Je parle d'internet car c'est aujourd'hui un puit de connaissance quasiment incontournable. S'en voir mis à l'écart est stigmatisant.

J'avais commencé à un écrire un blog uniquement sous Dragon, mais j'avais finalement mis le projet de côté après quelques articles.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE