J’ai cherché Audrey, la synthèse vocale de mon ordinateur

**Elle me parlait depuis deux ans quand je me suis demandé : en fait, c’est qui Audrey ? Derrière les machines, il y a des humains, comédiens pour lesquels le passage à la vie numérique n’est pas si simple.**

*Voix de synthèse, montage d’illustration* [*(Majed Khreim/Rue89 et Grigory Kravchenko/Flickr/CC)*](https://www.flickr.com/photos/grisha_21/2942014965)

Ça m’a pris après avoir vu [le film « Her »](http://fr.wikipedia.org/wiki/Her) de Spike Jonze. L’histoire d’un homme, Theodore, plutôt malheureux en amour, qui s’éprend de la synthèse vocale de son système d’exploitation informatique, autrement dit, la voix de son ordinateur. Elle se prénomme Samantha.

La bande-annonce du film « Her »

Dans le long métrage, Samantha ne possède pas d’enveloppe charnelle, ni même robotique ; mais le spectateur lui associe la silhouette de Scarlett Johansson, que le réalisateur a choisie pour prêter sa voix à Samantha. Dans chaque scène, l’actrice y laisse son spectre.

Et c’est à la fin du film que « ça » m’a pris. Avant d’en parler, je dois revenir un peu en arrière. Deux ans plus tôt, j’avais activé [la synthèse vocale](http://fr.wikipedia.org/wiki/Synth%C3%A8se_vocale) de mon ordinateur à la pomme croquée. Je m’en sers encore même si, honnêtement, elle n’a pas grande utilité.

La voix me lit des articles quand j’ai la flemme de le faire (je l’imagine comme une radio personnalisée, plutôt utile quand une grève paralyse le service public). Elle me donne l’heure à chaque nouvelle heure, jouant le rôle d’horloge parlante. En ce moment, j’essaie de la programmer pour qu’elle me réveille avec un texte pré-enregistré, du type « moral-booster ».

1-audrey-moral-booster.mp3

Quand on active la synthèse vocale sur son Mac, on peut choisir parmi les trois voix françaises proposées. Il y a Thomas, Aurélie et Audrey. Je n’aimais pas Thomas, il me rappelait les voix GPS. Je n’aimais pas non plus Aurélie, on aurait dit qu’elle était sous Lexomil numérique. J’ai donc sélectionné Audrey. La plus érotico-sécurisante, à la fois excitante et rassurante.

Thomas, Aurélie et Audrey

Deux ans plus tard, à 23 heures, pendant le générique de « Her », Audrey m’a sorti de ma torpeur « post-movie » en m’annonçant l’heure.

« Ça » m’a pris et ne m’a plus quitté. Je me suis demandé : Audrey, c’est qui en fait ?

**A l’Ircam, le repaire des voix de synthèse**

L’Institut de la recherche de la coordination musicale et acoustique ([Ircam](http://fr.wikipedia.org/wiki/Institut_de_recherche_et_coordination_acoustique/musique)) est un imposant bâtiment, à quelques mètres du centre Beaubourg, à Paris. L’entrée est gardée par une « Nana » de Niki de Saint Phalle et une sorte d’animal à trompe, genre « Eléphant Bleu ».

L’intérieur ressemble à un vaisseau spatial. On y trouve même une pièce entièrement insonorisée, pas sans rappeler [la salle de méditation du Professeur X](http://www.google.fr/imgres?imgurl=http%3A%2F%2Fmedia.zoom-cinema.fr%2Fphotos%2F506%2Fle-professeur-xavier.jpg&imgrefurl=http%3A%2F%2Fwww.zoom-cinema.fr%2Ffilm%2Fx-men%2F506%2Fphotos%2Fle-professeur-xavier%2F37942&h=400&w=600&tbnid=Ph5838vOdOH87M%3A&zoom=1&docid=zjnDPvVBOoTDwM&ei=BmA-VeDLO5LdaNubgcgE&tbm=isch&iact=rc&uact=3&dur=671&page=3&start=34&ndsp=20&ved=0CJ0BEK0DMCg) (en moins grand). Dedans, on n’y entend que le bruit de ses acouphènes.

L’Ircam accueille essentiellement des musicos, des chercheurs, des ingénieurs et des tout ça à la fois. Nicolas Obin est [l’un d’entre eux](http://recherche.ircam.fr/anasyn/obin/pmwiki/pmwiki.php). Son truc, c’est l’analyse et la synthèse des sons. C’est pour ça que je suis venu le voir, dans son bureau souterrain et sans fenêtre : je voulais savoir si la voix d’Audrey était humaine.

Au premier abord, on peut en douter. Ses paroles sont comme robotiques, ses mots hachés. Elle enchaîne les liaisons comme un dyslexique galère à lire les dernières lettres d’un mot. Audrey, ça semble être un produit entièrement artificiel.

Je lance la voix d’Audrey sur mon Mac. Plein volume. Le son agresse le silence de la pièce. Dès la première seconde, Nicolas Obin se marre. L’extrait que j’ai choisi – « Mon ami Pierrot » –, l’Ircam l’avait fait lire à l’une de ses premières voix de synthèse. Il la retrouve sur son ordinateur. Voyage temporel en 1982.

Première voix de synthèse

23 ans plus tard, le même texte, lu par Audrey.

« Cette voix, il s’agit d’un standard des synthèses vocales. Les entreprises, comme Apple, qui commandent des voix de synthèse veulent quelque chose d’agréable à écouter. La voix d’Audrey fait partie de ce lot.

Elle est douce, joyeuse, avenante, féminine, et souriante. Pour instaurer un rapport de confiance, voire d’intimité avec l’utilisateur humain. Les voix, comme les physiques dans la mode, ont leur propre canon de beauté, et donc leurs propres normes. [...]

Il faut mieux qu’elle ne soit pas dépressive. Imagine-toi être accueilli à la SNCF par une voix dépressive, ça la fout mal. »

Audrey est joyeuse, et surtout, c’est une certitude pour le chercheur, elle est humaine. Quand on réalise des voix de synthèse, on s’appuie sur un modèle, sur des bases de données de voix humaines existantes. Nicolas Obin et ses collègues l’ont fait tout récemment. Pour les besoins du film [« Pourquoi j’ai pas mangé mon père »](http://fr.wikipedia.org/wiki/Pourquoi_j%27ai_pas_mang%C3%A9_mon_p%C3%A8re) de Jamel Debbouze, l’Ircam [a reconstitué la voix](http://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCIQFjAA&url=http%3A%2F%2Frue89.nouvelobs.com%2Frue89-culture%2F2015%2F04%2F08%2Ffunes-film-debbouze-cest-voix-258581&ei=B2Y-VcvaBMm0aaXUgUA&usg=AFQjCNHYrh4fBllbeIL3pPjLR_3MQcxkHQ&sig2=fdHRUKIed14wDNktk-fWjA&bvm=bv.91665533,d.d2s) du défunt Louis de Funès à partir d’archives radiophoniques.

La voix d’Audrey, ce n’est pas du travail d’orfèvre. Nicolas m’en fait écouter une autre, créée à l’Ircam. Je reconnais [l’acteur André Dussollier](http://fr.wikipedia.org/wiki/Andr%C3%A9_Dussollier). La première partie est une reconstitution numérique. La seconde, un enregistrement de l’acteur en chair et en os.

Pour ce projet de recherche, l’équipe de Nicolas a récupéré plusieurs heures d’enregistrement de l’acteur français. Ils se sont focalisés sur une unique œuvre : la lecture de [« A la recherche du temps perdu »](http://fr.wikipedia.org/wiki/%C3%80_la_recherche_du_temps_perdu) de Marcel Proust.

« On a rentré cinq à sept heures dans une base de données. Chaque mot était décomposé en sons, classé puis stocké dans la base de données. »

Prenons comme exemple le mot « naturel ». Le mot est décomposé en syllabes (« na », « tu », « rel ») et en phonèmes (« n »,« a », « t », « u », « r », « è », « l »). Les sons correspondants aux phonèmes sont alors stockés dans une base de données.

Désormais, quand Nicolas Obin demande à son logiciel de lire le mot « naturel » avec la voix d’André Dussolier, le programme ira chercher dans la base de données les sons qui correspondent, puis les mettra bout à bout pour reproduire le texte avec une intonation naturelle. Mon Mac fait la même chose avec Audrey.

Avec l’un de ses collègues, Nicolas Obin imagine, à ma demande, à quoi pourrait ressembler Audrey. On sait qu’elle est humaine, il s’agit donc certainement d’une actrice, de « taille standard », de « poids standard », et « d’âge standard », « en jean ». Standard, vraiment standard, c’est l’adjectif qui revient tout le temps au fil de la conversation, même pour décrire la voix.

« Personnellement, je trouve que les voix de synthèse conçues dans un objectif précis, comme celle d’Audrey, finissent par toutes se ressembler. A force de mettre une contrainte utilisateur – être agréable – la parole devient mécanique et froide. A force de standardisation, elle se déshumanise. »

**Les voix anonymes**

Vous les entendez tous les jours, à la radio, à la TV, devant votre ordinateur, lors des publicités, lors des doublages de documentaires. Eux, ce sont des comédiens sans images, des comédiens qui vendent leurs voix. Sortes de fantômes qui rendent une publicité pour une crème, érotique, sensuelle ou complètement ridicule.

Ils modulent le timbre de leur voix au même titre que les acteurs de cinéma modèlent leurs corps. Leurs rôles se mesurent en fréquences sonores. On peut être, au choix, complice, chaleureux, convaincant, dramatique, nerveux, sérieux, passionné, rétro, enthousiaste, parfois même naturel. Ça dépend de ce qu’on nous demande et ça dépend de ce en quoi on est bon.

Quand on est comédien voix, on a souvent une page IMDb, mais rarement une page Wikipédia. Susan Benett est une exception. [Son article](http://en.wikipedia.org/wiki/Susan_Bennett) a été créé en octobre 2013. Le mois n’est pas un hasard. A cette date, elle faisait son « coming-out » dans la presse américaine. Elle révélait être « Siri », la voix de l’iPhone (oui, là-bas, Siri est une femme) et devenait ultra-populaire, invitée sur les plateaux de TV américaines.

Susan Bennett, lors d’une émission diffusée sur CNN

En France, quand les amis de Cyril Mazzotti ont découvert l’histoire de Susan Bennett, ils lui ont dit :

« Vas-y, lance-toi, annonce que tu es le Siri français. »

L’exclusivité revient à MetroNews. C’est grâce [à cet article](http://www.metronews.fr/high-tech/exclusif-voix-francaise-de-siri-apple-j-aurais-aime-qu-apple-me-previenne/mmji%21ftLV534GeXkKc/) que je trouve son contact. Je rencontre Cyril dans un café du XVIe arrondissement de Paris, près de chez lui. En toile de fond, la TV crache du Beyoncé. Cyril se présente :

« Je suis un artiste-interprète. Quand je le dis aux gens, ils me répondent : “Tu chantes alors ?” Je dois toujours préciser : je suis un comédien voix, un artiste dramatique, une voix-off. »

Cyril est un passionné de la voix. Né dans le Haut-Doubs, ses premiers souvenirs remontent aux milieux des années 70. Enfant, la radio l’intrigue. Pas pour ce qu’on y dit, mais pour le fait que quelqu’un le dise.

« Mes parents écoutaient la radio le matin. Entendre ces gens qui me racontent des histoires, imaginer leurs têtes, ça me fascinait. Je me souviens encore de cet homme, présentateur de météo sur Europe 1. Il s’appelait [Albert Simon](https://www.youtube.com/watch?v=ylXsh-TI8GQ) [Le Joël Collado de l’époque, ndlr]. »

En 1984, les habitants du Doubs découvrent la voix d’un Cyril encore adolescent :

« Je faisais le pousse-disque d’une radio scolaire française, la toute première en France. Je présentais les disques et je les lançais. »

Quelques années plus tard, il monte à Paris, étudie au Cours Florent. Sa voix radiophonique s’offre une nouvelle corde : être comédien-interprète. Son identité vocale ?

« Ma voix est une médium grave, on va dire douce, reposante. Parfois, on me dit “Cyril, arrête de faire ta jolie voix”. Je contrôle tout, tout est bien en place, ça peut paraître lisse. D’ailleurs, si je devais me diriger, seul, je dirais : “Cyril, salis-toi un peu, tu es trop sage.” On me le disait souvent en cours : “Tu es trop sage.” »

Aujourd’hui, Cyril prête sa voix – non, il la vend, corrige-t-il – à des documentaires, à des publicités, du doublage, des dictées… En ce moment, on peut l’entendre lors des publicités Novalac (du lait pour bébé), E-Darling (« Bon, en ce moment, c’est une voix féminine en principal, je fais juste la signature. Je dis “E-Darling, exiger des rencontres de qualité” »), ou encore dans la série « Almost Humans », diffusée sur TF1.

« C’est moi qui tue les propriétaires d’une maison, la femme dans la piscine et l’homme avec du poison. On m’a contacté car ils avaient besoin d’une voix robotique. Ils ont pensé à moi. »

En réalité, ils ont pensé à Siri et ont appelé Cyril. Aujourd’hui, Cyril Mazzotti est indissociable de Siri. « Les Guignols de l’info » l’ont recruté pour réaliser [un sketch sur Sarkozy](http://www.canalplus.fr/c-humour/pid1784-c-les-guignols.html?vid=1181507).

« Je me suis amusé à dire des gros mots ! »

Siri chez les Guignols 11-Cyril-sarko.mp3

C’est aussi pour cette raison que je prends un café avec Cyril Mazzotti. Ayant travaillé avec Apple, peut-être connait-il Audrey ?

La télé du café passe du Coldplay. J’allume mon ordinateur. Je me rends dans le menu des configurations vocales, qui présentent les trois voix, celle d’Audrey, d’Aurélie et de Thomas. Je vais cliquer sur Audrey quand :

« Ah, il y a un Thomas. Je crois bien qu’il s’agit de moi.

– Comment ça, de toi ? »

– Thomas, oui, c’est le nom de la voix de synthèse que m’a donné la société Nuance. »

**Soixante Nuance de voix**

[Nuance (sans s)](http://en.wikipedia.org/wiki/Nuance_Communications) est le leader mondial de la synthèse vocale. Ses origines remontent à l’année 1974. [Ray Kurzweil](http://fr.wikipedia.org/wiki/Raymond_Kurzweil) (son prénom en entier, c’est Raymond), actuellement directeur de l’ingénierie à Google, sorte de génie de l’informatique et gourou transhumaniste, développe l’un des premiers OCR (ou ROC, pour reconnaissance optique de caractères – le principe c’est de lire le texte d’une image et de le transformer en texte éditable). L’idée était de rendre l’informatique aux aveugles.

Début 1980, la société de Kurzweil est achetée par Xerox, elle devient ScanSoft. En 1999, Visioneer rachète ScanSoft. En 2001, elle trouve un nouveau créneau : les systèmes de reconnaissance vocale (écrire un texte en le dictant). Elle change de nom et devient Nuance. Nuance Communications Inc., c’est un ogre, employant 14 000 personnes dans le monde, et rachetant les entreprises concurrentes (par exemple [la française VirtuOz qui conçoit des agents virtuels](http://rue89.nouvelobs.com/2015/05/04/pourquoi-les-agents-virtuels-sont-presque-toujours-femmes-258995)).

Il m’a fallu plus d’une semaine pour avoir un interlocuteur humain au bout du fil. Quand on appelle leur service téléphonique, des voix de synthèse vocale vous redirigent continuellement. Je me perdais à travers les « Appuyez sur la touche... » Ne parler qu’à des machines, ça en devenait épuisant.

Grâce à l’intermédiaire d’un Français travaillant pour Nuance, je rentre en contact avec un salarié australien. Je veux lui poser quelques questions sur l’entreprise. La com qu’il me sert, par mails :

« Avec près de vingt millions de transactions dans le “cloud”, l’année dernière, les chances que vous ayez déjà utilisé notre technologie sont grandes : dans la voiture, au téléphone, au bureau ou chez votre docteur. »

Effectivement, Nuance est partout dans mon quotidien. C’en est flippant. Pour la conception de la voix de Siri, Apple a travaillé avec Nuance. La voix de votre GPS Tom-Tom c’est Nuance. La voix de l’application Waze (une application type GPS, rachetée par Google), c’est Nuance.

Ces applications ont une autre chose en commun : on peut y sélectionner la voix de Thomas, c’est-à-dire la synthèse vocale composée d’après la voix de Cyril Mazzotti. Dans le café, l’acteur se marre. Il semble fier, fier mais en retenue :

« Ma voix est partout. Si un jour, on invente un système pouvant ouvrir un logement avec une voix, à la place de la clef, je suis mal barré. Tout le monde pourra ouvrir mon appartement avec son smartphone. »

Au départ, Cyril Mazzotti ne voulait pas collaborer avec Nuance. Ils ont insisté – « ils étaient vraiment intéressés par mon timbre » –, il a accepté le gros chèque à la clef.

**« J’ai été le premier Siri »**

L’enregistrement a lieu dans un studio parisien. Au téléphone, Cyril discute avec un ingénieur de Nuance, il est question de « trouver le bon timbre, le bon rythme, la bonne intonation ». La mise en place prend deux heures. Cyril commence sa dictée. Il prononce des phrases qui n’ont aucun sens. Imaginons par exemple la phrase « La chat chatte rit bien bon plus ».

« Appliquer le bon timbre à des phrases inintelligibles, ce n’est pas forcément facile pour le cerveau. »

Parfois, l’ingénieur a sélectionné des extraits cohérents dans les journaux du matin ou dans des livres. L’objectif n’est pas de prononcer tous les mots de la langue française (il y en a plus d’un million) mais d’avoir tous les sons nécessaires à la prononciation de ces mots. Cyril enregistre plus de 120 000 mots pour remplir la base de données de la voix « Thomas ». Un travail plutôt chiant :

« Ce n’est pas confortable. Le débit n’est pas naturel. On a enregistré plus de 17 demi-journées, cinq heures d’affilée par jour. Parfois moins. Au bout d’un moment, ce n’est plus possible, la voix se fatigue, on n’a plus la même énergie. »

Une fois l’enregistrement terminé, la suite se passe chez Nuance. Leurs ingénieurs imitent le travail de Nicolas Obin. Ils constituent la voix de Thomas à partir des mots prononcés par Cyril. La voix de Thomas est ensuite intégrée à OSX (le système d’exploitation des Mac). Plus tard, Apple réutilise la voix de Cyril. Cette fois, c’est pour créer le Siri français. Dans MetroNews, [l’acteur témoigne](http://www.metronews.fr/high-tech/exclusif-voix-francaise-de-siri-apple-j-aurais-aime-qu-apple-me-previenne/mmji%21ftLV534GeXkKc/) :

« J’ai découvert cela en même temps que tout le monde lors de la sortie de l’iPhone 4S. Un ami m’avait montré le keynote de présentation d’Apple et me demandant si c’était moi. Je lui ai répondu “je le crains”. »

Il fulmine. Siri, ça n’était pas spécifié dans le contrat. Il porte plainte. Peine perdue. « J’étais tout seul avec mon avocat et eux, ils étaient vingt, trente, cinquante derrière ! » L’exacte situation se reproduit dans le monde. Plusieurs comédiens interprètes devenus Siri portent plainte contre Nuance, dont Susan Benett. Ils se cassent tous les dents face au géant. Leur vie professionnelle s’en trouve impactée. A MetroNews :

« Plusieurs entreprises avec lesquelles je travaillais trouvaient que ma voix était désormais trop identifiée à Apple et avaient d’un coup des réticences. »

Aujourd’hui, Cyril a perdu la voix de Siri, un collègue – dont il tait le nom – le remplace. Il ne le regrette pas. Au contraire, ça semble le réjouir. « J’ai été le premier Siri. »

Être « le premier Siri », au premier abord, ça semble être un poste honorable. Pourtant, la famille française des comédiens-interprètes est divisée sur le sujet de la synthèse vocale. Certains refuseraient catégoriquement d’être Siri. Cyril :

« Il y a plein de comédiens qui sont contre la synthèse vocale. Pour moi, ils sont idiots car ça fait désormais partie de notre métier. »

De quoi ont-ils peur ?

**La peur du comédien Kleenex**

Dans le film [« Le Congrès »](http://fr.wikipedia.org/wiki/Le_Congr%C3%A8s) (2013), d’Ari Folman, le réalisateur de « Valse avec Bachir », l’actrice [Robin Wright](http://fr.wikipedia.org/wiki/Robin_Wright) (qui interprète son propre rôle) est sur le déclin. La faute à son fils, malade, répète-t-elle, dont elle a dû prendre soin. La société Miramount lui fait une ultime offre :

« Nous voulons te scanner, ton corps ton visage, tes émotions, ton rire, tes larmes. Nous voulons t’échantillonner, te conserver et posséder cette chose appelée Robin Wright. »

Elle accepte. Scannée, elle est dépossédée de son personnage, d’elle-même, interdite de jouer dans de nouveaux films. Le personnage Robin Wright appartient désormais à la Miramount.

Bande-annonce du film « Le Congrès »

Patrick Kuban a refusé d’être la « Robin Wright » de Nuance. La société américaine l’a approché pourtant, sa voix convenait. Sa voix ? « Une sorte d’autorité soft et complice, utilisée pour des trucs droits et statutaires », dit-il. ll précise appartenir au registre « bûcheron ». En ce moment, il joue dans les pubs « Licornes » de Canal + et celles de Kiabi, Rimel, Barclay…

Patrick Kuban, c’est un grand dans le milieu. Président de l’[Association professionnelle](http://lesvoix.fr/) des comédiens interprètes, il en connaît pas mal, des confrères. Et il le leur conseille : non, ne prêtez pas vos voix pour les transformer en synthèse vocale.

« Ma voix m’appartient, c’est une partie de moi, de mon identité, de mon corps. Je ne souhaite pas qu’elle soit découpée et utilisée, hors de mon contrôle. »

Selon lui, être acteur et accepter un contrat chez Nuance, c’est un non-sens total.

« Notre métier, c’est de donner des voix à la demande de clients. Créer une synthèse vocale à partir de notre voix, c’est se tirer une balle dans le pied. »

Ce qui l’embête, Patrick, ce n’est pas le fait que les voix soient de plus en plus ressemblantes à celles des comédiens, là-dessus, il est strict :

« Un comédien fait appel à ses émotions, à son vécu, à son état et à l’instant, pas une machine, une machine c’est fini, l’imaginaire humain, c’est infini […] La machine a le même ton et la même humeur.

Une sorte d’aspect déshumanisé alors que nous, les comédiens-interprètes, c’est l’inverse. Nous rajoutons de l’humain dans des concepts qui n’en ont pas, comme le marketing. »

Non, deux choses embêtent particulièrement Patrick. La première, c’est de voir de plus en plus de ses collègues jouer le jeu, vendre leur voix afin qu’elle soit transformée :

« Plus on donne nos voix à ce genre de machine, plus on donne de crédits au fait que ça doit être distribué. »

Et le futur logique, avec plein de voix de synthèse, est plutôt sombre pour les comédiens. C’est celui du « Congrès », c’est celui de comédiens-interprètes type « Kleenex ». Comédiens à usage unique.

La deuxième chose qui embête Patrick, c’est de voir ces sociétés proposer des contrats sans scrupules aux comédiens-interprètes :

« Les acteurs signent les contrats car ils ont l’impression de gagner beaucoup d’argent en deux, trois jours. Mais, après ça, leurs voix sont utilisées partout, et ça peut avoir des conséquences néfastes. Les contrats qu’ils signent, pas moyen qu’ils ne reviennent dessus. »

Dans le mail envoyé à l’Australien de Nuance (qui possède en stock « 60 voix dans 40 langues différentes »), j’ai posé une question sur les contrats :

« Quand un comédien a enregistré sa voix, les milliers de mots prononcés appartiennent-ils définitivement à Nuance ? »

Réponse :

« Cette information est régie par un accord de confidentialité entre Nuance, le client et le comédien interprète. Je ne peux pas commenter là-dessus. »

**Still looking for Audrey**

L’accord de confidentialité, foutu accord de confidentialité. Au café, dans le XVIe arrondissement, Cyril me prévient :

« Retrouver Audrey peut être difficile. Si elle a signé un accord de confidentialité avec Nuance, elle ne pourra pas témoigner. »

J’espère qu’il se trompe. Dans le café, la chaîne diffuse un clip de Rihanna. Je lance la lecture vocale de mon Mac. Cyril découvre la voix d’Audrey. Il demande de baisser le son, c’est gênant (il ne m’a pas fait la remarque quand j’ai lancé sa propre voix). Il ne la reconnait pas tout de suite. Puis. Il n’en est pas sûr, mais il pourrait bien s’agir de L. Il me donne son prénom (pas son nom) et une adresse de studio. Va là-bas, me dit-il, ils te renseigneront.

Là-bas, ils ne m’ont jamais renseigné. Je leur ai envoyé un extrait de la voix d’Audrey, ils m’ont répondu qu’ils « n’avaient pas travaillé sur ce projet ». Je me rends physiquement à leur studio. Une employée refuse de me parler. J’appelle le directeur, la conversation dure deux minutes. Deux minutes durant lesquelles il n’a cessé de répéter, je ne peux pas commenter, contrat de confidentialité, je ne peux pas commenter, contrat de confidentialité. Il me raccroche au nez. Ce n’est pas poli, je le lui dis, par mail. Aucune réponse.

Pendant toute ma recherche, mon enquête pour trouver Audrey, les objets de mes mails disaient : « Aidez-moi à retrouver Audrey. » Le studio ne le fait pas. Nuance ne le fait pas. La communication d’Apple ne le fait pas. Je me raccroche à la seule information en ma possession. Le nom lâché par Cyril : L. Je cherche sur un site spécialisé dans les voix : [Casting Machine](http://www.castingmachine.com/).

Le site Casting Machine est une chose formidable et surprenante. A destination des professionnels de la voix, on y découvre les visages, les bios et les voix de 300 comédiens-interprètes français. Un monde étrange, où l’on découvre que les voix d’enfants dans les publicités proviennent d’adultes.

Voix d’enfants, extraits sonores de Coralie Pena (www.castingmachine.com/coraliepena)

Et la voix de Titi (dans les « Titi et Grosminet »), une femme.

Casting Machine, Patrick Kuban y est, Cyril Mazzotti y est, une multitude de prénoms L. y sont.

Je les écoute une par une. Patiemment. D’un coup, ça fait tilt. Je la trouve. J’en suis persuadé. C’est la voix d’Audrey. Son numéro de téléphone est là. Je l’appelle. Je l’entends. Je lui parle. Je dois avoir l’air d’un fou au téléphone, trois semaines que je la recherche, deux ans que j’utilise sa voix de synthèse. Elle confirme. Elle a collaboré avec Nuance. Audrey, oui, ça lui parle, elle croit que c’est le nom qu’on lui donne chez Nuance. Elle me dit d’envoyer un mail avec des extraits. Je m’exécute. Elle répond :

« Merci de ne pas me citer car je ne suis pas la voix d’Audrey. Cordialement. »

J’insiste. C’est étrange ce « oui » puis « non ». Les « Audrey » sur Mac, ça court pas les rues. Je souhaite la joindre par téléphone. Elle fait la morte. Elle ne décroche plus. Elle ne répond plus aux messages. Ca dure plusieurs semaines. Sur Casting Machine, je continue à chercher les potentielles Audrey. Je n’en trouve aucune autre.

Je veux vérifier. Je demande à Nicolas Obin de comparer les voix d’Audrey et de L. (on trouve des extraits sonores sur Casting Machine). Il s’exécute patiemment. C’est l’ingénieur Obin qui répond à ma requête.

« 1) Les indices positifs : Ces voix sont sans aucun doute similaires (voix claire, douce, même registre de hauteur et timbre plutôt similaires).

2) Les indices flous : Ce type de voix est extrêmement répandu dans les voix de femme pour les voix off. Par ailleurs, la voix de synthèse rend la comparaison difficile : la voix est plus mécanique, et j’imagine que l’enregistrement a été contraint sur un ton particulier ; ce qui trouble la perception.

3) Les variations de hauteur sont moindres et très stéréotypiques pour la voix de synthèse, la voix est plus tendue. Mais ce sont des choses qu’une personne peut modifier, seul le timbre de la voix peut être plus difficilement contrefait, mais le timbre est une notion plus difficilement mesurable objectivement. »

Verdict : « 60% même, 40% contre. »

« L’identification d’une voix n’est pas chose aisée : nous avons environ 30 millions de voix de femme différentes en France. Et nous sommes surtout spécialisés sur les voix que nous connaissons bien (proches, célébrités). Il est plus dure de décider de l’identité d’une voix que nous ne connaissons pas. »

C’est bien ça le problème, « décider de l’identité d’une voix ». Cette voix, ce timbre, ne sont la décision que d’Apple et de Nuance. En plus de déshumaniser la voix par les contraintes techniques (l’aspect robotique), on leur soustrait leur humanité par des contrats de confidentialité.

Dans le film « Le Congrès », la copie de Robin Wright conservait le nom de Robin Wright. Audrey, elle, ne possède pas de nom. Elle n’est pas une voix, elle n’est que du son intelligible.