La production, la reproduction et la transmission de la voix humaine

La voix humaine tient une place particulière dans la recherche acoustique.
Tout d'abord il s'agit d'un instrument de musique "vivant", faisant partie intégrante du corps humain, dont l'observation en cours de production sonore reste, aujourd'hui encore, difficile d'accès.
Ensuite, les signaux produits par la voix humaine sont dotés d'une très forte prégnance cognitive, et forment une catégorie perceptive qui se structure très tôt dans la m auditive. S'inscrivant implicitement dans un schéma de communication impliquant au moins deux partenaires, ils activent alternativement deux attitudes d'écoute. La première est l'écoute sémantique, en quête de sens : identité du locuteur, signification des paroles prononcées, éventuellement signes sémantiques prosodiques; la deuxième est l'écoute qualitative, celle des variations esthétiques des paramètres sonores, mise en œuvre lors de la production chantée.

Comprendre le mécanisme de la voix

Doter la machine de la parole a toujours intéressé les savants. L'effervescence créative du siècle des Lumières a été marquée par la création d'automates. La volonté de les faire parler est apparue en parallèle.

Christian Gottlieb Kratzenstein (1723 – 1795)

La première machine de synthèse vocale mécanique connue à ce jour a été construite par le scientifique allemand Christian Gottlieb Kratzenstein. Né le 30 janvier 1723 à Wernigerode, il obtenait à l'âge de 23 ans des doctorats en médecine et physique à l'université de Halle pour ses thèses Theoria fluxus diabetici et Theoria electricitatis mores geometrica explicata concernant la nature de l'électricité. Deux ans plus tard, il a été appelé à l'Académie des Sciences à Saint-Pétersbourg. En 1753, il est engagé comme professeur de physique expérimentale à l'université de Copenhague où il fût nommé recteur quatre fois de suite.

Gottlieb Kratzenstein était un proche (certains disent un protégé) du mathématicien renommé suisse Leonhard Euler qui exerçait également à l'Académie des Sciences à Saint-Pétersbourg. Dans sa correspondance (conservée aux archives de l'université de Bâle) avec le mathématicien Johann Heinrich Lambert, Leonhard Euler exprimait dès 1758 ses réflexions sur la nature des voyelles a, e, i, o, u et proposait la construction d'une machine parlante pour mieux comprendre l'organisme du langage humain. Sur initiative de Leonhard Euler, l'Académie des Sciences à Saint-Pétersbourg organisait en 1778 une compétition pour réaliser un appareil de synthèse des voyelles de la voix humaine.

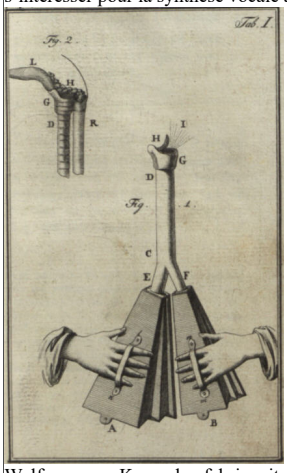
Après son départ à Copenhague, Gottlieb Kratzenstein avait gardé de bonnes relations avec ses pairs à l'académie et il était très intéressé à cette compétition. En se basant sur les études afférentes de Leonhard Euler, il a parfait sa machine parlante construite dès 1773, en utilisant des tubes et des tuyaux organiques pour créer des cordes vocales artificielles. Il s'est surtout inspiré du jeu vox humana de l'orgue.
En 1780, il a gagné le premier prix de la compétition avec son projet d'un orgue vocal. Une année plus tard il a publié une description en latin du projet : Tentamen Resolvendi Problema. Une version traduite en allemand de cette œuvre a été publiée en 2016.

Gottlieb Kratzenstein n'a pas seulement fourni une contribution appréciée pour la synthèse vocale, mais il a également fait avancer la science dans les domaines de la chimie, de la navigation, de l'astronomie, de la neurologie et de l'électricité. En 1743 il avait même rédigé une œuvre philosophique Beweis, dass die Seele ihren Körper baue. Dans la suite il publiait Abhandlung von dem Nutzen der Electricität in der Arzeneiwissenschaft concernant l'électrothérapie.
Gottlieb Kratzenstein était marié et il avait quatre enfants. Il est décédé en 1795. Un de ses petits-fils est Christian Gottlieb Stub, un peintre danois renommé qui porte les prénoms de son grand-père et qui a ajouté dans la suite également le nom de famille Kratzenstein à son nom. Né en 1783, il est décédé en 1816 à l'âge de 33 ans.
Dans ce contexte il convient de mentionner également l'échange de lettres entre Gottlieb Kratzenstein et l'astronome renommé Johann III Bernouilli après la mort de l'épouse du premier. Susan Splinter a rédigé une contribution Ein Physiker auf Brautschau à ce sujet.

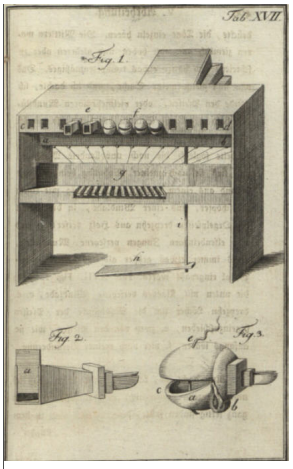
Johann Wolfgang von Kempelen (1734 – 1804)

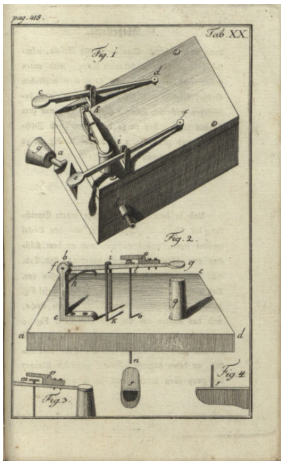
Si on s'intéresse pour l'histoire de la synthèse vocale on pense d'abord à Johann Wolfgang von Kempelen. Né le 23 janvier 1734 à Presbourg (aujourd'hui Bratislava) et décédé le 26 mars 1804 à Vienne, Wolfgang von Kempelen (Farkas Kempelen en hongrois) était ingénieur et conseiller aulique à la Cour impériale de Vienne. Il est surtout connu sous le nom de Baron von Kempelen pour l'invention du Turc mécanique en 1769, un automate célèbre qui avait l'apparence d'un Turc et actionnait les pièces d'un jeu d'échecs.

Wolfgang von Kempelen a commencé au début des années 1770 à construire une machine parlante, c.à.d. à la même époque où Gottlieb Kratzenstein commencait à s'intéresser pour la synthèse vocale à Copenhague.



Wolfgang von Kempelen fabriquait plusieurs prototypes qui menaient à des échecs. Pour la première version il utilisait un soufflet de cuisine, une anche (roseau) de cornemuse et une cloche de clarinette.
Pour la seconde version il utilisait une console d'orgue avec un clavier où les différentes touches étaient associés à des lettres. Les sons étaient produits avec des tubes de différentes formes et longueurs. Le problème était toutefois le chevauchement (co-articulation) des différents sons qui empêchait la génération de syllabes.





Wolfgang von Kempelen concluait qu'il fallait mieux comprendre le fonctionnement de l'appareil phonatoire humain pour progresser. Ce n'est qu'au début des années 1780 que son modèle réalisé lui donnait satisfaction et qu'il le présentait au public.

Contrairement à la machine de Gottlieb Kratzenstein, la construction de Wolfgang von Kempelen était la première à produire, non seulement certaines voyelles, mais surtout des mots entiers et des courtes phrases.
En 1791 Wolfgang von Kempelen a publié un livre Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine pour expliquer aux personnes intéressées les principes de sa machine. Le livre comprenait 456 pages. L'objectif de cet ouvrage n'était pas seulement d'élucider le mystère de l'étonnant appareil, mais aussi d'inciter le lecteur à le perfectionner de sorte qu'on puisse enfin en obtenir ce pour quoi il fut imaginé. On peut considérer la description de Wolfgang von Kempelen comme un premier projet "open-source".

L'auteur formule dans son livre quelques hypothèses essentielles sur la production de la parole humaine. Il proposait une conception de la langue qui n'était plus envisagée comme souffle de l'âme, mais tout simplement comme de l'air s'échappant à travers des fentes de formes variables. En s'observant lui-même, Wolfgang von Kempelen décrivait également les différents sons et les positions que devaient prendre les organes phonateurs pour les produire.

Le musée des sciences et de la technique à Munich (Deutsches Museum), créé en 1903, expose dans un coin du département des instruments de musique une machine parlante désignée comme celle construite par Wolfgang von Kempelen. Elle a été offerte au musée en 1906 par l'Académie de la Musique à Vienne. Suite à ses recherches, Fabian Brackhane conteste qu'il s'agit de l'original de Wolfgang von Kempelen. Il estime plutôt qu'il s'agit d'une réplique construite par Charles Wheatstone.

Il s'agit d'une construction composée d'un caisson en bois, d'un entonnoir en caoutchouc qui faisait office de bouche et d'un second, plus petit, divisé en deux, qui remplissait les fonctions d'un nez. Le mécanisme interne était un soufflet qui simulait les poumons. Le flux d'air était conduit dans la "bouche" par l'intermédiaire d'un couloir très étroit. Une hanche vibrante, sorte de glotte et de cordes vocales réunies, produisait un son. Celui-ci était ensuite modulable par différents petits leviers et l'utilisation des doigts de l'opérateur pour modifier l'air à la sortie de la "bouche" afin de simuler le mouvement des lèvres.

Grâce au livre de Wolfgang von Kempelen, plusieurs chercheurs ont réussi à construire une réplique de sa machine parlante. Un exemple a été créé au département des sciences et technologies du langage à l'université de la Sarre.

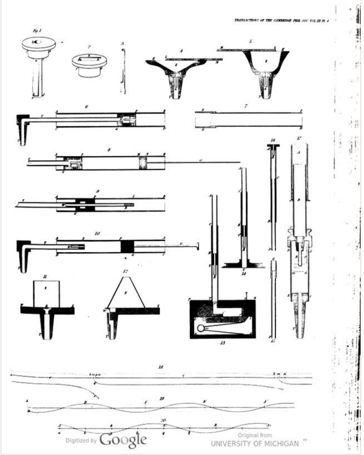
Depuis sa création, plusieurs chercheurs ont repris le travail de Wolfgang von Kempelen en y intégrant les nouvelles technologies développées à leur époque.

Serge DURIN, facteur d'instruments à vent, teste la machine parlante reconstruite d'après le traité écrit par le baron WOLFGANG VON KEMPELEN.
Le musée des sciences et de la technique à Munich y a présenté sa machine parlante de von Kempelen. Cette exposition virtuelle peut encore être visitée sur la plate-forme Arts & Culture de Google qui était partenaire de l'exposition.

Robert Willis (1800 – 1875)

Né en 1800, Robert Willis était un académicien anglais, renommé comme ingénieur en mécanique et pour ses publications au sujet de l'architecture. Il a été ordonné prêtre en 1827. Le révérend Willis a publié deux contributions très appréciées sur la mécanique de la parole humaine dans le journal Transactions of the Cambridge Philosophical Society : On vowel sounds, and on reed-organ pipes en 1828 et On the Mechanism of the Larynx en 1829.
La contribution de Robert Willis a été traduite en allemand en 1832 dans le journal scientifique Annalen der Physik und Chemie.

Robert Willis a repris les travaux pratiques de Gottlieb Katzenstein et il a testé plusieurs variantes de résonateurs pour parfaire la génération des voyelles.

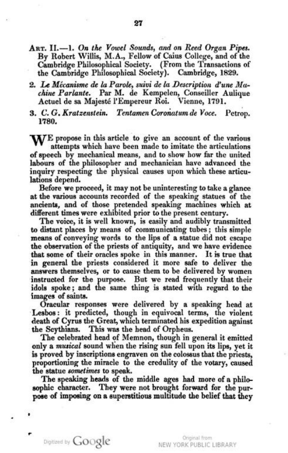


Transactions of the Cambridge Philosophical Society, 1830, Vol III

Charles Wheatstone (1802 – 1875)

Né en 1802, Charles Wheatstone est un physicien et inventeur anglais. On lui doit le premier télégraphe électrique au Royaume-Uni, le pont de Wheatstone, le stéréoscope, un microphone, et l'instrument de musique Concertina.

En 1835, Charles Wheatstone présentait une réplique de la machine parlante de Wolfgang von Kempelen. Il profitait des progrès technologiques réalisés dans les dernières décennies pour parfaire la construction.
En 1837 il publiait un article au sujet des inventions de Gottlieb Kratzenstein et de Wolfgang von Kempelen ainsi que sur la contribution “On vowel sounds” de Robert Millis, dans le journal The London and Westminster Review.



27

ART. II.—1. *On the Vowel Sounds, and on Reed Organ Pipes.* By Robert Willis, M.A., Fellow of Caius College, and of the Cambridge Philosophical Society. (From the Transactions of the Cambridge Philosophical Society). Cambridge, 1829.

2. *Le Mécanisme de la Parole, suivi de la Description d'une Machine Parlante.* Par M. de Kempelen, Conseiller Aulique Actuel de sa Majesté l'Empereur Roi. Vienne, 1791.

3. *C. G. Kratzenstein. Tentamen Coronatum de Voce.* Petrop. 1780.

WE propose in this article to give an account of the various attempts which have been made to imitate the articulations of speech by mechanical means, and to show how far the united labours of the philosopher and mechanician have advanced the inquiry respecting the physical causes upon which these articulations depend.

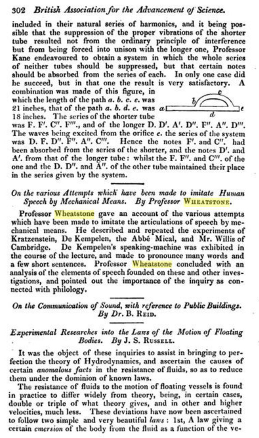
Before we proceed, it may not be uninteresting to take a glance at the various accounts recorded of the speaking statues of the ancients, and of those pretended speaking machines which at different times were exhibited prior to the present century.

The voice, it is well known, is easily and audibly transmitted to distant places by means of communicating tubes; this simple means of conveying words to the lips of a statue did not escape the observation of the priests of antiquity, and we have evidence that some of their oracles spoke in this manner. It is true that in general the priests considered it more safe to deliver the answers themselves, or to cause them to be delivered by women instructed for the purpose. But we read frequently that their idols spoke; and the same thing is stated with regard to the images of saints.

Oracular responses were delivered by a speaking head at Lesbos: it predicted, though in equivocal terms, the violent death of Cyrus the Great, which terminated his expedition against the Scythians. This was the head of Orpheus.

The celebrated head of Memnon, though in general it emitted only a *musical* sound when the rising sun fell upon its lips, yet it is proved by inscriptions engraven on the colossus that the priests, proportioning the miracle to the credulity of the votary, caused the statue *sometimes* to speak.

The speaking heads of the middle ages had more of a philosophic character. They were not brought forward for the purpose of imposing on a superstitious multitude the belief that they



302 *British Association for the Advancement of Science.*

included in their natural series of harmonics, and it being possible that the suppression of the proper vibrations of the shorter tube resulted not from the ordinary principle of interference but from being forced into unison with the longer one, Professor Kane endeavoured to obtain a system in which the whole series of neither tubes should be suppressed, but that certain notes should be absorbed from the series of each. In only one case did he succeed, but in that one the result is very satisfactory. A combination was made of this figure, in which the length of the path *a. b. c. e.* was 21 inches, that of the path *a. b. d. e.* was 18 inches. The series of the shorter tube was F. F′. C″. F‴., and of the longer D. D′. A′. D″. F″. A″. D‴. The waves being excited from the orifice *e.* the series of the system was D. F. D″. F″. A″. C‴. Hence the notes F′. and C″. had been absorbed from the series of the shorter, and the notes D′. and A′. from that of the longer tube: whilst the F. F″. and C‴. of the one and the D. D″. and A″. of the other tube maintained their place in the series given by the system.

On the various Attempts which have been made to imitate Human Speech by Mechanical Means. By Professor WHEATSTONE.

Professor Wheatstone gave an account of the various attempts which have been made to imitate the articulations of speech by mechanical means. He described and repeated the experiments of Kratzenstein, De Kempelen, the Abbé Mical, and Mr. Willis of Cambridge. De Kempelen's speaking-machine was exhibited in the course of the lecture, and made to pronounce many words and a few short sentences. Professor Wheatstone concluded with an analysis of the elements of speech founded on these and other investigations, and pointed out the importance of the inquiry as connected with philology.

On the Communication of Sound, with reference to Public Buildings. By Dr. B. REID.

Experimental Researches into the Laws of the Motion of Floating Bodies. By J. S. RUSSELL.

It was the object of these inquiries to assist in bringing to perfection the theory of Hydrodynamics, and ascertain the causes of certain *anomalous facts* in the resistance of fluids, so as to reduce them under the dominion of known laws.

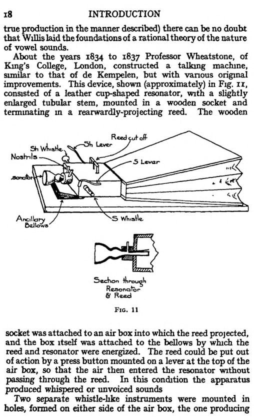
The resistance of fluids to the motion of floating vessels is found in practice to differ widely from theory, being, in certain cases, double or triple of what theory gives, and in other and higher velocities, much less. These deviations have now been ascertained to follow two simple and very beautiful *laws*: 1st, A law giving a certain *emersion* of the body from the fluid as a function of the ve-

-> Rapport de la réunion à Dublin en 1835

A gauche "The London and Westminster review, Vol 28, october 1837"

Deux années plus tôt, Charles Wheatstone avait déjà exposé ses études et présenté sa réplique d'une machine parlante à l'association britannique pour l'avancement des sciences lors de son assemblée à Dublin en août 1935.

En 1930 Richard Paget, un avocat anglais et amateur des sciences, publiait son livre Human Speech. C'est le dernier qui rapportait au sujet de la réplique de la machine parlante réalisé par Charles Wheatstone sur base de la description de Wolfgang von Kempelen. Dans son livre il a publié l'esquisse suivante de l'engin :



18 INTRODUCTION

true production in the manner described) there can be no doubt that Willis laid the foundations of a rational theory of the nature of vowel sounds.

About the years 1834 to 1837 Professor Wheatstone, of King's College, London, constructed a talking machine, similar to that of de Kempelen, but with various original improvements. This device, shown (approximately) in Fig. 11, consisted of a leather cup-shaped resonator, with a slightly enlarged tubular stem, mounted in a wooden socket and terminating in a rearwardly-projecting reed. The wooden

FIG. 11

socket was attached to an air box into which the reed projected, and the box itself was attached to the bellows by which the reed and resonator were energized. The reed could be put out of action by a press button mounted on a lever at the top of the air box, so that the air then entered the resonator without passing through the reed. In this condition the apparatus produced whispered or unvoiced sounds

Two separate whistle-like instruments were mounted in holes, formed on either side of the air box, the one producing

Livre Human Speech de Richard Page, 1930, page 18

Fabian Brackhane a signalé dans sa dissertation Kann was natürlicher, als Vox humana, klingen ? de 2015 que la réplique de Charles Wheatstone se trouve aujourd'hui dans un dépôt du Musée des Sciences à Londres avec l'attribution Wheatstone's artificial voice box. Il a même réussi à obtenir des photos de la construction qu'il a publiée dans sa dissertation.

Joseph Faber (1800 – 1850)

Joseph Faber est né aux environs de 1796 à Fribourg-en-Brisgau. Il a fait des études de physique, mathématiques et musique à l'Institut polytechnique impériale et royale à Vienne. Pour se remettre d'une grave maladie, il est retourné à sa ville de naissance en 1820. Pendant sa reconvalescence, il s'est mis à construire pendant 17 ans une machine parlante améliorée sur base du livre de Wolfgang von Kempelen.

Une ancienne gravure, dont on ignore l'auteur et la date de création (éventuellement 1835), montre une jeune dame (probablement l'épouse de Joseph Faber) qui manipule le clavier à 16 touches de la machine parlante. La face de la tête est posée sur la table et on peut voir la bouche de la machine avec des lèvres et une langue.

Gravure ancienne de la machine parlante Euphonia.

En 1840, il a présenté son invention, qu'il appelait Euphonia, au public à Vienne et au Roi de Bavière en 1841. L'instituteur Schneider de Bauernwitz décrit en 1841 dans un journal pédagogique (Der katholische Jugendbildner) le fonctionnement de la machine qu'il a pu voir lors de sa visite à Vienne la même année. L'existence de cette machine a été signalée en janvier 1841 dans le même journal.

En 1842 il exposait la machine à Berlin et à Dresde, une année plus tard à Leipzig. Comme il ne rencontrait pas l'intérêt souhaité, il décidait de l'exposer aux États-Unis. En 1844 il la présentait à New York, une année plus tard à Philadelphia (Musical Fund Hall), sans succès. Ce fût Phineas Taylor Barnum, l'entrepreneur américain de spectacles (cirque Barnum) qui la rendit célèbre. Il amena Joseph Faber à Londres et présenta la machine à l'Egyptian Hall à partir de 1846.

Le périodique Illustrated London News du 8 août 1846 (page 16) rapportait sur cette nouvelle attraction exposée à Londres.

receive our drawings of Fashions from the Editor of the *Modes Parisiennes*, the most elegant and distinguished of all the Journals of the Fashions, and patronised by the Royal Families of Europe. Subscriptions for this work are received by Madame Frederick, No. 40, Albemarle-street: it is forwarded direct from Paris, and is received in London every Tuesday morning, price, for six months, 12 shillings.

THE EUPHONIA.

We noticed this new Speaking Automaton a fortnight since, and then recorded its feats *arte loquendi*. It is stated by Professor Faber to be the result of twenty-five years' labour! It is well observed in the *Athenæum*:—"It is in vain to apply the 'Cui bono?' to a matter like this. It is quite true that mechanical figures, in heads and turbans, with their lungs in red baize and worked by machinery, are not in themselves utilities—the more particularly as their talking machinery requires the impulse of a real living and talking man, who might more conveniently have done the talking at first hand. As an example of inductive and mechanical skill, however, such an exhibition as this is well deserving of attention; and there is no difficulty, besides, in imagining a number of purposes to which the discovery of any artificial means for producing vocal articulation might be applied with valuable effect. It is, in any case, an old scientific problem; and anything that brings us nearer to its solution would have an interest, were it for that reason alone. We believe this invention of Professor Faber comes closer to that result than any previous 'instrument made with hands.'"

THE EUPHONIA, OR SPEAKING MACHINE.

Still, the writer continues: "This is, like all similar attempts which have preceded it, only an approximation, though a nearer approximation, to the thing proposed. It requires all our sense of the ingenuity and perseverance which have been bestowed on the work, to induce our assent to the proposition which calls the voice a human voice; but we recommend it to notice as a remarkable result of contriving skill and scientific patience."

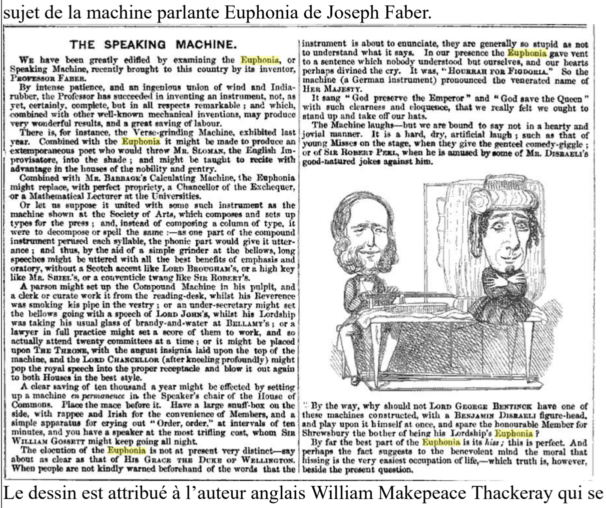
In this light, the Euphonia has been numerously visited and inspected.

The Speaking Head of Roger Bacon, and similar inventions of much earlier date, show that the idea of applying machinery to imitate life, is of some antiquity, and that considerable success was not deemed impossible. One of the greatest masters in this way was Vaucanson, with his duck, his flute and flageolet players; but even he must *duck* his diminished head before the Euphonia of our day.

London: Printed and Published at the Office, 198, Strand, in the Parish of St. Clement Danes, in the County of Middlesex, by William Little, 198, Strand, aforesaid.—Saturday, August 8, 1846.

La face de la tête a été appliquée et le mécanisme de la machine a été caché par un rideau et un buste de poupée pour faire ressembler la machine à un vrai personnage.

Le journal satirique anglais Punch (The London Charivari) a publié dans son volume 11, 1846, page 83, une contribution Speaking Machine avec un dessin humoristique au sujet de la machine parlante Euphonia de Joseph Faber.



THE SPEAKING MACHINE.

We have been greatly edified by examining the Euphonia, or Speaking Machine, recently brought to this country by its inventor, Professor Faber.

By intense patience, and an ingenious union of wind and India-rubber, the Professor has succeeded in inventing an instrument, not, as yet, certainly, complete, but in all respects remarkable; and which, combined with other well-known mechanical inventions, may produce very wonderful results, and a great saving of labour.

There is, for instance, the Verse-grinding Machine, exhibited last year. Combined with the Euphonia it might be made to produce an extemporaneous poet who would throw Mr. Sloman, the English Improvisatore, into the shade; and might be taught to recite with advantage in the houses of the nobility and gentry.

Combined with Mr. Babbage's Calculating Machine, the Euphonia might replace, with perfect propriety, a Chancellor of the Exchequer, or a Mathematical Lecturer at the Universities.

Or let us suppose it united with some such instrument as the machine shown at the Society of Arts, which composes and sets up types for the press; and, instead of composing a column of type, it were to decompose or spell the same:—as one part of the compound instrument perused each syllable, the phonic part would give it utterance; and thus, by the aid of a simple grinder at the bellows, long speeches might be uttered with all the best benefits of emphasis and oratory, without a Scotch accent like Lord Brougham's, or a high key like Mr. Shiel's, or a conventicle twang like Sir Robert's.

A parson might set up the Compound Machine in his pulpit, and a clerk or curate work it from the reading-desk, whilst his Reverence was smoking kis pipe in the vestry; or an under-secretary might set the bellows going with a speech of Lord John's, whilst his Lordship was taking his usual glass of brandy-and-water at Bellamy's; or a lawyer in full practice might set a score of them to work, and so actually attend twenty committees at a time; or it might be placed upon The Throne, with the august insignia laid upon the top of the machine, and the Lord Chancellor (after kneeling profoundly) might pop the royal speech into the proper receptacle and blow it out again to both Houses in the best style.

A clear saving of ten thousand a year might be effected by setting up a machine *en permanence* in the Speaker's chair of the House of Commons. Place the mace before it. Have a large snuff-box on the side, with rappee and Irish for the convenience of Members, and a simple apparatus for crying out "Order, order," at intervals of ten minutes, and you have a speaker at the most trifling cost, whom Sir William Gossett might keep going all night.

The elocution of the Euphonia is not at present very distinct—say about as clear as that of His Grace the Duke of Wellington. When people are not kindly warned beforehand of the words that the instrument is about to enunciate, they are generally so stupid as not to understand what it says. In our presence the Euphonia gave vent to a sentence which nobody understood but ourselves, and our hearts perhaps divined the cry. It was, "Hourrah for Fiodoria." So the machine (a German instrument) pronounced the venerated name of Her Majesty.

It sang "God preserve the Emperor" and "God save the Queen" with such clearness and eloquence, that we really felt we ought to stand up and take off our hats.

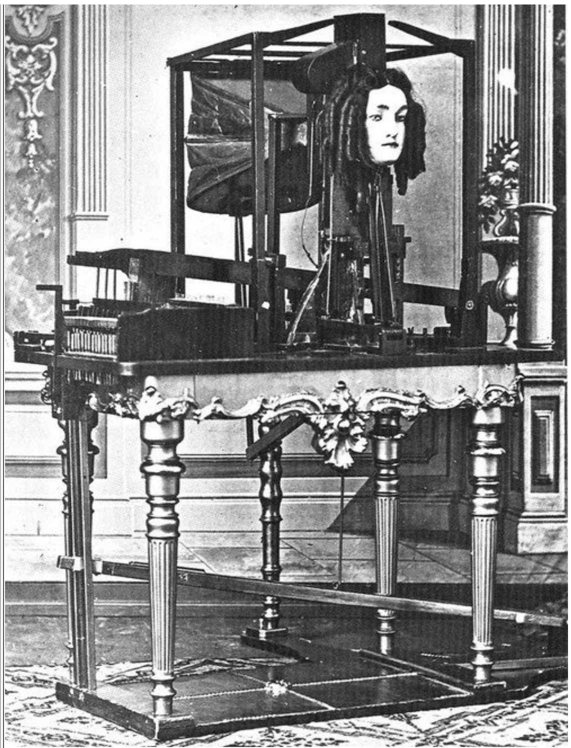
The Machine laughs—but we are bound to say not in a hearty and jovial manner. It is a hard, dry, artificial laugh; such as that of young Misses on the stage, when they give the genteel comedy-giggle; or of Sir Robert Peel, when he is amused by some of Mr. Disraeli's good-natured jokes against him.

∴ By the way, why should not Lord George Bentinck have one of these machines constructed, with a Benjamin Disraeli figure-head, and play upon it himself at once, and spare the honourable Member for Shrewsbury the bother of being his Lordship's Euphonia?

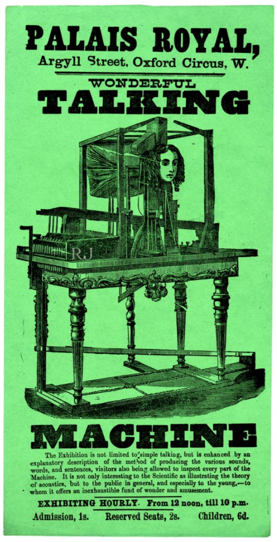
By far the best part of the Euphonia is its *hiss*; this is perfect. And perhaps the fact suggests to the benevolent mind the moral that hissing is the very easiest occupation of life,—which truth is, however, beside the present question.

Le dessin est attribué à l'auteur anglais William Makepeace Thackeray qui se moque des parlementaires Lord George Bentinck et Benjamin Disraeli. Richard Daniel Altick a repris dans son livre The Shows of London, publié en 1978, ces illustrations.

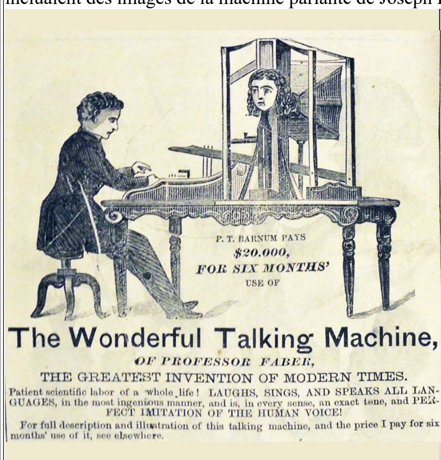
Après Londres, P. T. Barnum a exposé la machine parlante de Joseph Faber dans son musée américain des curiosités à New York. C'est là que le photographe américain Mathew Brady prenait la photo suivante vers 1860.

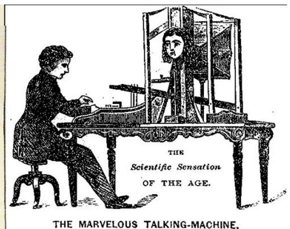


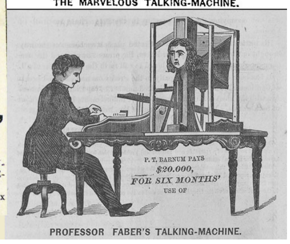
En novembre 1870, le périodique The London Journal a publié une nouvelle contribution sur la machine parlante.
L’article précise qu’une Talking Machine du Professeur Faber de Vienne est exposée au Palais Royal, Oxford-Circus, et qu’une visite vaut la peine. Il semble donc que la machine a été retournée des États-Unis à Londres. L’affiche suivante qui fait partie de la collection de Ricky Jay date probablement de cette époque.



En 1873, les affiches, courriers et annonces dans la presse concernant le nouveau cirque itinérant de P. T. Barnum, sous le nom de P. T. Barnum’s Great Traveling World’s Fair, incluaient des images de la machine parlante de Joseph Faber.







P. T. Barnum mettait la machine parlante en évidence avec les attributs wonderful, marvelous, amazing, greatest invention of modern times. P. T. Barnum se vantait même d’avoir payé 20.000 $ pour la présentation exclusive de cette attraction durant 6 mois.

Sur les publicités de P. T. Barnum pour la machine parlante du Professeur Faber on voit un jeune homme qui manipule le clavier. Il s’agit probablement du mari de la nièce de Joseph Faber. Après le décès de celui-ci le 2 septembre 1866 à Vienne, sa nièce avait hérité la machine parlante et son mari se faisait passer pour le professeur Faber. Le cirque Barnum exposait la machine jusqu’en 1875. Ensuite la nièce, avec la complicité de son mari, continuait à la présenter jusqu’en 1885 à Londres et Paris (Grand-Hôtel en 1877,

salle à proximité du théâtre Robert-Houdin en 1879) sous le nom de Amazing Talking Machine.

C'est dans cette période que A.G. Bell étudiait de son côté la reproduction de la voix et travaillait sur la télégraphie multiple.

Après 1885, on ne trouve plus trace de la machine.

Malgré tous les efforts, l'Euphonia n'a jamais dépassé le stade d'une curiosité. En réalité l'invention de Joseph Faber était vraiment une version améliorée de la machi... parlante de Wolfgang von Kempelen. Un clavier de 16 touches actionnait un mécanisme comprenant une série de six plaques de métal coulissantes qui avaient des ouvertures de formes variées à leurs extrémités. Quand ces plaques étaient soulevées ou rabaissées, elles créaient un courant d'air finement nuancé envoyé par un soufflet. L'air atteignait ensuite une cavité qui imitait l'anatomie humaine du palais avec des joues en caoutchouc, une langue en ivoire et une mâchoire inférieur. Avec cette construction Joseph Faber pouvait produire des voyelles et consonnes dans différentes langues. Il est rapporté que la machine pouvait même chanter "God save the Queen".

Parmi les visiteurs des présentations de la machine parlante de Joseph Faber figurent des personnalités comme Frédéric Chopin (lettre du 11.10.1846 à ses parents), Robert ' Patterson, Joseph Henry, Graham Bell et le Duc de Wellington.

Aujourd'hui Joseph Faber est considéré comme un vrai pionnier de la synthèse vocale.

L'analyse des articulations des sons du langage

Vers 1854, Manuel Garcia – frère de la Malibran et auteur de plusieurs ouvrages sur la voix et l'art du chant, dont Mémoire sur la voix humaine (1840) –, en se promenant dans les jardins du Palais Royal, eut l'idée de regarder ses cordes vocales par le biais de sa canne : la lumière solaire se reflétant dans le pommeau renvoyait un rayon au niveau de sa bouche. Garcia parvint ainsi à visualiser le jeu des cordes vocales. Après cette première « découverte », il plaça pour améliorer son observation un petit miroir au bout d'un long manche : ce sont les débuts de la laryngoscopie. L'ensemble des expériences de Garcia est recueilli dans ses Observations physiologiques sur la voix humaine et publié en 1855. Ses recherches sont complétées par les expérimentations sur le larynx et les cordes vocales d'un autre physiologiste, le médecin tchèque Czermack. Grâce au laryngoscope, Czermack explore en 1880 le « fonctionnement des cordes vocales et celui du voile du palais dans la production des nasales».

Les recherches sont approfondies par des physiologistes l'Autrichien Ernst von Brücke, définit les bases théoriques de cette nouvelle approche grâce à ses travaux sur l'analyse des articulations des sons du langage dans les principales langues anciennes et modernes. **Hermann von Helmholtz**, avec son *"Die Lehre von des Tonempfinungen"*, ouvrage fondamental paru en 1862, donne pour la première fois une théorie physique des voyelles et montre qu'elles se distinguent l'une de l'autre par leur timbre, d'où sa théorie de la résonance appliquée aux timbres des sons en harmoniques simultanées. À cet effet, Helmholtz met au point un instrument de mesure, les « résonateurs de Helmholtz » (des caisses de résonance sphériques ouvertes construites initialement en verre puis en laiton), fabriqué et commercialisé par Rudolph Koenig.

L'apprentissage de la langue la méthode Bell

Alexander Bell opte pour un procédé qui cherche à rendre la langue « visible » par l'utilisation d'un alphabet comportant dix symboles pour la langue, les lèvres, le larynx et les fosses nasales : le ***Visible Speech***.
Cet alphabet physiologique donnait la position des organes au cours de la prononciation et il permettait donc de transcrire « graphiquement pour chaque son du langage les composantes articulatoires qui les réalisent».
Un« Anglais, qui n'était ni physiologiste ni physicien, mais simplement professeur de diction», **Alexander Bell**, apporte sa contribution à une meilleure connaissance de l'articulation des phonèmes et des voyelles au moyen d'une étonnante méthode. Il l'explique dans un ouvrage intitulé ***Visible Speech : the Science of Universal Alphabetics, or Self-interpreting Physiological Letters for the Printing and Writing of all Languages in One Alphabet***, dont la première édition date de 1867.
Cette méthode connaîtra un grand succès dans les écoles d'Angleterre et des États-Unis et sera utilisée pendant une quinzaine d'années.

Les travaux de la famille Bell se situent à la croisée des différentes expérimentations et permettent donc de mettre à jour un certain nombre de relations qui réunissent plusieurs éléments au premier abord distincts : d'une part la phonétique, la physiologie, la surdité ; d'autre part l'acoustique, le téléphone, le phonographe et le microphone. Après avoir exercé l'activité de cordonnier, Alexander Melville Bell devient maître d'élocution au théâtre royal d'Edimbourg. C'est alors qu'il ouvre une école de diction et traitement des troubles de la parole. Il est aussi l'auteur de divers ouvrages sur le sujet. Son fils, Alexander Bell (1819-1905), dans la même lignée, est également professeur de diction. La parole joue un rôle essentiel dans sa vie professionnelle et privée, car sa femme Elisa était devenue sourde à l'âge de dix ans après une scarlatine. Son frère David était aussi professeur d'élocution dans une école de Dublin et c'est au cours de son enseignement de la diction qu'il songe à un système pour faciliter l'apprentissage de la langue et la correction des défauts de prononciation.

Le flambeau de la famille est repris par le fils d'Alexandre Melville : **Alexander Graham Bell** (1847-1922).
Celui-ci complète les études de phonétique, des langues grecque et latine, par l'acoustique et la physique. Il devient lui aussi professeur de diction et, comme son père, épouse une de ses élèves (une jeune fille devenue sourde à l'âge de cinq ans) : Mabel Hubbard, auteur d'un ouvrage sur la lecture consacré à la lecture les lèvres (The Subtil Art of Speechreading, 1895). Graham Bell poursuit donc l'œuvre de son père en travaillant lui aussi sur l'idée de la visualisation comme une solution pour l'éducation vocale de ses pensionnaires. « Il est bien connu – disait-il – que les sourds et muets ne sont muets que parce qu'ils sont sourds, et qu'il n'y a dans leur système vocal aucun défaut qui puisse les empêcher de parler ; par conséquent, si l'on parvenait à rendre visible la parole et à déterminer les fonctions du mécanisme vocal nécessaire pour produire tel ou tel son articulé représenté, il deviendrait possible d'enseigner aux sourds et muets la manière de se servir de leur voix pour parler. »

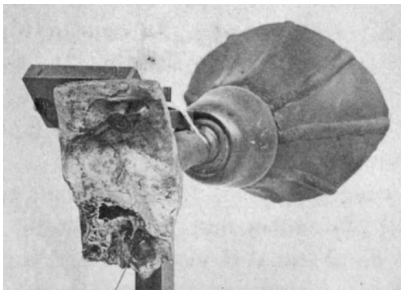
Pour comprendre la production de la voix humaine, il existe beaucoup d'ouvrages et de sites internet qui traitent sur l'organe vocal le **larynx**
Pour visualiser cet organe, voici une belle vidéo :

La représentation graphique des sons

Bell travaille à partir de 1874 sur la représentation graphique des sons produits, mais sa passion pour la physique et l'acoustique le conduira tout naturellement vers la recherche de systèmes plus avancés que ceux mis au point par ses prédécesseurs. Dans le but de visualiser la parole pour ses élèves, il reprend d'abord les expérimentations existantes, comme celles utilisant la capsule manométrique de Koenig ou le **phonautographe** de Scott de Martinville. Pour rendre ces systèmes plus efficaces, il en modifie certains composants : il remplace par exemple la membrane de Scott par une autre plus sensible, ou encore il perfectionne le phonographe d'Edison (vers 1886).

Ensuite, Bell commence à travailler sur un nouveau procédé et réalise l'appareil « à paroles visibles » (version manuelle). Ce système repose sur l'application d'une autre invention : celle de « l'**oreille artificielle** ». Graham Bell, en observant les tracés graphiques réalisés avec le phonautographe, a eu l'intuition de faire un rapprochement entre l'appareil et l'oreille humaine : « Je fus très frappé des résultats produits par cet instrument, et il me sembla qu'il y avait une grande analogie entre lui et l'oreille humaine.

Je cherchais alors à construire un « **phonautographe** » modelé davantage sur le mécanisme de l'oreille. »
Le phonautographe à oreille était un instrument macabre. Construit par Alexander Graham Bell et Clarence J. Blake en 1874, il était composé de parties d'une oreille humaine retirée chirurgicalement – un fragment de crâne, un canal auditif, un tympan et des osselets – et servait à « écrire » visuellement des ondes sonores.
Graham Bell construit son appareil en enduisant la membrane d'un tympan et d'un pavillon circulaire artificiels avec un mélange de glycérine et d'eau et en donnant à ces organes la souplesse suffisante pour qu'en chantant dans la partie extérieure de cette membrane artificielle le stylet qui lui était directement relié soit mis en vibration. Le tracé de ces vibrations était obtenu sur une plaque de verre noircie, disposée en dessous de ce stylet.



La section d'oreille humaine était fixée à la partie supérieure de l'appareil par boulon enfoncé dans le fragment de crâne. Une vis à oreilles maintenant le tout en place. L'instrument fonctionnait en canalisant les vibrations des ondes sonores produites en parlant dans l'embouchure située derrière le fragment de crâne, vers le canal exposé de l'oreille. En heurtant le tympan – sensible – situé à l'intérieur, ces vibrations déclenchaient une réaction en chaîne : le tympan vibrait d'abord, puis les osselets, suivis du stylet, un petit morceau de paille fixé au dernier os. Lorsqu'une pièce de verre recouverte d'une fine couche de suie était tirée rapidement sous le stylet vibrant, ce dernier gravait ou « écrivait » la forme des vibrations sur la surface du verre.

La reproduire les sons vocaux

Bell continue ses recherches dans cette voie et entame une étude sur les moyens de reproduire les sons vocaux en même temps que la manière de les transmettre électriquement. Sa première étape est celle de l'observation de l'oreille et l'étude de la transmission des vibrations sur le tympan d'un cadavre. Il simule alors le tympan en faisant vibrer une membrane métallique en fer-blanc près d'un barreau aimanté, entouré d'une bobine de fil de cuivre ; les variations du champ magnétique engendrées par les vibrations font naître dans la bobine un courant alternatif induit qui se transmet par un fil conducteur à la bobine du récepteur dont l'aimant fait vibrer une seconde membrane métallique de la même façon. C'était une première ébauche d'un appareil mieux connu sous le nom de téléphone. Cette invention fait écho à bien d'autres, « car le fait que Graham Bell, inventeur du téléphone en 1876, que Charles Cros et Thomas Alva Edison, inventeurs quasi simultanés du phonographe en 1877, aient tous trois, à un moment de leur vie, été éducateurs pour sourds ne doit évidemment rien au hasard.
Il y a là un rapport très étroit entre un questionnement sur la physiologie de la parole, la possibilité de l'apprentissage de la langue et la machine conçue comme artefact rédempteur.
Après l'invention du téléphone de Graham Bell, l'audiométrie connaît un large essor et par conséquent les recherches sur l'éducation auditive et vocale s'orienteront, de plus en plus, vers les méthodes « oralistes » par opposition aux méthodes basées sur la gestualité et les signes.

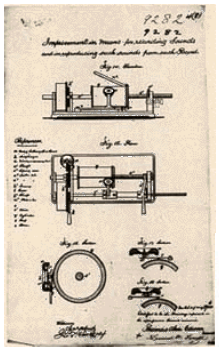
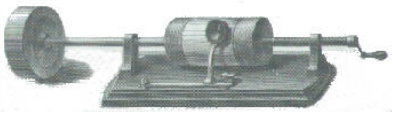
La transmission de la parole

À l'origine de ces nouvelles recherches se trouve encore la transmission de la parole.
Dès 1857, L. S. de Martinville met au point le phonotaugraphe destiné à produire des diagrammes afin d'étudier les vibrations de la voix. Cette ingénieuse sténographie acoustique n'est qu'une réussite partielle cependant. Son auteur parvient à enregistrer les sons mais échoue à les reproduire. Il lui reste cet invraisemblable désir de vaincre le temps, fixer l'immatériel, immortaliser la voix :
« *Y a-t-il possibilité d'arriver, en ce qui concerne le son, à un résultat analogue à celui atteint dès à présent pour la lumière par la photographie ? Peut-on espérer que le jour est proche où la phrase musicale échappée des lèvres du chanteur viendra s'inscrire d'elle-même, et comme à l'insu du musicien, sur un papier docile, et laisser une trace impérissable de ces fugitives mélodies que la mémoire ne retrouve plus alors qu'elle les cherche ?* »

Sur cette voie, Alexandre Graham Bell songe à traduire les vibrations mécaniques en courant électrique alternatif. Pour cela, il exploite le phénomène de la modulation d'un faisceau lumineux sous l'influence des ondes sonores. Cette modulation est obtenue soit par la variation de l'intensité du faisceau, soit par la déviation de son axe de propagation. Finalement, en 1880, Bell et son collaborateur Charles Summer Tainter appliquent ce principe en vue de la transmission directe des sons entre un poste transmetteur et un poste récepteur. Voici comment Jean Vivié présentait ce circuit dans son Traité général de technique du cinéma :
« Il s'agissait d'une transmission téléphonique entre un poste transmetteur et un poste récepteur pointés l'un sur l'autre ; le récepteur utilisait les propriétés photoélectriques du sélénium. Le transmetteur pouvait mettre en œuvre diverses méthodes ; dans le modèle primitif, les rayons lumineux étaient concentrés sur la plaque argentée d'un diaphragme téléphonique, et sous l'influence de la parole cette plaque vibrait en se bombant plus ou moins, donc en faisant varier la convergence du faisceau et par suite son intensité dans l'axe ; sous une seconde forme, la membrane téléphonique faisait vibrer une plaque légère percée de nombreuses fentes et placée en travers du faisceau parallèlement à une plaque fixe identique ; enfin les inventeurs du Photophone ont également mis en jeu dans un troisième transmetteur l'action des courants électriques microphoniques sur la lumière polarisée. »
Le photophone d'articulation, nom donné à ce système, a été présenté à l'Académie des sciences le 13 octobre 1880. Son objectif était de se servir de la lumière pour transmettre la parole à distance, grâce aux propriétés électriques du sélénium. Il comportait une lampe à arc (A), un miroir (B) réfléchissant le faisceau et concentrée par un condensateur (C) sur une membrane vibrante (D) d'un cornet téléphonique (O). Au moment de la transmission des sons, un objectif (E) envoyait vers le récepteur (M) un faisceau de lumière modulée. Le récepteur (M), constitué d'un miroir parabolique, comportait une cellule au sélénium (F) ; celle-ci recevait les variations de la lumière faisant varier à leur tour la résistance électrique de la cellule même. Enfin ces variations de courant étaient traduites en son dans deux récepteurs téléphoniques.
Si Bell et Tainter sont donc à l'origine de l'utilisation de la lumière pour la transmission des sons, ils ne pensent ni à la possibilité de fixer sur un support sensible les variations de courant, ni à utiliser le faisceau comme une « écriture » permettant de réentendre les sons. Pourtant leur application du sélénium, dont les propriétés étaient déjà connues auparavant, constitue une grande avancée pour les futures recherches sur « l'écriture des sons » par la lumière.

Le 30 avril 1877, c'est au tour de **Charles Cros** de déposer auprès de l'Académie des Sciences de Paris un pli intitulé ***Procédé d'enregistrement et de reproduction des phénomènes perçus par l'ouïe***.
Quand l'auteur propose à ses pairs le projet, fragmentaire et malheureusement jamais breveté, d'une machine appelée **paléophone** (« la voix du passé »), c'est dans l'idée d'obtenir là encore « des photographies de la voix, comme on en obtient des traits du visage et ces photographies serviront à faire parler, ou chanter, ou déclamer les gens, des siècles après qu'ils ne seront plus».
Mais dans ce cas l'analogie avec la lumière est doublement motivée puisque C. Cros lui-même avait rendu public quelques années auparavant un court texte consacré à la Solution générale du problème de la photographie en couleurs . Face à l'évanescence, à la perte et au négatif, qu'il s'agisse d'un photogramme ou d'un phonogramme, une identité doit être restituée. L'appareil doit produire de l'authentique, se faire trace ou empreinte. Le paradoxe de la technique est qu'elle se jauge selon sa capacité à perpétuer du vivant. Car la fabrication du disque phonographique chez C. Cros ne relève pas seulement d'une archéologie des voix défuntes et oubliées. Elle procède d'un désir de résurrection.
L'appareil devient un sujet second ou l'agent du sujet disparu, comme en témoigne la tournure factitive « faire parler ».

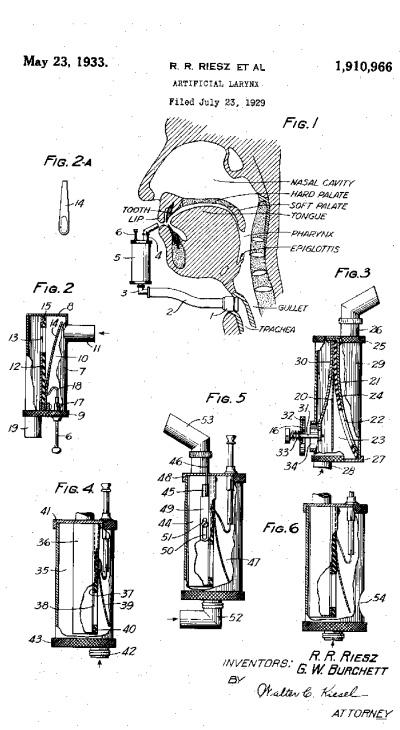
Le brevet Edison du phonographe fut accepté le 17 fèvrier 1878 et décrivait un appareil très simple.

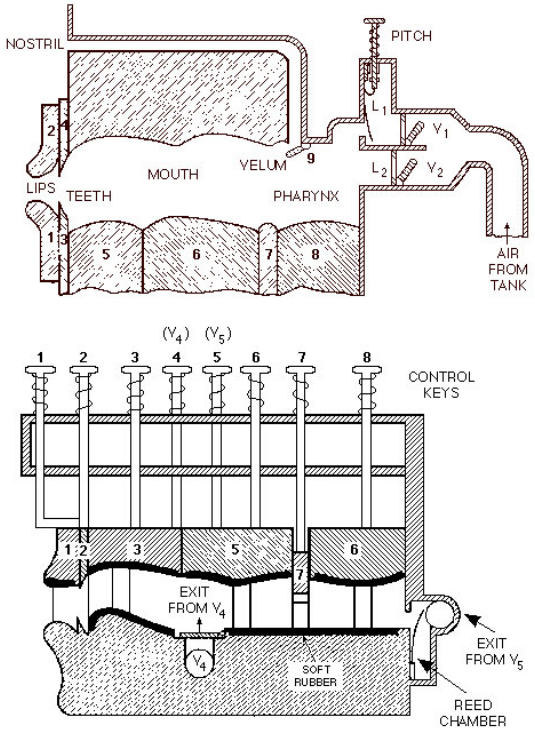


Archives Edison " The Edison papers "

Ainsi doit-on à un poète ce modèle que **T. A. Edison** parviendra à confectionner avec la complicité de J. Cruesi, huit mois plus tard sous une forme différente, cylindre, manivelle et diaphragme enregistreur capable, si on lui ajoute un pavillon, de reproduire le son cette fois.
Il n'est pas inutile de rappeler toutefois que la démonstration du phonographe Edison se solda le 11 mars 1878 par un échec devant l'Académie des Sciences. En entendant le son nasillard qui sortait de l'étrange boite, on crut d'ailleurs à une plaisanterie et les plus malveillants subodorèrent un subterfuge de ventriloque. Pourtant, assez vite, des séances d'écoute payantes sont organisées boulevard des Capucines, signe du succès populaire que rencontre l'appareil. En 1889, à la galerie des machines de l'Exposition universelle, l'objet est finalement présenté sous une version modernisée aux visiteurs puis commercialisé avant d'être détrôné en 1895 par le graphophone de C. S. Tainer et G. Bell.

Robert R. Riesz (1903 –)

Né en 1903 à New-York, Robert R. Riesz a été engagé dans les laboratoires Bell Téléphone en 1925, après avoir effectué des études en mathématiques et physique. En 1929 son innovation Artificial Larynx a été breveté au profit de Bell. Il s'agissait d'une prothèse pour des personnes ayant subi une ablation du larynx. Le projet a été décrit en 1930 dans le journal de la société acoustique américaine.





En 1937 Robert Riesz présentait un modèle réaliste de l'appareil vocal humain. Abstraction faite de quelques projets récents entrepris au Japon dans le domaine de la robotique, le modèle vocal de Robert Riesz est la dernière tentative de construire une machine de synthèse vocale mécanique.
Dans la suite Robert Riesz contribuait au développement des premiers projets de synthèse vocale électroniques.

Beaucoup d'inventions se succédèrent pour contrôler mémoriser la parole :
- Le Télégraphone
- Le Dictaphone pour enregistrer et re écouter des messages parlés.
- Le Directaphone pour parler entre pièces distantes.
- Le dictographe britannique
...

La synthèse vocale

C' est une discipline dont l'objectif est de produire de façon artificielle (mécanique ou électronique) des effets sonores imitant la voix humaine. Elle permet de convertir des textes écrits en une forme vocalisée. Les anglo-saxons parlent alors de Text to Speech ou TTS.
Outre la lecture de textes à destination de personnes malvoyantes ou non voyantes, la synthèse vocale entre notamment en application dans le cadre d'interfaces hommes machines sonores. Elle est dans ce cas utilisée conjointement avec une technologie de reconnaissance vocale, dont le but est de retranscrire un message sonore sous une forme intelligible pour l'ordinateur (qui consiste donc à réaliser l'opération inverse du point de vue conceptuel, bien que le processus soit totalement différent).
La synthèse vocale fait désormais partie de notre quotidien, elle est aujourd'hui une science informatique, Jusqu'à la fin du 19e siècle la synthèse vocale était basée uniquement sur des constructions mécaniques. Elle est présente dans nos smartphones, nos GPS ou encore dans nos salons avec les enceintes connectées. Elle est également plébiscitée pour la simplification des interactions qu'elle permet. Les voix d'Alexa ou de Google Home sont aujourd'hui très proches de réelles voix humaines. Mais obtenir un résultat aussi naturel et agréable à l'oreille a nécessité des dizaines d'années de recherche.

Suite à l'introduction du téléphone en 1876, les synthétiseurs vocaux ont évolué vers des équipements électromécaniques, pour devenir purement électriques dès la fin des années 1930, et des équipements électroniques 40 années plus tard. Pour manipuler les synthétiseurs mécaniques et électriques il fallait être un expert qui maîtrisait la génération de sons.

Pour tracer l'histoire de la synthèse vocale électrique, il faut retourner dans la seconde moitié du 19e siècle.

Après l'introduction de la télégraphie électrique, l'invention du téléphone en 1876 et du phonographe en 1877 a interrompu l'intérêt pour la synthèse vocale d'un jour à l'autre, et ceci pendant une longue période. Il convient donc de jeter également un regard sur les pionniers de la téléphonie et de la reproduction du son pour comprendre la renaissance de l'intérêt pour la synthèse vocale soixante ans plus tard, à la fin des années 1930.
Joseph Henry est un physicien américain qui découvrit l'auto-induction et le principe de l'induction électromagnétique des courants induits, ce qui contribuait à la fabrication du premier télégraphe électromagnétique.
L'invention du télégraphe électromagnétique est attribuée à Samuel Morse qui présentait sa première version opérationnelle le 6 janvier 1838 à Speedwell Ironworks p Morristown aux États-Unis.
Toutefois l'idée du télégraphe électromagnétique revient à Joseph Henry qui montrait dès 1831 aux étudiants du collège à Albany, où il enseignait, un concept d'un télégraphe qui fonctionnait avec un aimant électrique.En décembre 1845, juste avant l'exposition publique de la machine parlante de Euphonia au Musical Fund Hall à Philadelphia, Joseph Henry avait visité Joseph Faber en privé, accompagné de son ami Robert M. Patterson. Joseph Henry était impressionné par l'ingéniosité de la machine et il envisageait de transmettre des mots via le télégraphe en actionnant les touches par des électro-aimants. Avec une petite astuce, des mots générés moyennant le clavier à une extrémité du télégraphe pourraient être reproduits à l'autre extrémité.

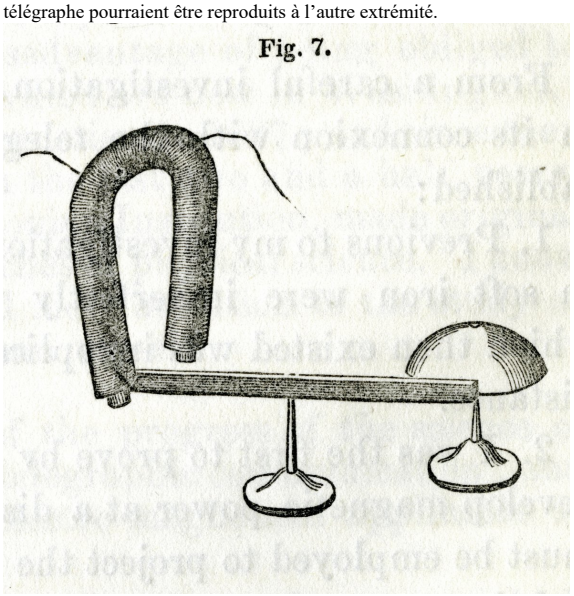


Fig. 7.

Alors qu'il avait les compétences requises, Joseph Henry n'avait jamais mis en pratique cette idée. Trente ans plus tard, en février 1875, le jeune Alexandre Graham Bell, qui avait l'idée du téléphone en tête, venait demander conseil à Joseph Henry, qui à l'époque approchait ses 80 ans. Devenu en 1846 le premier secrétaire de la Smithsonian Institution, une institution américaine de recherche et d'éducation scientifique, Joseph Henry était alors reconnu comme sommité dans le domaine de l'électromagnétisme et de la télégraphie. Il encourageait Graham Bell de persévérer et confirmait qu'il était sur le bon chemin.
En 1863, Graham Bell a été emmené par son père à une visite à Londres de la machine parlante reconstruite par Charles Wheatstone. Il était fasciné par la machine. Charles Wheatstone lui prêtait la description de Wolfgang von Kempelen et, avec son frère Melville, il réalisa sa propre version de cet automate. Dans la suite, Alexandre Graham Bell pratiqua des expériences sur la physiologie de la parole et étudia la hauteur et la formation des voyelles à l'université d'Édimbourg. La fascination par la machine parlante semble donc être à l'origine de l'intérêt pour la parole humaine qui ont conduit Graham Bell à faire breveter le téléphone en 1876.

Né le 11 février 1847 à Milan dans l'Ohio, Thomas Edison a travaillé au début comme télégraphiste. À l'âge de 19 ans il déposait son premier brevet pour l'invention d'un transmetteur-récepteur duplex automatique de code Morse. En 1874 il devenait patron de sa première entreprise grâce aux fonds récoltés pour son brevet. Avec deux associés, il dirigeait un laboratoire de recherche avec une équipe de 60 chercheurs salariés. Il supervisait jusqu'à 40 projets en même temps, et se voyait accorder un total de 1.093 brevets américains. Sa société a donné naissance à la General Electric, aujourd'hui l'une des premières puissances industrielles mondiales.
En 1877, Thomas Edison a achevé la mise au point de son phonographe, capable non seulement d'enregistrer mais aussi de restituer toute forme de sons, dont la voix humaine. Les premiers phonographes étaient munis d'un cylindre phonographique d'acier en rotation, couvert d'une feuille d'étain, et la gravure était effectuée par une aiguille d'acier. L'enregistrement était lu par la même aiguille dont les vibrations sur un diaphragme mince étaient amplifiées par un cornet acoustique.

Né en 1896 en Virginie, Homer Dudley était un pionnier de l'ingénierie acoustique. Après ses études il a joint la division téléphonie des laboratoires Bell en 1921. Quelques années plus tard le premier service téléphonique commercial transatlantique entre l'Amérique du Nord et l'Europe a été établi. C'était le 7 janvier 1927, la liaison passait par ondes radio.
Le premier câble transatlantique a déjà été posé en 1858 et il était utilisé pour la télégraphie. Le premier message a été envoyé le 16 août 1858. Le câble s'est rapidement détérioré et la transmission d'un message d'une demi-page de texte prenait jusqu'à un jour. Après 3 semaines, le câble est tombé en panne et n'a pas pu être réparé. Dans la suite d'autres câbles ont été posés qui étaient plus durables. En 1866 on arrivait à transmettre 8 mots par minute sur un câble télégraphique transatlantique. Au début du XX siècle la vitesse dépassait 120 mots par minute.
La performance des câbles télégraphiques a été améliorée par l'alliage magnétique permalloy qui compense l'induction électromagnétique. Inventé par le physicien Gustav Elmen en 1914 dans les laboratoires Bell sur base des théories de Oliver Heaviside, un nouveau procédé de fabrication de câbles sous-marin afférent a été ttiavec succès en 1923. Les nouveaux câbles télégraphiques construits avec ce procédé ont atteint des bandes passantes de 100 hertz, ce qui permettait de transmettre jusqu'à 400 mots par minute.

L'objectif de Homer Dudley était de comprimer la voix humaine de façon à pouvoir la transmettre sur un tel câble télégraphique ayant une bande passante de 100 hertz. Homer Dudley se disait que la langue, les lèvres et les autres composants du conduit vocal ne sont rien d'autre que des clés télégraphiques qui bougent à une cadence maximale de quelques dizaines de hertz. Il suffit d'analyser les principales composantes spectrales de la voix et de fabriquer un son synthétique à partir du résultat de cette analyse.

Il faut se rappeler que Joseph Henry faisait les mêmes réflexions 80 ans plus tôt.

Aujourd'hui on sait que la limite théorique minimum de débit pour un codage conservant l'information sémantique contenue dans la parole est d'environ 60 bits par seconde, si l'on compte environ 60 phonèmes dans une langue et une vitesse d'élocution moyenne d'une dizaine de phonèmes par seconde. Pour un débit aussi faible, les informations concernant le locuteur et ses émotions sont perdues.

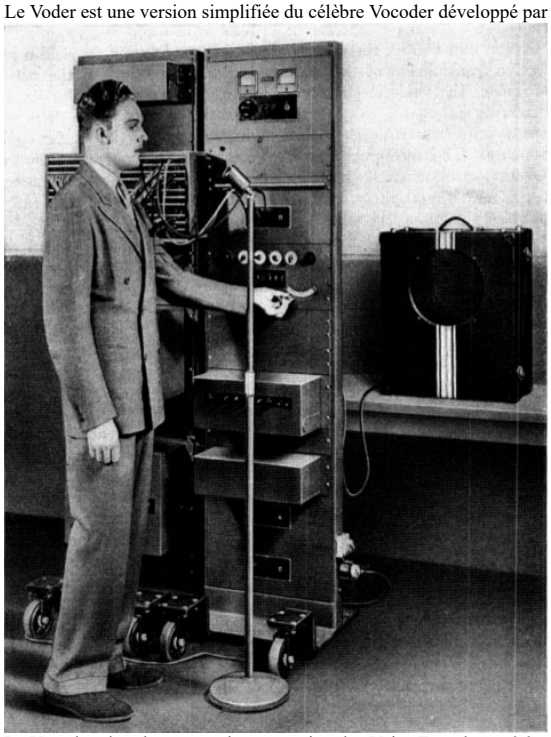
Homer Dudley a pu profiter des études effectuées par les chercheurs Harry Nyquist et Ralph Hartley qui travaillaient également dans les laboratoires Bell. Le premier est à l'origine du théorème d'échantillonnage de Nyquist-Shannon, le deuxième a contribué à la fondation de la théorie de l'information. En octobre 1928, il a esquissé les premiers circuits et édité une description de sa future invention : le **VOCODER**.

Synthèse vocale électronique : le Voder et le Vocoder

Durant tout le début du 20e siècle, plusieurs chercheurs travaillent sur la synthèse vocale. Parmi eux, les laboratoires Bell, qui vont marquer l'histoire de l'informatique et de la synthèse vocale.
De 1936 à 1939, les laboratoires Bell développent, sous la direction de l'ingénieur acoustique et électronique Homer Dudley, le premier synthétiseur vocal électronique. La synthèse vocale de cette machine se fait via une interface rappelant celle d'une machine à écrire. Les commandes étaient constituées d'un clavier ainsi que de pédales

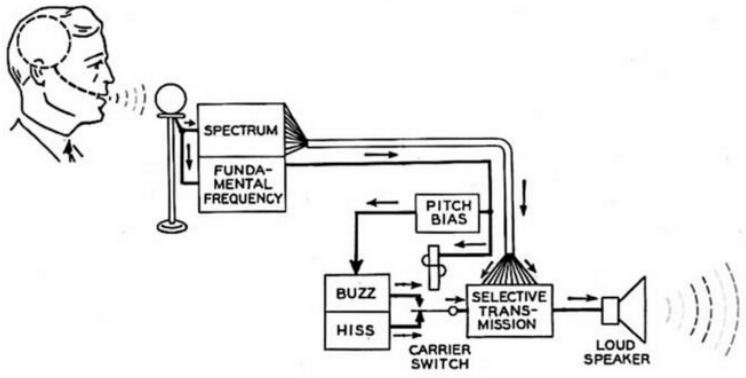
permettant de moduler les effets sonores.

Le Voder est une version simplifiée du célèbre Vocoder développé par Homer Dudley de 1926 à 1939.



Le Vocoder, dont le nom est la contraction de “Voice Encoder” a été conçu suite à la volonté des laboratoires Bell de réduire le coût des appels téléphoniques transcontinentaux. Pour cela, il effectuait une opération d’encodage du côté de la personne parlant et décodait le signal du côté de la personne qui l’écoutait. Cela permettait de faire transiter un minimum d’informations et donc d’économiser de la bande passante.
L’astuce consistait à découper le signal sonore en une multitude de plages de fréquences grâce à des filtres passe-bande. Ainsi il était possible d’analyser l’amplitude du signal de chacune de ces plages de fréquences. Ces caractéristiques étaient ensuite appliquées à une fréquence fondamentale transformée en y appliquant les modulations provenant des différentes bandes. (Pour en savoir plus, voir la présentation rédigée par Thomas Carney dans le cadre du Graduate Program in Audio and Acoustics, de l’université de Sidney)

Illustration provenant de la vidéo “The secret history of Vocoder”



La qualité de la voix reproduite par le vocodeur de Homer Dudley (aujourd’hui connu sous le nom de spectrum channel vocoder) était insuffisante pour envisager une introduction commerciale. Avec l’aide de Robert R. Riesz, Homer Dudley s’est alors focalisé sur la partie décodeur avec la parole synthétique. Le codeur a été remplacé par une console avec un clavier pour générer les paramètres de la voix. Cet appareil a été nommé VODER et breveté en 1937 au profit de Homer Dudley et des laboratoires Bell. Il a été présenté au grand public lors de l’exposition mondiale à New York en 1939.

Le Vocoder a été utilisé à partir de 1943 par l’armée américaine dans le cadre du système SIGSALY. Il a succédé au système A-3 dont les fonctionnalités de cryptage commençaient à être jugées insuffisantes pour les transmissions audio durant la Seconde guerre mondiale. Les sonorités synthétiques et métalliques du Vocoder sont désormais de notoriété publique. Elles ont en effet été réutilisées dès la fin des années 1960 dans de nombreux films (notamment pour faire parler des robots) et en musique à des fins artistiques. Il s’agit encore aujourd’hui d’un effet très utilisé dans de grands hits musicaux d’artistes comme Daft Punk par exemple.

Nous vous invitons à visionner cette vidéo anglophone publiée par The New Yorker, intitulé The Secret History of the Vocoder (L’Histoire secrète du Vocoder). Elle illustre la diversité des usages de cet appareil.

Pendant la deuxième guerre mondiale, Homer Dudley a contribué aux recherches du projet SIGSALY concernant les transmissions sécurisées sur base du vocodeur. Il est resté auprès de Bell jusqu’au début des années 1960. Le dernier projet de Homer Dudley était le développement d’un kit de synthèse vocale pour des étudiants et hobbyistes.

Fabriqué à partir de 1963, le kit a été commercialisé jusqu'à la fin des années 1960.

La recherche sur la synthèse vocale a été poursuivie dans les laboratoires Bell. Au début des années 1960 c'étaient John L. Kelly, Carol Lockbaum, Cecil Coker, Paul Mermelstein et Louis Gerstman qui étaient engagés dans le développement de systèmes de simulation de l'appareil vocal humain.
Le vocodeur inventé par Homer Dudley est devenu célèbre dans les années 1970 comme instrument de création sonore utilisé par des groupes de musique électronique Kraftwerk, The Buggles et d'autres. En réalité, Kraftwerk utilisait surtout le synthétiseur **Votrax**, inventé par Richard T. Gagnon.
En 1941, pendant la deuxième guerre mondiale, les chercheurs W. Koenig, H. K. Dunn et L. Y .Lacy des laboratoires Bell ont inventé le premier spectrographe acoustique qui convertit une onde sonore en spectre sonore.
Franklin Seaney Cooper a eu l'idée de faire le processus inverse, c'est-à-dire convertir un spectre sonore en onde sonore. Né en 1908, il a obtenu un doctorat en physique au MIT en 1936. Une année plus tôt, il a fondé ensemble avec Caryl Parker Haskins les laboratoires Haskins, une institution de recherche sans but lucratif, spécialisée dans le langage oral et écrit. Franklin S. Cooper a été le président et directeur des recherches des laboratoires Haskins de 1955 à 1975.
À la fin des années 1940, Franklin S. Cooper, assisté de John M. Borst et Caryl P. Haskins, a développé la machine appellée Pattern Playback pour convertir des spectrogrammes en paroles.
L'appareil pouvait reproduire des images de spectrogrammes réels ou bien des modèles artificiels recomposés à la main avec une peinture blanche sur une base d'acétate de cellulose. La production des sons se faisait par l'intermédiaire d'un faisceau lumineux modulé par une roue tonale et réfléchie par les parties blanches du spectrogramme. Comme le spectrogramme se déroulait dans l'appareil à une vitesse uniforme, on pouvait dessiner sur le modèle synthétique les trois paramètres fondamentaux du son : le temps (axe horizontal), la fréquence (axe vertical) et l'amplitude (épaisseur des traits)..
La machine Pattern Playback a été perfectionnée au fur et à mesure et a servi à faire de nombreuses recherches au sujet de la perception de la voix et de la synthèse et reconnaissance de la parole, notamment par Alvin Liberman, Pierre Delattre et d'autres phonéticiens.
La machine a été utilisée une dernière fois en 1976 pour une étude expérimentale réalisée par Robert Remez et elle se trouve maintenant au musée des laboratoires Haskins.

Carl Gunnar Michael Fant était professeur émérite à l'Institut royal de technologie de Stockholm (KTH). Gunnar Fant obtint une maîtrise en génie électrique en 1945. Il s'était spécialisé dans l'acoustique de la voix humaine, en particulier dans la constitution des formants des voyelles. Il a effectué ses recherches auprès de Ericsson et du MIT à ce sujet et créait un département des sciences vocales (Speech Transmission Laboratory) auprès du KTH. Gunnar Fant développait en 1953 le premier synthétiseur par formants Orator Verbis Electris (OVE-1). Des résonateurs ont été connectés en série, le premier résonateur pour le formant F0 a été modifié avec un potentiomètre, les deux autres pour les formantes F1 et F2 avec un bras mécanique. En 1960 Gunnar Fant publiait la théorie associée à ce synthétiseur, le modèle source-filtre pour la production de la parole. Après sa retraite, Gunnar Fant a continué ses recherches dans le domaine de la synthèse vocale en se focalisant notamment sur la prosodie. Pour son 60e anniversaire en 1979 ses collaborateurs avaient édité un numéro spécial du journal interne de son département.

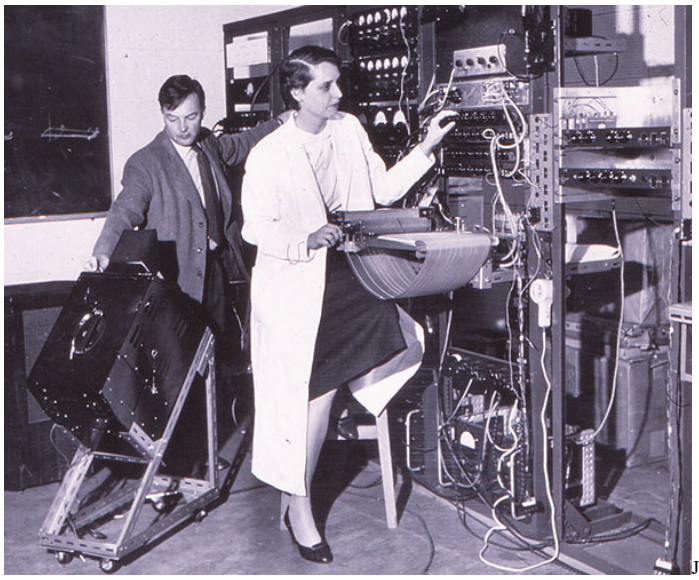
En 2000 Gunnar Fant a tenu un discours au congrès FONETIK à Skövde en Suède sur ses travaux effectués dans le domaine de la phonétique durant un demi-siècle. Il a décrit sa fascination pour la machine Pattern Playback de Franklin S. Cooper qu'il a pu voir en 1949. Son propre synthétiseur en cascade OVE II utilisait une technique similaire pour générer les signaux d'entrée qui varient les fréquences et amplitudes des oscillateurs et des générateurs de bruits afin de produire des voyelles et des consonnes pour parler des phrases complètes. Cette deuxième machine, réalisée en collaboration avec Janos Martony et qui fonctionnait avec des tubes à vide électriques, a été présentée en 1961.


Janos Martony et Gunnar Fant (à droite) avec le synthétiseur OVE-II en 1958

Une troisième version électronique du synthétisuer OVE III a été réalisée par Johan Liljencrants en 1967 avec des transisteurs. Elle a été raccordée à un ordinateur et gérée par un programme informatique.
Après sa mort en 2009, un portrait en mémoire de Gunnar Fant a été publié sur le site web du KTH.

Walter Lawrence était un ingénieur anglais qui a travaillé après ses études au SRDE (Signals Research and Developement Establishment), une institution militaire créée en 1943 (pendant la deuxième guerre mondiale) par le ministère de l'approvisionnement du gouvernement britannique. Walter Lawrence était en charge d'examiner des moyens de réduction de la bande passante pour améliorer la sécurité de l'encryption des messages militaires. Un deuxième objectif était de rendre la transmission des communications téléphoniques plus efficace et plus économique sur les grandes distances, en particulier sur les nouveaux câbles transatlantiques.
Pour ce faire, il s'est basé sur l'invention du vocodeur de Homer Dudley et il se focalisait sur la partie VODER. Il publiait en 1953 un article The Synthesis of Speech from Signals which have a low Information Rate dans lequel il proposait six paramètres pour transmettre et restituer la parole : les trois fréquences des résonances dominantes (formants), l'amplitude et la fréquence de l'excitation du larynx et l'intensité de l'excitation fricative.
Walter Lawrence appelait son synthétiseur basé sur ces principes Parametric Artificial Talker, abrégé en P.A.T. Aux fins de parfaire sa première construction réalisée en 1952 au SRDE qui n'exploita que quatre paramètres, il contacta l'université d'Edimbourg qui avait une bonne renommée dans le domaine des sciences linguistiques. Un synthétiseur qui utilisait les six paramètres décrits était opérationnel en 1956 dans les laboratoires de l'université. L'entrée des paramètres se faisait par des plaques en verre sur lesquelles on dessinait le flux des six paramètres qui étaient scannées par un faisceau lumineux.
L'installation est conservée aujourd'hui dans le musée national de l'Ecosse.

La même année (1956), Walter Lawrence et Gunnar Fant ont présenté leurs projets respectifs à une conférence au MIT à Cambridge. Le sommet de la présentation a été une conversation en temps réel entre les deux synthétiseurs P.A.T. et OVE.

 James Tony Anthony et Frances Ingemann avec le P.A.T.en 1958

Le P.A.T. a évolué en une version avec huit paramètres au début des années 1960, notamment grâce à l'apport de Frances Ingemann qui assistait le constructeur principal de la machine, James Tony Anthony, à la parfaire. Frances Ingemann a publié en 1960 une contribution Eight-Parameter Speech Synthesis dans le journal de la société acoustique américaine.
Le synthétiseur P.A.T. a connu une certaine notoriété auprès du grand public suite à la diffusion d'un film afférent à la télévision. Eye en Research était une série de télévision produite par la BBC au sujet de la recherche scientifique dans les années 1957 à 1962. La transmission se faisait en direct à partir des laboratoires de recherche avec une caméra mobile. Un journaliste de la BBC présentait et commentait le sujet. Le 28 octobre 1958 le sujet portait sur les six paramètres du synthétiseur P.A.T., développé dans les laboratoires de l'université d'Édimbourg.
Après sa retraite au SRDE en 1965, Walter Lawrence a rejoint l'université d'Édimbourg comme maître de conférences. Il est décédé en 1984.

John N. Holmes est un ingénieur en génie électrique britannique qui a obtenu son diplôme en 1953 à Londres. En 1956 il a rejoint la Joint Speech Research Unit (JSRU) qui a été formée la même année par la fusion de plusieurs agences gouvernementales en charge de la recherche dans le domaine vocal. L'intérêt de la JSRU se focalisait sur les télécommunications, aussi bien pour des applications militaires que civiles. La JSRU était affiliée administrativement au Post Office qui était responsable à l'époque pour la téléphonie.
L'analyse de la parole était l'activité principale de la JSRU. La synthèse vocale était un outil important à ces fins. John Holmes visitait en 1960 le KTH à Stockholm pour se familiariser avec le synthétiseur OVE II de Gunnar Fant. Après son retour à Londres en 1961 il a construit son propre synthétiseur désigné dans la suite comme JSRU formant synthesizer. Au lieu de la cascade de résonateurs connectés en série appliquée pour OVE, il utilisait un arrangement parallèle qui fournissait une meilleure qualité de la voix synthétique. Avec un jeu de paramètres optimisés manuellement à partir de valeurs déduites de la voix naturelle, John Holmes était capable de reproduire des phrases pour lesquelles des interlocuteurs ne pouvaient plus faire la distinction entre la parole naturelle originale et la parole synthétique.Il appelait ce principe la copie-synthèse.
En 1970 John Holmes est devenu directeur de la JSRU. À l'occasion de la conférence internationale sur la communication vocale qui a eu lieu en avril 1972 à Boston, il a présenté les conclusions de ses travaux. Il était d'avis qu'avec la technologie actuelle il suffisait d'avoir les bons paramètres d'entrée pour produire de la parole synthétique qu'on ne peut plus distinguer de la parole naturelle. Pour le prouver il faisait des démonstrations lors de la conférence. Avec des paramètres configurés manuellement (peintes à la main) les phrases prononcées étaient très proches de la parole naturelle, mais avec des paramètres extraits à partir des enregistrements de la parole naturelle la situation était une autre.
À partir du début des années 1970 les scientifiques concentraient leurs recherches sur la production automatique des paramètres de contrôle des synthétiseurs vocaux à formants. On commençait à générer les paramètres à partir de textes et la linguistique prenait un rôle de plus en plus important. En outre le but de la synthèse vocale changeait. Au lieu d'être un outil d'analyse de la voix humaine pour comprendre comment fonctionne l'appareil phonatoire humain, le nouvel objectif était de produire des interfaces pour faciliter la communication avec les ordinateurs qui faisaient leur entrée à grands pas dans les entreprises et institutions.
En 1977 John Holmes publiait l'article Extension of the JSRU synthesis by rule system, en 1984 l'article Use of flexible voice output techniques for machine-man communication. Après son départ de la JSRU en 1985 John Holmes travaillait comme consultant indépendant et il continuait à publier des articles et livres au sujet de la synthèse et reconnaissance vocale. Il est décédé en 1999. Sa dernière contribution porte le nom de Robust Measurement of Fundamental Frequency and Degree of Voicing.

Forrest Mozer est un physicien expérimental, inventeur et entrepreneur américain, reconnu pour ses travaux novateurs sur les mesures de champ électrique dans un plasma spatial. Il est en outre le développeur de circuits électroniques pour la synthèse et la reconnaissance de la parole.
Auteur de plus de 300 publications et détenteurs de 17 brevets, il a reçu de nombreuses reconnaissances pour ses travaux scientifiques. Né en 1929 à Lincoln au Nebraska, il a travaillé comme chercheur nucléaire dans différentes institutions après ses études et son doctorat. En 1970 il a été nommé professeur en physique à l'université de Californie à Berkeley.
Parmi ses étudiants se trouvait un aveugle qui demandait son assistance, ce qui incitait Forrest Mozer à s'intéresser pour la synthèse de la parole. Il a alors inventé le codage de la parole connu sous le nom de Mozer Compression pour comprimer la voix enregistrée par microphone de façon à pouvoir la reproduire moyennant les microprocesseurs à 8 bits, disponibles en mi-1970. Cette technique était un précurseur de l'ADPCM.
Le brevet américain afférent (US4214125A) a été introduit en 1974 et modifié plusieurs fois. Le brevet a été licencié d'abord à Telesensoriy Systems Inc pour l'utilisation dans le calculateur TSI Speech+, destiné aux personnes aveugles. Le circuit intégré S14001A, développé en 1975 par Silicon Systems Inc. pour ce calculateur est considéré comme le premier circuit intégré de synthèse vocale.
Dans la suite le brevet a été licencié à l'entreprise National Semiconductor qui a créé le circuit intégré DigiTalker MM54104 qui a été utilisé dans différents ordinateurs personnels, des jeux d'arcade et des interfaces de synthèse de la parole.
En 1984, Forrest Mozer a co-fondu l'entreprise Electronic Speech System pour commercialiser son système de synthèse de la parole breveté. Dix ans plus tard, il a fondé, ensemble avec son fils Todd Mozer, l'entreprise Sensory Circuits Inc., qui est spécialisée dans la reconnaissance de la parole.

Richard Thomas Gagnon était un ingénieur électronique qui a travaillé pour l'entreprise Federal Screw Works. Cette entreprise a été fondée en 1917 comme fournisseur de pièces métalliques pour l'industrie automobile. Depuis sa jeunesse Richard Gagnon était fasciné par les sciences et les technologies audio. Il s'est intéressé tôt à la synthèse de la parole parce qu'il avait des problèmes avec la vue et il craignait de devenir aveugle.
Pendant son temps libre, il développait à partir de 1970 un appareil de synthèse vocale dans son laboratoire, situé au sous-sol de sa maison à Michigan. Il s'est basé sur sa propre voix pour définir les phonèmes. Il a réalisé plusieurs versions comme prototype et comme modèle de démonstration sous les noms VS1, VS2 et VS3. Il s'agissait de synthétiseurs à formants avec architecture parallèle.
Il avait soumis une demande de brevet auprès des autorités compétentes en 1971 pour son invention qui a été approuvée en 1974, après quelques modifications, sous le numéro US 3.836.717.
Richard Gagnon licenciait le brevet à son patron qui créa la division Vocal Interface au sein de l'entreprise Federal Screw Works pour fabriquer la version VS4 du synthétiseur au nom de VOTRAX. Les circuits électroniques fabriqués étaient encapsulés par de la résine pour éviter une rétro-ingénierie du produit. Une centaine d'équipements a été vendue. La division a été renommée dans la suite en Votrax Division.
D'autres versions ont été réalisées, dont les modèles VS6.x qui ont été vendues à plusieurs milliers d'exemplaires jusqu'à la fin des années 1970. Les versions supportaient entre 32 et 128 phonèmes.

En 1980, la fabrication d'un circuit intégré SC-01 a été commandée auprès de la société Silicon Systems Inc. pour le synthétiseur vocal, suivi du circuit SC-02 en 1983. Ces

circuits ont été utilisés par des producteurs tiers pour la fabrication de terminaux parlants et de consoles arcades. A côté de l'anglais, quelques autres langues ont été supportées. À partir de 1984 le synthétiseur Votrax a été commercialisé comme carte PC sous la désignation Votalker pour les ordinateurs IBM PC, Appel II et Commodore 64. L'œuvre de Richard Gagnon est considéré comme une One Man Show. C'est le dernier inventeur dans le domaine de la synthèse vocale qui a agi seul et sous sa propre responsabilité. Il détenait 18 brevets et il était également l'auteur de quelques publications scientifiques.
En 1994, Richard Gagnon a subi un accident vasculaire cérébral et il ne pouvait plus communiquer avec ses proches, ce qui explique que ses contributions à la synthès parole étaient longtemps ignorées par la communauté scientifique et par le public.
Grâce au témoignage de la fille de Richard Gagnon, Sheila Janis Gagnon, qui a établi en 2013 l'inventaire des documents et équipements qui se trouvaient encore dans le laboratoire au sous-sol de la maison familiale, pour les remettre au musée de l'institution Smithsonian, il a été possible de donner le crédit mérité au développeur ingénieux du Votrax.

En juin 1978 Texas Instruments (TI) présentait à la CES un équipement, appelé Speak & Spell, destiné aux enfants pour apprendre l'orthographie. Ce jouet prononçait environ 200 mots qu'il fallait épeler correctement moyennant le clavier intégré.
Speak & Spell marque la transition entre la synthèse vocale électronique et informatique et constituait une étape importante (milestone) dans l'évolution des circuits intégrés.
Le premier circuit intégré a été développé par Jack Kilby auprès de Texas Instruments en 1958. Il a reçu le prix Nobel de physique en 2000 pour son invention.
Au milieu des années 1970, les ingénieurs du département Produits de Consommation de Texas Instruments cherchaient une idée pour créer un équipement grand public pour mettre en valeur les mémoires à bulles, une technologie de stockage électromagnétique miniaturisée, sans parts mouvantes, qui était en vogue à l'époque, mais qui a été abandonnée à la fin des années 1980. L'ingénieur Paul Breedlove de TI proposait le développement d'un jouet didactique utilisant la synthèse de la parole, dans la tradition de la calculette Little Professor qui a connu un grand succès, suite à son introduction par TI en 1976.
L'idée a été approuvée et une équipe projet de quatre ingénieurs a été constituée au sein de TI sous la direction de Paul Breedlove; les autres membres étaient Gene Frantz, Richard Wiggins et Larry Branntingham.

Richard Wiggings était en charge de concevoir le système de synthèse de la parole, Larry Branntingham était responsable pour le développement des circuits intégrés et Gene Frantz gérait la collaboration avec d'autres développeurs de TI qui contribuaient au projet. Il s'est avéré rapidement que la digitalisation et le stockage de mots parlés dans des mémoires bulles ou dans des ROM's (Read Only Memories) était trop couteux et qu'il fallait recourir aux méthodes de synthèse existantes à l'époque. La solution retenue était basée sur le codage prédictif linéaire (LPC) qui permet de comprimer sensiblement la voix, de stocker pour chaque mot à prononcer un nombre très limité de données et de reproduire la parole enregistrée moyennant un circuit intégré de traitement numérique du signal (DSP).
Fondamentalement il s'agissait toujours d'une simulation du fonctionnement de l'appareil phonatoire humain, sauf que les dispositifs pour synthétiser des sons à partir de paramètres spécifiques étaient mécaniques il y a 250 ans, puis électriques analogiques il y a 100 ans, dans la suite électroniques analogiques il y a 50 ans, et donc pour la première fois numériques dans le cas de Speak & Spell. Bien sûr le nombre et la qualité des paramètres de commande pour produire un son a augmenté au fil des années, mais le principe de base restait le même.

Le circuit intégré LPC développé par TI pour réaliser la synthèse vocale portait la désignation TMC0281, renommé plus tard en TMS5100. L'architecture a fait l'objet de plusieurs brevets et le circuit a été inclus dans le temple de la renommée (hall of fame) des circuits intégrés de l'IEEE en 2017, bien que ce n'était pas le premier circuit intégré de synthèse vocale. A côté de ce circuit le jouet Speak & Spell comportait un microprocesseur TMS1000 pour gérer le clavier, l'affichage, le processus LPC et l'interface avec l'usager. Les paramètres des mots à prononcer étaient stockés dans deux ROM's à 128 Kbits. C'était la capacité de ROM la plus élevée disponible à l'époque. Pour enregistrer ces mots on faisait recours à un modérateur radio, Hank Carr.
Speak & Spell comportait également un dispositif pour connecter des cartouches de jeu échangeables. Son prix de lancement était de 50 $ US. Le jouet a été commercialisé avec grand succès pendant plus de 10 ans au monde entier, en plusieurs variantes (Speak & Read, Speak & Math), avec des voix pour différents langages et avec de nombreuses cartouches de jeux. En France, le jouet a été vendu sous le nom de Dictée magique. En 2009 Speak & Spell a été nommé comme milestone IEEE.
En 1982, l'équipement Speak & Spell a été utilisé par E.T. dans le film culte de Steven Spielberg pour téléphoner avec les habitants de sa planète. C'est une des raisons pourquoi le jouet lui-même est devenu un objet culte qui bénéficie encore aujourd'hui d'une large communauté de fans qui maintiennent des émulateurs du synthétiseur, des cartouches de jeu et des instructions comment pirater le système pour l'utiliser à d'autres fins.
Le circuit intégré TMC0281 et ses successeurs ont également été vendus séparément par Texas Instruments sous la désignation TMS5100, etc. Ils ont été implémentés dans des ordinateurs personnels, des consoles de jeu et des voitures. La technique LPC constituait la base pour le codage et la compression de la voix dans les terminaux GSM au début des années 1990 et elle est encore utilisée aujourd'hui

Synthétiseurs vocaux informatiques

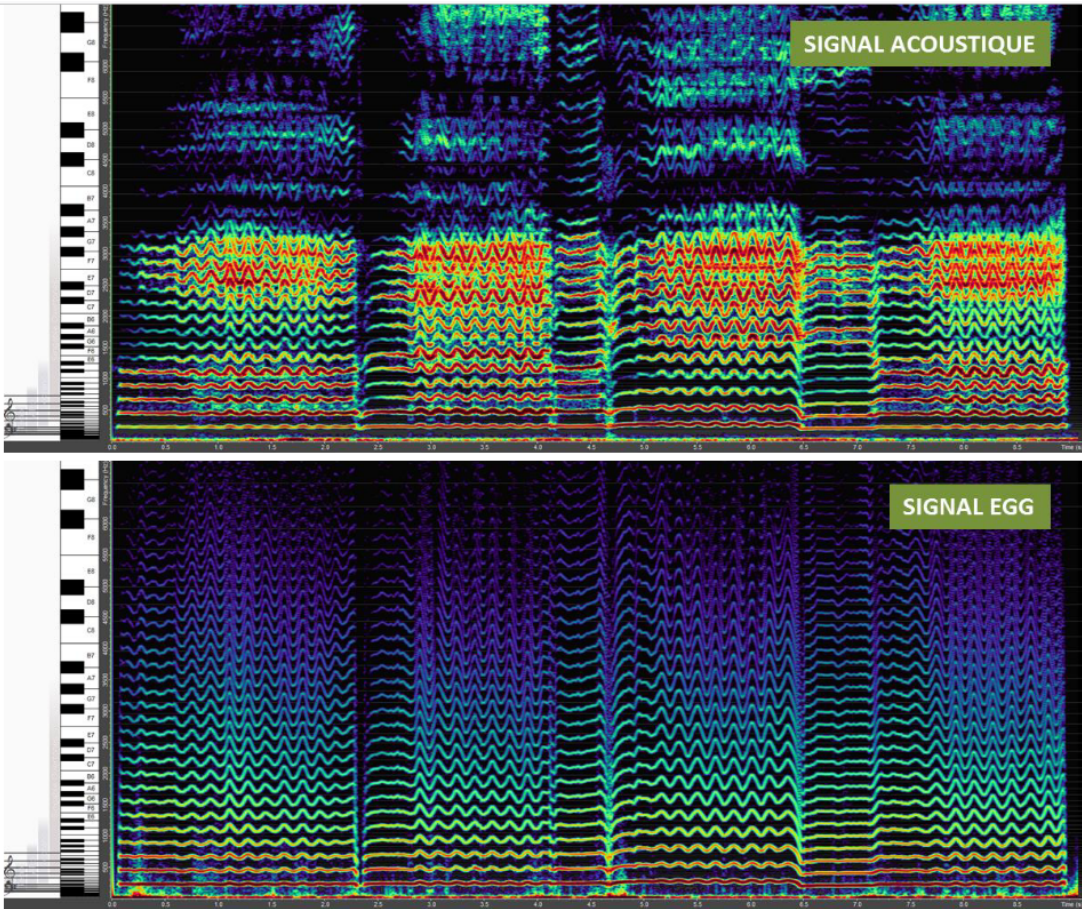
Le lecteur attentif a certainement compris que les inventeurs des synthétiseurs vocaux mécaniques et électriques ont tous essayé de simuler le fonctionnement de l'appareil phonatoire humain. Bien qu'il y a eu des progrès dans la qualité de la voix synthétique au fil du temps, le principe de base est resté le même. Il n'est donc pas étonnant que le portage de la synthèse vocale sur les premiers ordinateurs est resté au début dans la même lignée. Les oscillateurs et résonateurs sont devenus numériques et ont été réalisés moyennant des programmes informatiques pour générer les formants des voyelles et les sons des consonnes. Au lieu d'une cascade sérielle ou d'un arrangement parallèle on distinguait la synthèse sonore additive, soustractive ou FM. Comme la génération des sons est commandée par les valeurs de plusieurs paramètres on parle de synthèse par règles.

L'évolution des ordinateurs avec des processeurs de plus en plus puissants et des capacités de stockage de plus en plus grandes a permis de remplacer la simulation classique de l'appareil phonatoire humain par des procédés plus modernes. Une première méthode consistait à enregistrer des bribes de parole (phonèmes) et de les enchaîner. Les phonèmes sont les éléments minimaux de la parole. La langue luxembourgeoise comporte par exemple 55 phonèmes, en tenant compte des mots étrangers. Mais comme la parole est un processus temporel continu, les phonèmes sont trop courts et inappropriés pour obtenir une synthèse vocale par concaténation de qualité acceptable. La prochaine étape consistait à utiliser des diphones qui sont définis comme la portion du signal de parole comprise entre les noyaux stables de deux phonèmes consécutifs. Le nombre de diphones est théoriquement le carré du nombre de phonèmes, c'est-à-dire 3.025 pour la langue luxembourgeoise. Mais comme certaines transitions n'ont pas lieu en pratique, on obtient environ 2.000 diphones à considérer. Si on utilise des unités plus longues que les diphones, on parle de synthèse par sélection d'unités. Ces unités sont extraites automatiquement par des programmes utilisant des méthodes statistiques, à partir d'enregistrements vocaux avec annotation des textes correspondants, pour les sauvegarder dans une base de données indexée. Le nombre d'unités peut dépasser les 15.000 pour une voix ...

Exemples de nouveaux outils informatiques

L'assistant personnel intelligent ALEXA et l'enceinte y associée ECHO, annoncés par Amazon en novembre 2014, sont probablement les outils vocaux les plus connus par le grand public. Il est moins connu que Amazon propose aux développeurs dans son portefeuille des services cloud AWS des outils très performants en relation avec la parole : Amazon Polly, un service qui transforme le texte en paroles réalistes, Amazon Lex, un service de reconnaissance automatique de la parole (RAP) et de compréhension du langage naturel (CNL), Amazon Translate, un service de traduction automatique neuronale offrant des traductions linguistiques rapides, Alexa Voice Services, un service de synthèse et reconnaissance vocale et Alexa Skills Kit, un service d'intelligence derrière les appareils de la gamme Amazon Echo.

Bien souvent, en cabinet, l'orthophoniste n'a pas à disposition les outils technologiques pour analyser la voix de son patient, il s'appuie sur la puissance informative de son écoute analytique. Néanmoins, comment garder trace de cette écoute d'une séance à l'autre ?
Comment vérifier objectivement les ressentis subjectifs ?
A peu de frais, il est possible de s'équiper d'un système d'**enregistrement des signaux audio** : un microphone adapté, une carte son, un ordinateur. Le choix du matériel est important, en particulier celui du microphone. Suivant les recommandations publiées par Svec et Granqvist (2010), la réponse en fréquence du microphone doit être plate (à 2dB près) dans la zone de fréquence d'intérêt (dans l'idéal de 20Hz à 20kHz), la dynamique appropriée pour permettre l'enregistrement sans distorsion des productions les plus sonores et le rapport signal sur bruit suffisamment élevé (au moins 15dB) pour permettre l'enregistrement des productions les moins sonores.



Analyse temps-fréquence du signal acoustique en sortie des lèvres et du signal électroglottographique (EGG) correspondant, à partir du logiciel OvertoneAnalyzer. Phrase chantée par un baryton ("ave maria")**.**

L'analyse la plus simple à effectuer à partir d'un enregistrement du signal audio est de représenter visuellement le son par les fréquences acoustiques qu'il contient et leur évolution au cours du temps, comme le montre la figure ci dessus.
L'oreille humaine est intégrative et elle ne permet pas toujours de distinguer avec précision les zones fréquentielles où l'énergie acoustique est renforcée ou atténuée. L'analyse temps-fréquence d'un son met en évidence les fréquences qui le constituent, leurs niveaux d'amplitude et leurs variations temporelles. Cette représentation visuelle d'un son est appelée spectrogramme ou sonagramme. De nombreux logiciels permettent cette visualisation de façon plus ou moins automatique. Ils permettent de mesurer, sur le signal audio, des paramètres acoustiques d'intérêt pour l'analyse de la voix parlée ou chantée, interprétables pour un clinicien et complément indispensable de l'analyse perceptive. Des caractéristiques acoustiques de la voix dans la parole peuvent être objectivées, telles la fréquence fondamentale, l'intensité vocale, la coordination pneumophonatoire, les fréquences formantiques et la richesse harmonique.
Quel logiciel choisir ? Le logiciel WaveSurfer est un logiciel gratuit et simple d'utilisation pour visualiser et analyser le son, la fréquence fondamentale et la richesse harmonique. Le logiciel Praat est également un logiciel gratuit d'édition et d'analyse du son, mais il diffère par la complexité de son usage. Une connaissance préalable de l'outil est nécessaire pour pouvoir en faire un bon usage. Une fois maîtrisé, le logiciel Praat est un outil complet et paramétrable par l'utilisateur. Conçu pour l'analyse phonétique de la parole, il permet l'annotation des corpus. Le logiciel Overtone Analyzer, développé initialement comme un outil de pédagogie vocale, se distingue par une interface très conviviale complétée d'un clavier et d'une portée musicale, pour un coût modéré . Il présente l'avantage de pouvoir filtrer visuellement des fréquences harmoniques dans le signal analysé pour un travail ciblé sur l'écoute du timbre par exemple.
Le premier paramètre d'importance est la fréquence fondamentale, qui renseigne sur la hauteur de la voix, sa stabilité au cours de la production, sa plage de variabilité. La fréquence fondamentale se mesure sur les parties voisées du signal acoustique, c'est-à-dire pour la production vocale qui met en jeu la vibration des plis vocaux. Sa définition et

son calcul requièrent une stabilité de la durée du cycle vibratoire glottique sur plusieurs cycles consécutifs. Quand cette durée est modifiée de façon notable d'un cycle glottique à l'autre lors de la production de voix pathologique, la mesure de fréquence fondamentale perd de son sens. Il peut être alors intéressant de comparer les durées de cycles glottiques successifs. C'est ce que propose le paramètre vocal connu sous le nom de jitter, qui représente une mesure des perturbations à court terme de la fréquence fondamentale du signal sonore exprimée en pourcentage. Le jitter se calcule comme le rapport entre la moyenne de toutes les différences de durées entre deux cycles glottiques successifs (en valeur absolue) et la durée moyenne d'un cycle. Selon le manuel du logiciel Praat, le seuil normal/pathologique de jitter est fixé à 1,04%.
Un jitter élevé reflète une variabilité importante dans la durée du cycle glottique. Un autre paramètre de perturbation, le shimmer, reflète les perturbations à court terme l'amplitude du signal sonore. La moyenne des différences entre l'amplitude maximale de deux cycles glottiques successifs (en valeur absolue) est divisée par la moyenne des amplitudes maximales de chaque cycle. Le seuil normal/pathologique est fixé à 3,81 %. Ces deux paramètres de perturbation sont mesurés lors de la production d'une voyelle tenue. La pertinence de ces mesures dans l'analyse des
voix pathologiques est souvent questionnée (Bielamowicz et al., 1996). Comme le soulignent Baken et Orlikoff (1997), les mesures acoustiques de la voix, et en particulier les mesures de perturbation, ne présentent pas de corrélation cliniquement utile avec des catégories de troubles vocaux spécifiques.
Elles ne permettent en aucun cas le diagnostic. La capacité à parler fort est reflétée par la mesure de l'intensité moyenne. Seule une intensité calibrée, indépendante du volun d'enregistrement, permet une mesure comparative entre enregistrements.
La coordination pneumo-phonatoire peut être évaluée à travers la mesure du temps maximum de phonation sur une voyelle (TMP en moyenne de 15s pour les femmes et de 20s pour les hommes), le rapport de durée de la consonne sourde /s/ divisé par celui de son équivalent sonore /z/ (rapport équivalent à 1 dans le cas d'une coordination optimale).
Le rapport de durée entre parties voisées et parties non voisées est également informatif de l'usage vocal du sujet ou du patient.
L'analyse du signal audio mesuré en sortie des lèvres permet aussi d'estimer la fonction de transfert acoustique du conduit vocal et d'en déduire les fréquences et largeurs de bande des formants. Les formants sont des zones spectrales d'énergie renforcée par l'action de résonance des cavités qui constituent le conduit vocal. Leur positionnement conditionne notre perception des voyelles. Pour l'analyse formantique, le logiciel Praat est le logiciel d'analyse de la voix le plus approprié, car il permet de tracer l'évolution des fréquences formantiques sur l'analyse spectrographique du signal.
D'autres paramètres de timbre reflètent la richesse harmonique, à travers la mesure de rapports d'amplitude entre les différents harmoniques d'un son voisé.

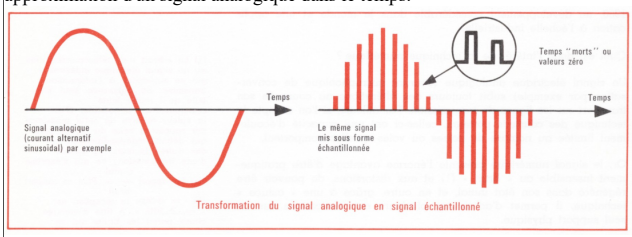
Il y a très certainement un intérêt à s'inspirer des approches développées dans le milieu scientifique pour effectuer des mesures objectives du comportement vocal d'un patient. Même si certains paradigmes expérimentaux nécessitent un équipement sophistiqué et coûteux, c'est le cas de l'IRM par exemple, de nombreux protocoles reposent sur des évaluations parfaitement réalisables en cabinet orthophonique. L'usage du logiciel Praat qui se répand de plus en plus dans la pratique orthophonique en est un bel exemple illustratif. Il permet à la fois de prendre les données et de les analyser. Corrélée à l'analyse perceptive de la voix du patient et à l'auto-évaluation de la qualité de voix, l'analyse acoustique apporte des éléments quantitatifs nécessaires à l'abord de la pathologie vocale. L'analyse de scènes vidéo permet de conserver une trace de l'évaluation du patient et d'évaluer à posteriori les gestes posturaux, respiratoires et articulatoires du patient.

Nous vivons à l'ère du numérique et les signaux analogiques captés par ces différents outils de mesure sont convertis en signaux numériques avant d'être sauvegardés sur un ordinateur ou dêtre transmis à l'autre bout du monde avec nos téléphones mobiles.
Cette opération de conversion analogique/numérique a un impact sur les signaux qu'il est important de connaître.
Le premier aspect de cette conversion est l'**échantillonnage** du signal : il existe une durée non nulle entre deux mesures successives. La fréquence de prise de mesure, qu'on appelle fréquence d'échantillonnage, va définir la précision de l'information enregistrée. Plus cette fréquence sera élevée, plus les variations rapides du signal (fluctuations hautes fréquences) seront prises en compte.
Dans le cas d'un signal audio de parole, il est nécessaire d'avoir de l'information fréquentielle dans toute la bande audible, donc de préférence jusqu'à des fréquences de 16kHz à 20kHz. Ceci impose d'avoir une fréquence d'échantillonnage au moins deux fois supérieure à la fréquence limite d'intérêt (Théorème de Shannon). Les cartes d'acquisition proposent des fréquences d'échantillonnage à 44,1kHz ou 48 kHz, ce qui permet de couvrir la gamme des fréquences audibles. Si ces fréquences d'échantillonnage conviennent bien à l'enregistrement de signaux audio, il n'est parfois pas nécessaire de recourir à une telle précision temporelle pour des signaux qui évoluent lentement au cours du temps. Certains signaux, les signaux de débit ou de pression aérodynamique par exemple, ne demandent pas de fréquence d'échantillonnage très élevée car ils évoluent lentement au cours du temps.
Le second aspect de cette conversion est la *quantification* du signal : le signal est décrit par une quantité finie de valeurs du fait de la capacité de codage (généralement sur 16 bits). La quantification entraîne, comme l'échantillonnage, une perte de données et un bruit éventuel (bruit de quantification). La quantification des données entraîne une imprécision sur les données inhérentes à ce processus.
Cette imprécision peut également dépendre de l'outil de mesure, des conditions d'acquisition. Aucune mesure ne permet d'approcher la réalité de façon exacte. Evaluer la précision d'une mesure et l'intervalle d'incertitude reflète la qualité d'une approche expérimentale, gage d'une démarche
scientifique rigoureuse et réfléchie. Nombreuses sont les études qui donnent des mesures à 2 ou 3 chiffres après la virgule, alors que l'outil de mesure ne permet pas, et de loin, une telle précision.
L'évaluation de l'incertitude d'une mesure nécessite de connaître les caractéristiques de précision de l'outil de mesure et celles de la conversion analogique-numérique. Il est à mentionner ici que la sauvegarde de données sous des formats compressés, comme par exemple l'encodage mp3 de signauxaudio, est à proscrire car il y a toujours une perte d'information dans ces encodages.

Cette technique est toujours employée dans la téléphonie d'aujourd'hui.

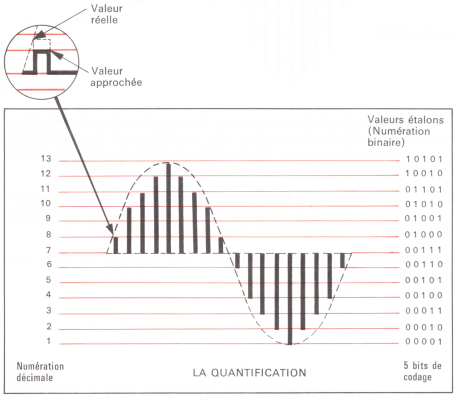
Voici un rappel quelques informations essentielles sur l'utilisation des techniques numériques en téléphonie à propos du multiplexage 32 voies appelé par la suite **système MIC (Modulation par Impulsion et Codage).**

Le théorème d'échantillonnage de Nyquist stipule qu'un signal analogique à bande passante limitée peut être représenté pratiquement parfaitement si le signal est échantillonné à une fréquence de deux fois la bande passante. Un signal vocal est considéré comme ayant une bande passante d'environ 3 kHz (300 Hz à 3 400 Hz) et donc, en principe, pourrait être représenté par une séquence d'impulsions résultant d'un échantillonnage à environ 6 kHz. Des considérations pratiques dictent cependant l'utilisation d'une fréquence d'échantillonnage plus élevée et en téléphonie 8 kHz est devenu la norme. L'amplitude des impulsions échantillonnées est quantifiée (logarithmiquement) et représentée par un nombre binaire à huit chiffres, sept chiffres indiquant le niveau et le huitième le signe. Un canal téléphonique numérisé est alors un flux binaire de 64kb/s dans chaque sens. Le simple remplacement d'un canal analogique par un canal numérique offre très peu d'avantages mais l'approche numérique permet l'utilisation du multiplexage temporel (TDM) et il est possible de multiplexer jusqu'à 30 canaux, formant un flux binaire de 2Mb/s.
- Étape 1 : Échantillonnage.
C'est un peu e même principe que le cinéma 24 photos par seconde suffisent pour tromper l'oeil et voir la scéne avec une bonne fluidité.
En téléphonie classique avec des téléphones basiques qui existent depuis l'invention du téléphone, les signaux analogiques vocaux (ainsi que les tonalités transmises) d'une conversation en cours entre deux abonnés sont tout d'abord échantillonnés à la fréquence de 8.000 Hz. (Un échantillon vocal est prélevé et mesuré toutes les 125 µs. Ceci signifie que l'on effectue 8.000 mesures de tension à chaque secondes.)
Un tel échantillonnage permet de pouvoir reconstituer à chaque extrémité de la chaîne de commutation et de transmission les conversations de manière fidèle jusqu'à une fréquence maximale audible de 4.000 Hz, limite suffisante pour reconstituer des conversations en cours qui soient compréhensibles. L'échantillonnage est en fait une approximation d'un signal analogique dans le temps.

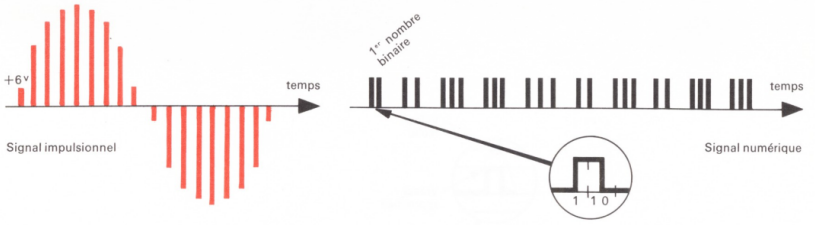


- Étape 2 : Quantification.
Une fois les échantillons vocaux prélevés toutes les 125 µs, il est nécessaire de procéder à une seconde approximation : l'approximation en niveau de tension.
En effet, un signal analogique étant susceptible de prendre une infinité de valeurs entre une tension A et une tension B, cet aspect impose de réduire les valeurs de tensions possibles de ces échantillons en un nombre limité de valeurs-étalons. La valeur de sortie de l'étage de quantification est la valeur-étalon de référence la plus proche de la valeur réelle de la tension d'échantillonnage d'entrée.

Il a été retenu, en norme téléphonique, que les niveaux de tensions échantillonnées seraient compris entre 256 niveaux de tensions différents (256 valeurs-étalons). (Chaque échantillon est donc systématiquement arrondi en une valeur numérique comprise entre une valeur comprise entre 0 et 255.)
Une telle quantification, même s'il ne s'agit pas de Haute-Fidélité telle que l'on pourrait la qualifier en acoustique, permet en norme téléphonique, le codage de suffisamment d'états d'amplitude possibles des signaux vocaux.



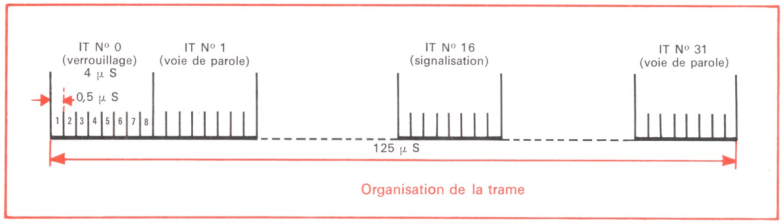
Étape 3 : Codage.
Puis ces échantillons vocaux, qui peuvent prendre 256 valeurs différentes sont convertis en numération binaire (en base 2) sur des mots d'une longueur de 8 bits. À partir de là, les échantillons sont devenus des nombres exprimés en base 2, c'est à dire par un nombre au format de 8 chiffres, dont chaque chiffre peut prendre la valeur 0 ou 1.
Comme ces signaux codés sont échantillonnés à la fréquence de 8.000 Hz, sur un mot binaire de 8 bits, le débit équivalent en éléments binaires par secondes (e.b/s) sera de 8.000 Hz x 8 bits = 64.000 bits/s. Bit se traduit par Élément Binaire : 0 ou 1.

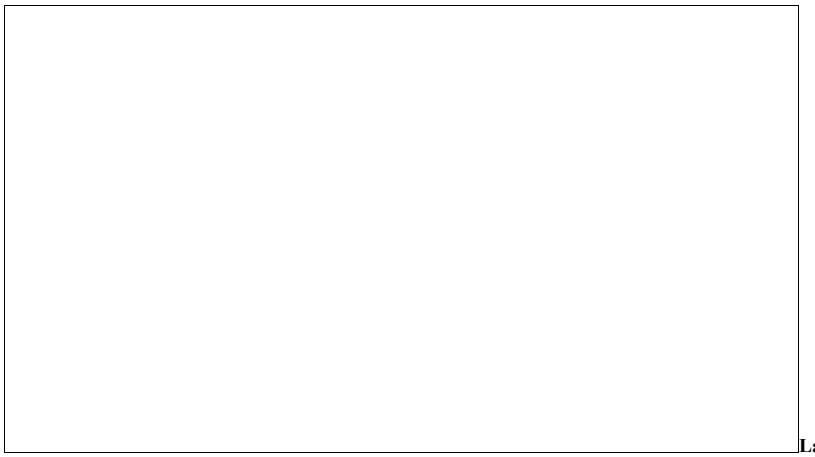


Il serait déjà avantageux de réaliser des transmissions sur de longues distances sous forme numérique, car l'intérêt premier serait de pouvoir amplifier de manière peut coûteuse la liaison numérisée, étant donnée que nous savons à l'avance qu'à un instant donné, la valeur théorique transportée est soit égale à 0, soit égale à 1. Par contre, nous ne pourrions transporter sur de longues distances qu'une seule voie téléphonique simultanément, ce qui finalement ne s'avérerait pas très avantageux... Il faut donc trouver un moyen supplémentaire.
Le Multiplexage Numérique.
Lorsque nous avons échantillonné à chaque instant T, toutes les 125µs, en fait, cet instant T a duré 3,90µs. (durée fixée par les normes téléphoniques : il faut l'instant le plus court possible, mais tout en gardant une durée suffisamment longue de sécurité, eu égard aux tolérances des composants électroniques, qui eux, sont bien réels, et ne sont pas des formules mathématiques parfaites...)
Donc, sur une liaison numérique, nous voyons qu'il y a un temps mort de 125µs - 3,90µs = 121,10µs.
Puisqu'il existe un si grand temps mort entre deux échantillons numériques vocaux, pourquoi ne pas y insérer d'autres échantillons vocaux émanant d'autres conversations téléphoniques ?
Ainsi nous pourrions transmettre sur une même liaison numérique 125µs/3,90µs = 32 conversations téléphoniques numérisées à la fois ! En fait, si la durée d'échantillonnage est de 3,90µs, nous avons 32 Intervalles de Temps disponibles (IT) pour faire circuler à la fois successivement et simultanément 32 conversations téléphoniques.
C'est ce que l'on appelle le Multiplexage Numérique : à partir d'une simple liaison numérique, nous pouvons acheminer simultanément 32 voies téléphoniques, de quoi faire disparaître la pénurie de capacités de voies de transmissions de conversations, en réutilisant les liaisons métalliques existantes, qui ne peuvent acheminer en basses fréquences qu'une seule conversation à la fois...
Le Multiplexage Numérique est en fait un système Multiplex à répartition dans le temps.

Ces signaux numérisés sous forme de mots binaires de 8 bits, émanant d'une conversation en cours, avec un débit binaire de 64.000 bits/s, sont ensuite insérés dans une voie d'un Circuit MIC, et ce côte à côte avec d'autres signaux provenant d'autres conversations en cours. (jusqu'à 30 conversations téléphoniques simultanées peuvent circuler sur une même liaison MIC.)
Un Circuit MIC est équipé de 32 voies, car une Liaison MIC est "découpée" en 32 Intervalles de Temps de 3,90µs chacun.
Mais seulement 30 voies sont en réalité réservées au transport des conversations téléphoniques, car 2 voies sont notamment affectées à la synchronisation et au contrôle d'erreur. En effet, parmi les 32 voies, numérotées de 0 à 31,
- la voie 0 est destinées à la synchronisation : qui doit permettre d'indiquer aux équipements de multiplexage (ou de démultiplexage) quel est le premier Intervalle de Temps parmi les 32 possibles,
- la voie 16 est destinée par convention à l'échange de signaux de signalisation (dialogues) entre équipements téléphoniques, pour permettre l'aiguillage des conversations, le contrôle d'erreurs etc...
Le risque de diaphonie (mélange) entre plusieurs conversations est quasiment inexistantant.

Une fois multiplexés, les signaux des 30 voies de conversations téléphoniques sortent sur une Liaison M.I.C.



 **La reconnaissance vocale**

Les synthétiseurs vocaux électroniques et informatiques disposaient d'une interface permettant d'entrer du texte et de le convertir en parole. On commençait à parler de synthèse de la parole et de TTS (Text to Speech). L'utilisation de ces équipements était à la portée de tout le monde, tandis que les développeurs de ces machines devaient avoir des connaissances approfondies en linguistique, phonologie, phonétique et sémantique.
À partir des années 1980 les scientifiques qui faisaient des recherches dans le domaine de la synthèse vocale, respectivement de la synthèse de la parole, se sont focalisés uniquement sur l'informatique. Depuis quelques années c'est l'intelligence artificielle qui domine le domaine. L'apprentissage approfondi permet aux ordinateurs d'apprendre à parler de la même manière que les petits enfants, sans se soucier de grammaire, d'orthographe ou de syntaxe.

Les travaux sur la reconnaissance de la parole datent du début du XXe siècle.
Le premier système pouvant être considéré comme faisant de la reconnaissance de la parole date de 1952.
Ce système électronique, développé par Davis, Biddulph et Balashek aux laboratoires Laboratoires Bell, était essentiellement composé de relais et ses performances se limitaient à reconnaître des chiffres isolés.

En 1952, alors que la recherche financée par le gouvernement américain prenait de l'ampleur, les laboratoires Bell développèrent un système de reconnaissance automatique de la parole capable d'identifier les chiffres de 0 à 9 prononcés au téléphone.
Des progrès majeurs suivirent au MIT. En 1959, un système identifia avec succès les voyelles avec une précision de 93 %. Sept ans plus tard, un système doté d'un vocabulaire de 50 mots fut testé avec succès.
Au début des années 1970, le programme SUR donna ses premiers résultats substantiels. Le système HARPY, à l'université Carnegie Mellon, pouvait reconnaître des phrases complètes constituées d'un nombre limité de structures grammaticales. Mais la puissance de calcul nécessaire était prodigieuse ; il fallait 50 ordinateurs contemporains pour traiter un canal de reconnaissance.

La recherche s'est ensuite considérablement accrue durant les années 1970 avec les travaux de Jelinek chez IBM (1972-1993).
La société Threshold Technologies fut la première à commercialiser en 1972 un système de reconnaissance d'une capacité de 32 mots, le VIP100. Aujourd'hui, la reconnaissance de la parole est un domaine à forte croissance grâce à la déferlante des systèmes embarqués.

Une évolution rapide :

1952 : reconnaissance des 10 chiffres par un dispositif électronique câblé.
1960 : utilisation des méthodes numériques.
1965 : reconnaissance de phonèmes en parole continue.
1968 : reconnaissance de mots isolés par des systèmes implantés sur gros ordinateurs (jusqu'à 500 mots).
1970 : Leonard E. Baum met au point le modèle caché de Markov, très utilisé en reconnaissance vocale1.
1971 : lancement du projet ARPA aux États-Unis (15 millions de dollars) pour tester la faisabilité de la compréhension automatique de la parole continue avec des contraintes raisonnables.
1972 : premier appareil commercialisé de reconnaissance de mots.
1978 : commercialisation d'un système de reconnaissance à microprocesseurs sur une carte de circuits imprimés.
1983 : première mondiale de commande vocale à bord d'un avion de chasse en France.
1985 : commercialisation des premiers systèmes de reconnaissance de plusieurs milliers de mots.
1986 : lancement du projet japonais ATR de téléphone avec traduction automatique en temps réel.
1993 : Esprit project SUNDIAL2
1997 : La société Dragon lance « NaturallySpeaking », premier logiciel de dictée vocale.
2008 : Google lance une application de recherche sur Internet mettant en œuvre une fonctionnalité de reconnaissance vocale
2011 : Apple propose l'application Siri sur ses téléphones3.
2017 : Microsoft annonce égaler les performances de reconnaissance vocale des êtres humains4.
2019 : Amazon lance la reconnaissance vocale en consultation de médecine5
2023 : Nabla lance la transcription puis synthèse de consultation6 et classification CISP2 CIM10 du résultat de consultation.

Depuis 2024, de nombreux logiciels de transcriptions utilisent l'**intelligence artificielle : l'IA**

Les systèmes de reconnaissance vocale modernes utilisent des modèles du langage qui peuvent nécessiter des gigaoctets de mémoire ce qui les rend impraticables, en particulier sur les équipements mobiles. Pour cette raison, la plupart des systèmes de reconnaissance vocale modernes sont en fait hébergés par des ***serveurs distants*** **SVI** et nécessitent une connexion internet et l'envoi à travers le réseau du contenu vocal.
- Cortana (Microsoft)
- Siri (Apple)
- Google Now (Google)
- Alexa (Amazon)
- Vocapia Research (VoxSigma suite)
- Vocon Hybrid et Dragon (respectivement dictée par grammaire et dictée libre par Nuance Communications)
- LinTO (logiciel libre développé sous licence open source par Linagora).
- Mozilla a lancé un projet communautaire, Common Voice, visant à recueillir des échantillons de voix dans une base de données libres, pour entraîner des moteurs de reconnaissance vocale non-propriétaires...

Un serveur vocal interactif ou SVI (en anglais, interactive voice response ou IVR) est un système informatique capable de dialoguer avec un utilisateur par téléphone. Il est capable de recevoir et d'émettre des appels téléphoniques, de réagir aux actions de l'utilisateur (appui sur des touches du téléphone, reconnaissance vocale ou reconnaissance de son numéro téléphonique d'appel) selon une logique préprogrammée, de diffuser des messages préenregistrés ou en synthèse vocale, et d'accéder à des bases de données d'autre part. Un serveur vocal interactif est généralement capable de traiter de nombreux appels simultanés indépendants.

FERMA a été le premier à fournir des systèmes où la parole était créée par ***Text To Speech*** ("synthèse à partir du texte") avec la technologie de diphones du **CNET** développée à Lannion et aussi donnant la possibilité de dialogue à partir à la fois de "postes à cadran" et de postes à touche **DTMF**.

La technologie originale de traitement des signaux transitoires envoyés par les cadrans était importante compte-tenu du parc limité des postes DTMF et de l'absence de reconnaissance de parole multi-locuteur de performance suffisante, les utilisateurs en étaient si convaincus cette fonctionnalité "reconnaissance décimale" faisait partie des obligations dans les appels d'offres publics audiotels de la fin des années 1990.
De nombreuses applications vocales basées sur l'interactivité par "téléphone décimal" ont pu se développer à Taïwan et en Chine,pays où il y avait très peu de DTMF à cette époque.

L'application IA pour la reproduction et la traduction de lavoix

Le clonage de voix gratuit est une technologie basée sur l'IA qui permet de reproduire la voix d'une personne à l'aide d'algorithmes d'apprentissage automatique. L'application IA pour la reproduction de voix en ligne vous permet de créer des audios de haute qualité qui ressemblent étroitement à la voix originale.

Lancé en septembre 2024 en dehors de l'Union européenne, le nouveau mode « Avancé » de **ChatGPT Voice** permet de discuter avec un assistant vocal futuriste qui compr les émotions, peut les imiter, accepte qu'on lui coupe la parole et peut même faire des accents ou se lancer dans un jeu de rôle. La France y a accès depuis le 22 octobre 2024

Traduction instantanée et évolutive avec Language Weaver
La traduction automatique (TA), une forme précoce d'intelligence artificielle linguistique et une ressource fiable dans le processus de traduction, est disponible sur la plateforme Trados depuis des décennies. Elle fournit une traduction instantanée que vous pouvez utiliser de différentes manières : utilisez-la de manière indépendante pour la traduction automatique, intégrez-la dans vos processus pour une utilisation interactive ou choisissez d'en affiner le résultat par le biais de la post-édition.
Si la traduction automatique vous intéresse, Language Weaver est notre solution de traduction automatique évolutive, basée sur l'IA, offrant les dernières avancées en matière de traduction automatique sécurisée. Trados étant au cœur d'un riche capital technologique, nous vous offrons également la possibilité de vous connecter à des dizaines de fournisseurs de traduction automatique tiers, afin que vous puissiez personnaliser et compléter votre solution en fonction de vos besoins.

AI Phone est une application d'appel téléphonique alimentée par l'intelligence artificielle avec traduction en direct. La traduction de conversation téléphonique en direct élimine les barrières linguistiques et d'accent, vous permettant de communiquer sans effort dans différentes langues pendant vos appels.

Les voix légendaires s'expriment sur l'IA dans le doublage
CHATTANOOGA, TN – Pour les enfants et parents de ma génération, rien ne pouvait rivaliser avec les dessins animés du samedi matin, accompagnés d'un bol de céréales au chocolat.
Lorsque nous réentendons ces voix d'autrefois, cela ravive en nous des souvenirs de ces matinées lointaines. Avec l'intelligence artificielle en plein débat, la question se pose : s'agit-il vraiment de la voix de notre enfance, ou est-ce simplement le produit d'une IA ?
Au Comic Con de Chattanooga, deux voix emblématiques ont partagé leurs expériences et leurs réflexions concernant l'IA dans le monde du doublage.
Selon Scott Innes, voix de Scooby-Doo, Shaggy et d'autres personnages, « *C'est effrayant, vous savez ? Quand vous entrez dans un magasin et que vous touchez un jouet qui danse en émettant des sons que vous reconnaissez. Vous vous dites 'oui, c'est ma voix, mais je ne suis pas allé en studio pour cela.' Ensuite, vous contactez votre agent, qui vous informe qu'ils ne se sentent pas redevables car cela a été généré sans que vous fassiez le moindre travail. C'est du vol, mais il n'y a pas encore de loi solide qui empêche cela.* »

Rob Paulsen, qui incarne des personnages comme Pinky dans « Pinky et le Cerveau », a une vision légèrement différente. « *Pour moi, Warner Bros possède Yako dans Animaniacs. S'ils reproduisent des segments existants pour d'autres usages, cela leur appartient. Je sais qu'ils sont impérativement tenus de me verser une rémunération supplémentaire si cela est stipulé dans notre contrat. Cependant, l'arrivée de l'IA complique les choses. Si Warner Bros possède la voix et le personnage et souhaite créer quelque chose de nouveau avec, quelles sont les limites ? C'est un monde audacieux qui s'ouvre à nous.* »

Un exemple marquant est celui de James Earl Jones, qui a signé des droits d'utilisation de sa voix pour le personnage de Dark Vador.
Il est difficile d'imaginer un film Star Wars sans cette voix iconique, et Jones a été généreusement compensé pour cela avant sa retraite. Cela soulève des interrogations quant à ce que nous laisserons à l'IA pour la suite – après tout, qui aurait pu anticiper les événements du film Matrix sorti en 1999 ?

Merci à Marco Barnig qui a publié "Synthèse de la parole" dans lequel j'ai extrait de nombreux passages.

Notre Vision
À l'heure où la technologie, notamment l'intelligence artificielle, redéfinit des domaines créatifs comme le doublage, il est essentiel de réfléchir aux implications de cette évolution. La voix humaine, avec toutes ses nuances et son expressivité, représente un aspect fondamental de la narration. L'IA peut certainement imiter, mais peut-elle vraiment capturer l'essence même de l'émotion humaine ?
En tant que professionnels, nous devons rester vigilants quant à la manière dont ces technologies interfèrent dans nos métiers et réfléchir à des approches qui préservent notre savoir-faire tout en intégrant de manière éthique les innovations technologiques. C'est avec une telle perspective que nous pourrions naviguer dans ces nouveaux territoires sans compromettre la richesse de notre art.

Demain ?