

Projet Common Voice : pour que la voix soit libre

On nous demande souvent comment contribuer au Libre sans être un cador en informatique. Voici un projet utile, d'une grande importance et auquel il est très simple de contribuer : il suffit de savoir lire, parler et écouter. On vous explique tout.

On voit émerger à grande allure des objets avec lesquels l'interaction repose sur la reconnaissance vocale : smartphones, assistants connectés, dispositifs de réalité virtuelle... Selon les experts de ce marché, 50 % des recherches toutes plateformes confondues passeront par la voix d'ici 2 à 3 ans. Quant aux objets dits « intelligents », ils atteignent pour les pays favorisés un niveau de prix qui les rend accessibles à un nombre croissant de consommateurs. On peut trouver dès aujourd'hui dans la grande distribution des « enceintes connectées » à l'écoute de vos questions pour moins de 100 euros... Autant dire que ces produits sont en passe d'être des objets de consommation de masse.



Les services vocaux, au-delà des fonctions d'espionnage qui suscitent la méfiance, apporteront des avantages sensibles aux usages numériques du grand public. Ils

abaisseront la barrière d'accès à des fonctionnalités utiles pour les personnes handicapées, en difficulté avec la lecture, dont les mains sont occupées ou pour celles qui ont besoin d'assistance immédiate. Dans bien des cas de figure il est ou sera plus efficace ou rapide d'utiliser la voix plutôt qu'une interface tactile ou souris/clavier¹



Le problème hélas a un air de déjà-vu : aujourd'hui les systèmes de reconnaissance vocale sont essentiellement propriétaires et reposent sur 4 ou 5 bases de données vocales propriétaires : Cortana (Microsoft), Siri (Apple), Google Now (Google), Vocapia Research (VoxSigma suite)... En d'autres termes, tout est prêt pour assurer à quelques géants du numérique, toujours les mêmes, une suprématie commerciale et technologique. Et l'histoire récente prouve qu'ils n'hésiteront pas longtemps avant de capter les données les plus précieuses, celles de notre vie dans la bulle privée de notre habitation.

Il se trouve qu'un projet qui repose sur des ressources libres (données et code informatique) a été lancé par l'un des acteurs majeurs du monde du Libre : la fondation Mozilla.

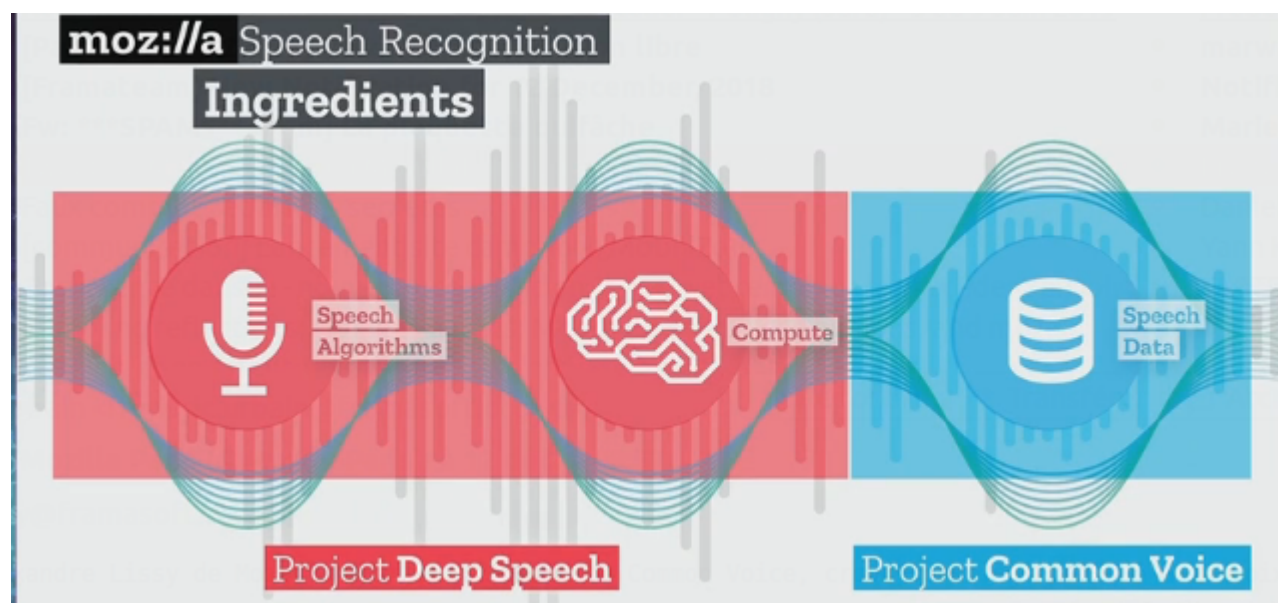
Pourquoi Mozilla s'en mêle ?

Parmi les principes qui guident Mozilla et qu'on retrouve dans son manifeste, la santé d'un Web ouvert et l'inclusivité sont des valeurs essentielles. Cette ressource numérique dont l'usage est appelé à se développer rapidement doit être à la disposition du plus grand nombre, à commencer par les entreprises innovantes (déjà sur la brèche par exemple Mycroft et Snips) qui n'ont pas les moyens financiers d'accéder aux bases propriétaires et qui seraient tout simplement marginalisées par les grandes entreprises. Au-delà, bien sûr, c'est

pour que des produits reposant sur la reconnaissance vocale soient accessibles à tous, quelle que soit leur langue, leur genre, leur accent local etc.

De quoi s'agit-il ?

De constituer **la plus riche base possible d'échantillons sonores** qui seront mis à la disposition des développeurs sous une licence libre (licence CC0). Le projet global s'appelle *Deep Speech* et Mozilla fait travailler des ingénieurs à traiter les données collectées avec des algorithmes, et ainsi alimenter un dispositif d'apprentissage machine.



Comment ça peut marcher ?

Ici nous tentons une description simple donc forcément approximative...

Un réseau de neurones va « apprendre » si...

1. On donne à la machine des fichiers audio en *entrée*
2. On calcule la *sortie*, c'est-à-dire le texte
3. On compare au texte d'origine et... ben non c'est pas tout à fait ça.
4. On ajuste *un petit peu* des millions de paramètres internes pour essayer de se rapprocher de la sortie voulue
5. On répète sur des milliers d'heures...



Alexandre Lissy, ingénieur Mozilla qui travaille au bureau de Paris pour le projet Deep Speech. Les autres membres de l'équipe sont à Berlin, au Brésil et à San Francisco... (Photo de Samuel Nohra publiée dans Ouest France)

Pourquoi est-ce difficile à réaliser ?

- L'entraînement des machines et la transcription nécessitent une grosse puissance de calcul.
- Un nombre très important d'heures d'enregistrements valides est nécessaire pour que la reconnaissance vocale soit la plus efficace possible. C'est une somme d'environ 10 000 heures qui est considérée comme souhaitable pour obtenir un résultat.
- Il existe peu de *gros* jeux de données publiquement accessibles en CC0 pour construire des modèles de reconnaissance 100% libres.

Les principes de Common Voice

Tout d'abord le projet est **mondial** et vise à fonctionner pour le plus grand nombre de langues possible. Le projet est assez récent et pour l'instant, 16 langues seulement sont actives dont bien sûr le Français. On remarque que le projet a de l'importance pour les langues qui peuvent se sentir menacées : le Catalan, le Breton et le Kabyle par exemple sont déjà lancés !



Mais euh... On n'en est que là nous autres les francophones ? Vous avez compris : il est temps de nous y mettre tous ! (copie d'écran de la page des langues)

C'est aussi un projet **inclusif** pour lequel les intonations diverses sont bienvenues, avec une insistance particulière des ingénieurs responsables du projet pour qu'il y ait une grande diversité de voix : locuteurs et locutrices, de tous âges, avec tous les accents régionaux (oui les accents du sud, du nord etc. sont tout à fait bienvenus, une trentaine d'accents sont retenus), car les machines devront traiter la voix pour le plus grand nombre et pas exclusivement pour une prononciation standard appliquée. il est important de prononcer distinctement, mais le projet n'a pas besoin de textes déclamés professionnellement par des acteurs.

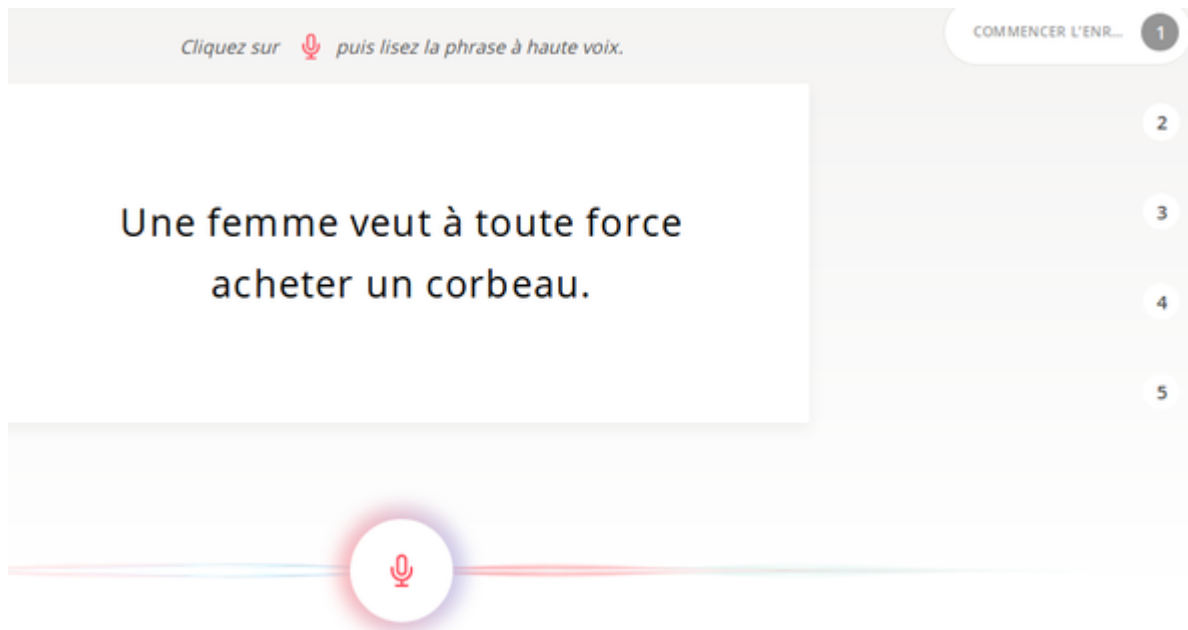
C'est surtout un projet **communautaire** et collaboratif. Il s'est construit dès le départ avec la communauté Mozilla et ses contributeurs et contributrices. Il fait maintenant appel à la communauté plus large de... tous les francophones. Car comme vous allez le voir, tout le monde ² peut y participer !

Simplicité de la contribution

Le mode d'emploi détaillé figure dans cet article de blog de Mozilla francophone. Mais pour l'essentiel, avec un micro-casque de qualité ordinaire vous pouvez :

1. PARLER

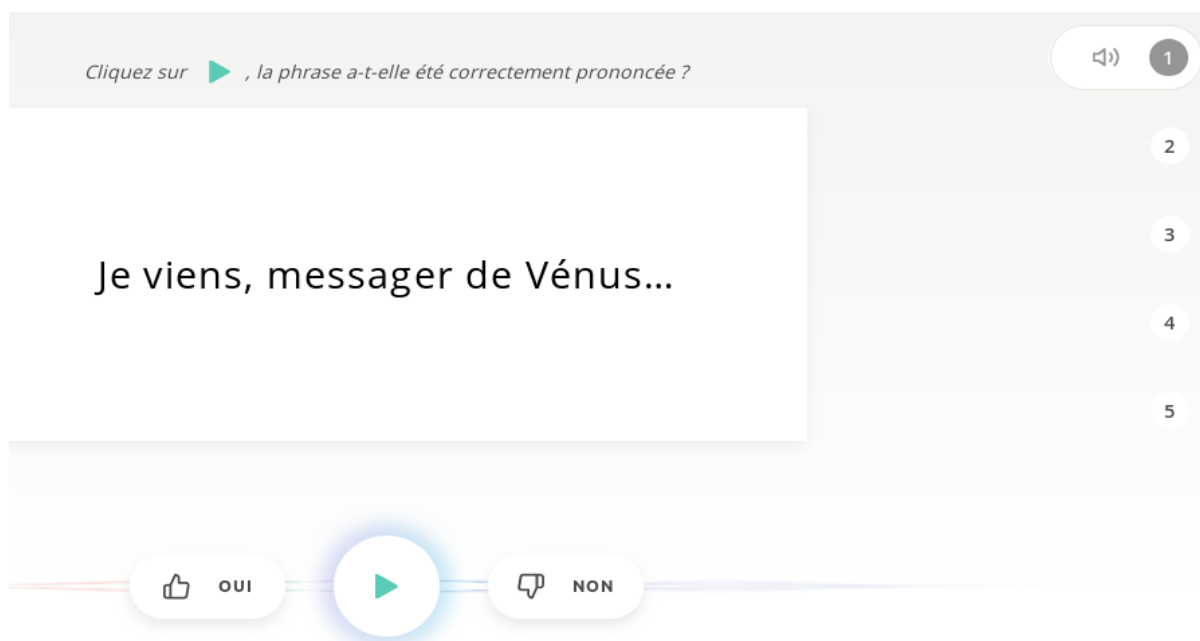
et enregistrer tour à tour une série de 5 brèves phrases



Pas d'inquiétude : on peut choisir « passer » avec un bouton dédié si on rencontre une difficulté de compréhension, prononciation ou autre. Et bien entendu si en se réécoutant on constate que ça peut être mieux, un simple clic permet de recommencer.

2. ÉCOUTER

pour valider ou non une série de 5 phrases



Vous vous demandez peut-être d'où viennent tous ces textes parfois bizarres

qu'on doit lire à haute voix ou écouter. Eh bien ce sont des textes dont la licence permet l'utilisation, qui viennent de transcriptions de débats de l'assemblée Nationale, de quelques livres du projet Gutenberg, de quelques pièces de théâtre, d'adresses françaises... Ah tiens, mais s'il faut enrichir la base, les romans de Pouhiou et celui de Frédéric Urbain publiés chez Framabook sont en CC0 ! Ça pourrait donner des phrases de test assez rigolotes.

Quand t'as eu des hémorroïdes, tu peux plus croire à la réincarnation.

Vingt-cinq ans, j'ai été bignole. Vous pensez si je retapisse un poulet.

Le coin des pas contents

Ah et puis ça ne va pas rater, les chevaliers blancs du Libre ne vont pas manquer d'agiter leur drapeau immaculé, donc on vous le dit d'avance : oui malheureusement la page de Common Voice contient du Google Analytics.



Et oui encore, amis sourcilleux sur les licences libres, la CC0 va permettre la réutilisation commerciale des voix enregistrées bénévolement. C'est dommage ou pas de placer un travail collectif sous une licence très permissive ?


CC0 1.0 universel (CC0 1.0)
Transfert dans le Domaine Public

Ceci est un résumé explicatif de la Licence ([lire la version complète](#)). [Avertissement](#)

Pas de droit d'auteur

 La personne qui a associé une œuvre à cet acte a **dédié** l'œuvre au domaine public en renonçant dans le monde entier à ses droits sur l'œuvre selon les lois sur le droit d'auteur, droit voisin et connexes, dans la mesure permise par la loi.

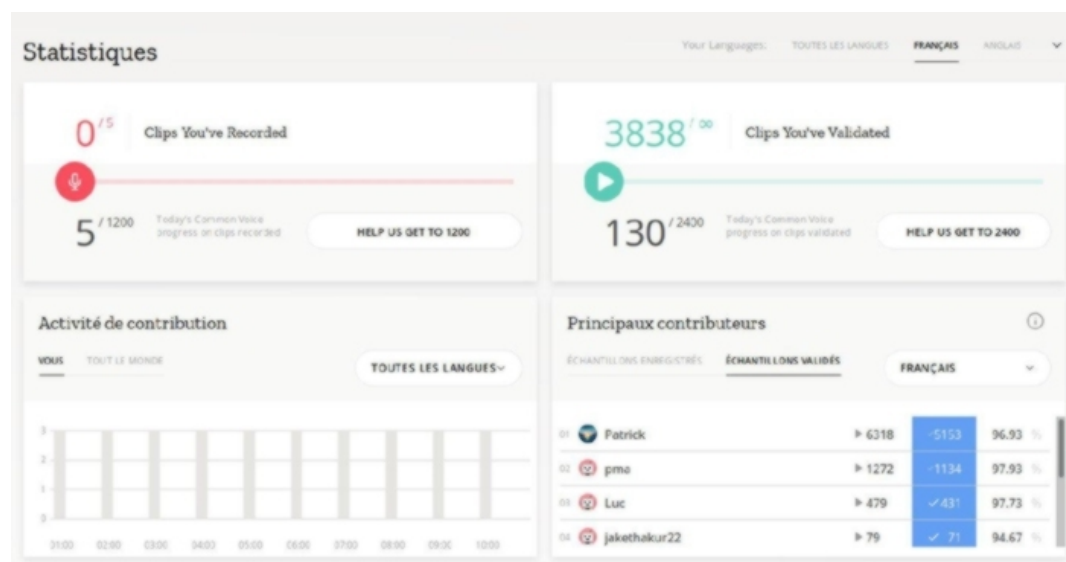
Vous pouvez copier, modifier, distribuer et représenter l'œuvre, même à des fins commerciales, sans avoir besoin de demander l'autorisation. Voir d'**autres informations** ci-dessous.

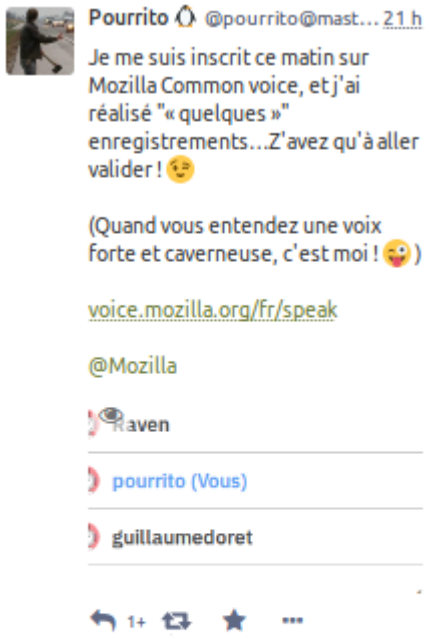


Cliquez sur l'image pour parcourir la page d'informations sur la CC0

À vous de jouer !

Chaque contribution est précieuse, surtout si elle est réitérée un grand nombre de fois ☐ Pour s'encourager, on peut suivre la progression des objectifs quotidiens, et certain·e·s ne manquent pas de se prendre au jeu...





Vous l'avez compris, ce projet réclame une participation massive pour avancer, alors si chacun y contribue ne serait-ce que modestement, nous pourrions dire que la voix est libre !

Liens utiles

- <https://voice.mozilla.org/fr/new>
 - Écouter pour valider ou non <https://voice.mozilla.org/fr/listen>
 - Lire à haute voix pour ajouter aux échantillons sonores <https://voice.mozilla.org/fr/speak>
 - La page de progression suivant les langues : on est loin du compte ! <https://voice.mozilla.org/fr/languages>
- Ludo répond à quelques questions pratiques sur un fil mastodon
- Une vidéo de conférence présentée par Alexandre sur le projet Deep Speech - Common Voice
- Le github du projet Deep Speech seulement pour ceux qui veulent se plonger dans sa dimension technique (on vous aura prévenu·e·s...)

- Des articles :
 - <https://blog.mozfr.org/post/2017/07/Haussons-la-voix-tous-ensemble-pour-le-Web-Common-Voice>
 - Un article qui résume le projet <https://blog.mozfr.org/post/2018/09/perdez-pas-voix-common-voice>
 - Un article récent : <https://www.blogdumoderateur.com/common-voice-mozilla-reconnaissance-vocale/>