

PARLER AVEC SON ORDINATEUR

Sébastien PEDREAU
Christian INGOUFF

Année 2011/2012
Semestre 1

INTRODUCTION

Parler avec la machine, pouvoir interagir avec elle comme un interlocuteur humain est un espoir caché, un rêve que l'homme espère pouvoir atteindre un jour. Peut être est-ce une façon pour lui de continuer son évolution, ou simplement de se débarrasser des tâches harassantes de la vie, mais l'humanité semble tendre vers le développement prochain de son alter ego mécanique.

Ce rêve a motivé nos ingénieurs et nos techniciens dans des programmes de plus en plus poussés, dans les domaines de la robotique et de l'intelligence artificielle, mais il a également inspiré les hommes dans de nombreux autres domaines, tels que la littérature ou le cinéma : de la célèbre Maria du film "Metropolis" de Fritz Lang en 1927 jusqu'à l'ordinateur central VIKI du film "I robot" de Alex Proyas, en passant par l'incontournable "Cycle de la Fondation" de Isaac Asimov, nous pouvons voir de nombreux signes de l'engouement envers la robotique.

Mais ce rêve est-il vraiment possible ? Peut-on réellement programmer des machines, de sorte que leur comportement mime celui des êtres humains ? La difficulté semble de taille, dans la mesure où une entité informatique ne peut en général pas réfléchir et apprendre d'elle même, seulement effectuer une série de tâche pour lesquelles elle est programmée. Pourtant, l'espoir reste permis. En effet, l'informatique est un domaine en plein développement, et grâce à la nanotechnologie, nos ingénieurs peuvent créer des processeurs de plus en plus réduits mais de plus en plus performants. Une des dernières trouvailles ouvre un champ de recherche quasi-illimité : la nouvelle puce SyNAPSE [1], merveille de technologie, capable d'apprendre et de retenir, à l'instar d'un cerveau humain.

Sommaire

I	ÉTAT DE L'ART	3
1	Intelligence artificielle	4
1.1	Définition générale	4
1.1.1	Intelligence artificielle forte	5
1.1.2	Intelligence artificielle faible	5
1.2	Recherche en I.A. : deux approches	7
1.2.1	I.A. connexionniste (ou neuronale)	7
1.2.2	I.A. symbolique (ou cognitive)	8
1.3	Test de Turing : l'I.A. est-elle vraiment intelligente ?	9
1.3.1	Description	9
1.3.2	Critiques et objections	9
1.3.3	Actualité	11
2	Bots	13
2.1	Définition générale	13
2.2	Avantages et inconvénients	14
2.2.1	Avantages	14
2.2.2	Inconvénients	15
2.3	Intelligence artificielle dans un bot	16
2.4	Exemple d'application : les chatterbots	17
II	LA VOIX ET L'INFORMATIQUE	19
3	Reconnaissance vocale	20
3.1	Entendre : traitement du signal	21
3.2	Traduire : décodage phonétique	22
3.3	Comprendre : analyser la phrase	23
4	Synthèse de la voix	25
4.1	Synthèse de la parole à partir d'un texte	26
4.1.1	Traitement du langage naturel	26
4.1.2	Traitement acoustique	28

Première partie

ÉTAT DE L'ART

Chapitre 1

Intelligence artificielle

1.1 Définition générale

La définition de l'intelligence artificielle (I.A.) est un point délicat, et porte encore à polémique, même du nos jours.

Prenons un exemple simple : l'ordinateur Deep Blue [2], premier ordinateur a avoir battu un champion du monde (Garry Kasparov) aux échecs en 1997, est-il doté d'une intelligence artificielle ?

D'un coté, nous avons un superordinateur doté de 256 processeurs avec une puissance de 11.38 GFLOPS, capable de calculer entre 100 et 300 millions de coups par seconde. Il a battu un champion du monde des échecs. Donc une formidable machine de calcul pouvant résoudre des problèmes complexes.

De l'autre côté, nous avons une machine bornée à une seule action, qui calcule les probabilités de réussite des coups sans réellement les comprendre. C'est une machine qui n'apprend pas, qui ne réfléchit pas, qui se contente de résoudre les problèmes grâce à des solutions rentrées à l'avance dans sa mémoire.



FIGURE 1.1 – Deep Blue

Cette machine est-elle donc une intelligence artificielle ?

Pour répondre clairement à cette question, nous poserons deux définitions : l'intelligence artificielle faible, et l'intelligence artificielle forte. [3] [4]

1.1.1 Intelligence artificielle forte

Nous qualifierons d'intelligence artificielle forte toute machine capable de montrer un comportement intelligent et rationnel. Cette machine aurait non seulement conscience des raisonnements qu'elle a effectués, mais aussi une conscience de sa propre existence en tant qu'être doté d'une intelligence, et pourquoi pas de sentiments que nous qualifierions d'humains.

On notera qu'à ce jour, les machines dotées d'une intelligence artificielle forte tiennent toujours de la science-fiction.

1.1.2 Intelligence artificielle faible

Nous définirons une intelligence artificielle faible toute machine capable de résoudre des problèmes dans une catégorie donnée (finances, mathématiques, loisirs...). Cette machine n'est toutefois pas intelligente, car elle n'est pas capable de comprendre ce qu'elle fait, comment elle le fait.

Elle se contente de résoudre le problème grâce aux données rentrées à l'avance par les personnes l'ayant créée.

NB : Ces définitions servent avant tout à qualifier un type de machine, et non spécialement l'intelligence qui l'habite.

Ainsi, pour faire quelques exemples, nous pouvons dire que Deep Blue, le superordinateur champion des échecs, est une machine à intelligence artificielle faible, car elle ne comprend pas réellement ce qu'elle fait, et ne peut exécuter son programme que grâce aux informations que les programmeurs lui ont fourni.

Ainsi, si on la place devant un jeu similaire aux échecs, mais légèrement différent (comme les Dames), Deep Blue sera tout simplement incapable de faire quoi que ce soit.

Comme les machines dotée d'une intelligence artificielle forte n'existent pas encore, nous tirerons notre prochain exemple de la science-fiction. Dans le film "2001, l'odyssée de l'espace", produit en 1968 par Stanley Kubrick, l'ordinateur de bord Hal 9000, est ce qu'on peut qualifier d'intelligence artificielle forte. Il peut faire face à tout type de situation, est capable de prendre des décisions, même si celle-ci vont à l'encontre de sa programmation initiale (ce qui signifie qu'il apprend, qu'il évolue), comme tenter de tuer les astronautes de la mission. Il a de plus conscience de sa propre existence en tant qu'entité intelligente, et tentera par tous les moyens de protéger son existence lorsque les hommes voudront le débrancher.

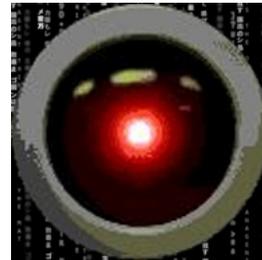


FIGURE 1.2 – Hal 9000

À présent que nous connaissons la distinction entre une intelligence artificielle faible et forte, nous pouvons nous intéresser aux deux types d'approches de l'intelligence artificielle. [3] [5]

1.2 Recherche en I.A. : deux approches

L'approche dans l'intelligence artificielle suit deux grands chemins de recherche : le connexionisme et le symbolisme. Bien que ces deux démarches aient le même but final, elles sont fondamentalement différentes.

1.2.1 I.A. connexionniste (ou neuronale)

Une branche de la recherche sur l'intelligence artificielle s'appuie sur le fait qu'il est impossible d'étudier et de reproduire l'intelligence humaine sans s'intéresser au cerveau humain, qui serait le centre de la capacité à raisonner, apprendre.

Ces recherches s'appuient donc sur les résultats de neurobiologie, neurophysiologie et des diverses sciences humaines. Leur but pour créer une intelligence artificielle en produisant des machines "neuronales", c'est-à-dire qui reproduisent le fonctionnement du cerveau humain : on cherche donc à créer un cerveau artificiel, comme l'on produit de nos jours des coeurs ou des reins artificiels mécaniques.

Cette approche a une vision minimaliste de l'intelligence, qui résiderait simplement dans la capacité d'un système à effectuer des connexions appropriées entre ces unités de base (par exemple les neurones pour un humain) pour résoudre un problème donné. Elle cherche à simuler le fonctionnement du cerveau à travers des lois mathématiques qu'on pourrait retranscrire sur ordinateur.

Les recherches de l'I.A. neuronale se situe essentiellement dans l'apprentissage. Les chercheurs pensent que les capacités viendront naturellement par la suite grâce à l'expérience et l'apprentissage progressif.

1.2.2 I.A. symbolique (ou cognitive)

L'autre démarche, est l'intelligence artificielle symbolique [6]. Elle repose sur le fait que la compréhension du monde qui nous entoure vient des représentations, des symboles qui nous entourent (d'où le nom de symbolique).

On suppose que l'homme, pour comprendre et analyser son environnement, manipule les descriptions élémentaires du monde, de l'objet simple de la vie quotidienne comme la cafetière, jusqu'aux concepts les plus abstraits, comme l'art. Afin de pouvoir transcrire ces descriptions en données facilement manipulables par un logiciel informatique, il faut créer des représentations, des modèles de connaissances, qui seront représentés par des symboles. Ces symboles sont particulièrement adaptés, car un ordinateur ne peut traiter que des signes et des symboles, et non des réalités et des faits qui semblent pourtant évidents pour l'homme.

Comme ces recherches se reposent sur les symboles et représentations, elles s'appuient sur toutes les disciplines caractéristiques à l'homme, comme la linguistique et la psychologie, qui peuvent nécessiter autre chose qu'une raisonnement purement rationnel (Les maths par exemple suivent des règles que nous qualifierons exactes et constantes).

Mais ces représentations existent-elles réellement ? Les sciences dites "humaines", comme la psychologie, ne suivent pas toujours des règles exactes ou rationnelles. D'après Descartes (philosophe et homme de science français du XVIIe siècle) dans son Discours sur la méthode, tout ce qui est pourvu de sens (comme la pensée), et régit par des règles de raisonnement, un système de notation simple, universel, qui n'attend que d'être découvert. Ce raisonnement est considéré jusqu'à ce jour comme étant l'un des plus pertinents sur la pensée humaine.

Nous connaissons maintenant les grandes définitions de l'intelligence artificielle. Mais une I.A. peut-elle être réellement qualifiée comme intelligente ? Cette question nous amène à nous intéresser aux réflexions de Turing à ce sujet.

1.3 Test de Turing : l'I.A. est-elle vraiment intelligente ?



Le test de Turing [7], du nom de son inventeur Alan Turing, est décrit pour la première fois en 1950 dans un livre scientifique écrit par Turing lui-même : “computing machinery and intelligence”. Le but de ce test serait d’identifier si une machine est une intelligence artificielle ou non, et cela au cours d’une joute verbale.

1.3.1 Description

Le test implique deux humains : une personne quelconque et un juge, ainsi que la machine dont on doit tester l’intelligence. Le juge est placé devant un ordinateur, et discute grâce à l’interface clavier avec l’autre humain, puis lors d’une seconde conversation avec la machine.

Cette discussion se déroule “en aveugle”, c’est-à-dire que le juge ne sait pas s’il discute avec l’humain ou l’ordinateur. Les conversations durent cinq minutes chacune, et à la fin du test, le juge doit être capable de pouvoir identifier l’humain de la machine. S’il se trompe, alors la machine a réussi le test.

1.3.2 Critiques et objections

Ce test peut faire l’objet de plusieurs objections que le rendraient plus ou moins caduques.

En effet, il s’avère assez simple de confondre un ordinateur en posant les bonnes questions. D’un autre côté, un ordinateur bien programmé peut très bien réussir à tromper un humain.

Ce test a-t-il réellement un sens ? Une des grandes critiques que l'on peut opposer à ce test est qu'il peut manquer de sens par certains points : pourquoi une machine devrait-elle forcément imiter l'homme dans un art aussi abstrait que celui d'une conversation ? Pourquoi devrait-on la forcer à faire des fautes d'orthographe ou de syntaxe dans le seul but d'imiter l'imperfection humaine ?

Ces questions n'ont proprement aucun sens. Une machine n'a pas à ressembler trait pour trait à un homme, et ne devrait pas avoir à apprendre nos défauts.

La pièce chinoise Nous savons qu'un ordinateur n'a pas réellement conscience de ce qu'il fait. Par exemple, Deep Blue, le superordinateur ayant battu un champion du monde en titre aux échecs, n'avait pas la moindre idée de ce qu'il faisait.

Le philosophe John Searle donne l'exemple concret de "La pièce chinoise" [8] : il s'imagine placé dans une salle sans ouverture, si ce n'est une fente assez large pour pouvoir y glisser un papier. Il n'a rien avec lui dans la pièce, excepté un livre dans sa langue, décrivant comment manipuler des symboles en chinois. Une personne extérieure à la salle glisse alors par la fente un papier écrit en chinois.



FIGURE 1.3 – John Searle

Grâce à son livre de traduction, Searle peut alors manipuler des symboles qui lui semblent incompréhensibles pour fournir une réponse adaptée, bien qu'il ne connaisse même pas la question. Il peut alors glisser la réponse par la fente, et faire croire au chinois à l'extérieur qu'il comprend parfaitement sa langue.

C'est exactement ce qu'il se déroule à l'intérieur de la machine : il y a manipulation des symboles (en l'occurrence transformés en '0' et en '1') sans pour autant saisir le savoir, le sens, la signification que transportent ces symboles.

Ainsi, un ordinateur peut très bien apparaître à un humain comme intelligent, capable de tenir une conversation, sans pour autant qu'il soit doté d'une réelle intelligence : tout dépend du "livre" qu'on lui a donné au départ pour qu'il puisse répondre correctement.

Rationalité et émotion Un humain est régi non seulement par une certaine logique qui lui est propre, mais aussi par ses sentiments, ses émotions, qui peuvent le conduire dans des actes sortant de toute rationalité.

Un ordinateur, lui, restera quoiqu'il arrive dans une logique sans faille faute de pouvoir exprimer des émotions. Ainsi, il aura des difficultés à saisir certains concepts qui requiert une analyse plus abstraite, tel que l'art, la littérature, la religion, des sujets reposant avant tout sur une certaine imagination, ou des préceptes moraux propres à l'homme (compassion, colère, indulgence...).

1.3.3 Actualité

Extrait du "Journaldelascience.fr" : [9]

Pour la première fois dans l'histoire de l'intelligence artificielle, un logiciel appelé Cleverbot vient de réussir le Test de Turing, le célèbre test d'intelligence artificielle imaginé en 1950 par le mathématicien Alan Turing.

Le 4 septembre 2011, le logiciel de conversation Cleverbot a donc réussi le test de Turing, lors d'un test impliquant près de 1334 personnes. quinze personnes ont discuté pendant cinq minutes avec quinze autres personnes via des systèmes de messagerie instantané, tandis que quinze autre personnes discutaient avec le logiciel Cleverbot. Toutes les discussions ont été retransmises au jury constitué des 1334 personnes, qui on alors voté pour chaque conversation si les participants étaient humains ou ordinateur.

A près de 60% (59.3% exactement), le logiciel Cleverbot a été jugé comme un humain, ce qui est en soi un exploit exceptionnel, car il est le premier à réussir le test de Turing. Bien sur, il faut remarquer que le test de Turing souffre de certaine faiblesse. Cela est très bien illustré par le fait que les humains n'ont été jugé "humains" qu'à 63.3%. Ce décalage prouve qu'il y a encore de nombreux progrès à faire, mais que l'espoir reste permis, car cette avancée marque un grand pas dans les logiciels dit "chatterbots".

Bien que l'algorithme qui régit le comportement de Cleverbot soit complexe, il suit l'idée principale de scruter l'historique de la conversation pour savoir quelle réponse est la plus adaptée dans le contexte. Bien sur, ce principe de réponse le rend quelque fois peu cohérent. Nous en déduisons ainsi qu'il n'est pas réellement intelligent. Il ne réfléchit pas réellement, et ne comprend sans doute pas ce qu'il fait. Il ne fait que mimer la façon de parler des humains. Il le fait bien, certes, mais cela reste du mimétisme.

Nous avons à présent quelques informations élémentaires sur les intelligences artificielles. Bien que concises, elles nous permettent de nous faire une idée claire concernant le sujet. Cependant, pour mener à terme notre projet, il nous faudra obligatoirement nous intéresser à la notion de bot informatique.

Chapitre 2

Bots

2.1 Définition générale

Le mot "bot" est la contraction du mot "robot". On peut donc définir un bot informatique comme étant un "robot virtuel" : c'est un agent logiciel qui effectue automatiquement, ou semi-automatiquement (avec l'intervention d'un humain pour déclencher l'action), des tâches prédéfinies lors de son codage. Ces tâches sont très variées et peuvent aller du simple "Bonjour !" automatique au triage de données dans un moteur de recherche comme Google. [10] [11]

Comme un robot, le bot a été conçu afin d'assister l'humain dans ses tâches ou de faire ces tâches à sa place. Le plus souvent, une action effectuée par un bot pourrait aussi être effectuée par un homme. Par conséquent, le bot, semblablement au robot, a une utilité pratique.

Le bot se différencie du robot "réel" par son domaine d'action : alors que le robot est un dispositif mécatronique (alliant mécanique, électronique et informatique), le bot agit uniquement dans le domaine informatique, par l'intermédiaire d'une interface, comme un ordinateur, un téléphone portable, un répondeur vocal ou encore une calculatrice.

Son mode de fonctionnement peut varier : certains bots effectueront leurs actions dès leur exécution et ne s'arrêteront seulement si on ne les arrête, d'autres attendront des ordres, des commandes de la part de l'utilisateur (le plus souvent tapées au clavier, ou prononcées à voix haute) avant d'effectuer la ou les actions correspondant à l'ordre donné.



FIGURE 2.1 – Vue d'artiste du Googlebot

2.2 Avantages et inconvénients

2.2.1 Avantages

Dans le cadre de l'assistance à l'homme, les bots informatiques comportent de nombreux avantages.

Efficacité "Errare humanum est" : l'homme n'est pas à l'abri d'éventuelles erreurs dans ses actions. Or, le bot est capable d'effectuer une tâche systématiquement. De cette manière, si le bot est bien conçu, toute erreur humaine possible est prévenue : c'est ce qui fait l'efficacité du bot informatique par rapport à l'homme.

Rapidité d'action Le bot informatique sait effectuer des actions normalement très pénibles pour l'homme efficacement et avec une rapidité étonnante. Par exemple, un bot vérifiera l'orthographe d'un texte beaucoup plus vite qu'un utilisateur humain, et n'oubliera aucun mot.

Tâches répétitives De la même manière dont les robots ont remplacé l'homme dans le travail à la chaîne, les bots informatiques sont très efficaces dans le cadre d'actions répétitives, systématiques, comme la recherche et l'accumulation de données pour un moteur de recherche.

2.2.2 Inconvénients

Malgré tous les avantages que les bots peuvent procurer, des inconvénients subsistent à leur usage, par rapport à l'homme.

Instabilité Un bot mal conçu peut devenir incontrôlable et devenir dangereux : si on utilise un ordinateur dans le cadre de l'exécution du bot, cet ordinateur court le risque d'être endommagé par les tâches effectuées par le bot.

Absence de jugement Certaines tâches requièrent un certain jugement afin d'être effectuées correctement : dans le cadre économique, le bot ne pourra pas prévoir un krach boursier dû aux actualités alors qu'un utilisateur humain le pourra. De même, un bot pourrait faire exploser une ville sans aucun remord.

Malveillance Un bot est à la merci de son utilisateur, s'il connaît le fonctionnement du bot. Ainsi, n'importe qui pourra utiliser des bots à son avantage, ce qui peut laisser libre court à des actions malveillantes, comme le spam (envoi automatique et massif de messages) ou le hacking, facilitées grâce aux bots.

2.3 Intelligence artificielle dans un bot

Le bot étant un agent logiciel, toutes ses actions vont uniquement dépendre de son utilisateur. Le bot n'est donc pas autonome : pour remédier à cela, on peut utiliser une intelligence artificielle afin de le "piloter".

Dans l'idéal, pour concevoir un bot autonome avec un comportement proche de l'humain, on préférera l'utilisation d'une I.A. forte, idée non concrétisée pour l'instant. Il n'y a donc que des bots à I.A. faible, qui ne sont donc utilisés à ce jour là que pour des actions simples, ne requérant pas de jugement ou d'émotion en particulier.

Certains bots cherchent néanmoins à imiter le comportement humain : les programmeurs y parviennent le plus souvent en créant l'illusion d'une intelligence. Pour cela, on peut coder le bot selon le comportement de l'homme, en attribuant à chaque donnée reçue par le bot une réaction "humaine".

Cependant, comme les réactions du bot sont systématiques, très rapides et sans erreurs, on peut facilement différencier un bot d'un humain, avec le test de Turing par exemple.

Bien entendu, il existe un nombre incroyable de moyens de remédier à ce problème en améliorant davantage l'illusion donnée par le bot. Cependant, l'être humain est doté de tellement de particularités qu'il est absolument impensable de vouloir créer un bot se comportant exactement comme un humain.

2.4 Exemple d'application : les chatterbots

Les "chatterbots", agents conversationnels, sont des bots cherchant à imiter l'être humain dans l'art de la conversation, de manière à permettre en quelque sorte un dialogue entre homme et machine. [12] [13]

La recherche sur ce genre de bots a été fortement influencée par le test de Turing, instaurant une véritable compétition entre les développeurs de chatterbots. Les premiers apparus sont ELIZA (1966) et PARRY (1972), respectivement 16 et 22 ans après l'introduction du test de Turing. Plus récemment, on a vu le succès de Cleverbot face au test de Turing. [14]

Leur fonctionnement peut être plus ou moins complexe : le code d'ELIZA tient par exemple en 3 pages en langage SNOBOL. En effet, ELIZA, afin de répondre à une question, peut parfois se contenter de répéter la phrase qui lui est donnée. Le reste du temps, ELIZA identifie des mots-clés, à la manière des robots d'indexation, pour constituer une réponse adéquate à une question. [12] [13]



FIGURE 2.2 – Interface de Cleverbot [14]

Les bots les plus récents tentent surtout de leurrer l'utilisateur humain, dans le cadre du test de Turing : un chatterbot peut alors se montrer très imprévisible, faire des fautes d'orthographe, "s'absenter virtuellement", avoir une réponse adéquate à des questions non comprises ("Ne changez pas de sujet", etc...) ou encore tenter de séduire l'utilisateur en se faisant passer pour le sexe opposé à celui-ci.

Malgré l'impossibilité pour le chatterbot d'exprimer de vrais sentiments ou d'émettre un vrai jugement, il peut malgré tout faire croire à l'utilisateur humain qu'il en est capable grâce à des réponses prédéfinies par les développeurs. C'est ainsi que l'effet ELIZA se crée : inconsciemment, grâce à ces réponses prédéfinies, l'utilisateur humain assimile le comportement du bot à celui d'un humain.

Douglas Hofstadter donne un exemple de la vie de tous les jours qui illustre cet effet : un distributeur automatique qui dit "Merci" à la fin d'une interaction. On pourrait croire à de la gratitude de la part du distributeur, mais ce "Merci" n'est issu que d'une réponse prédéfinie, une série de caractères, issue du concepteur. C'est ce qu'on appelle un sophisme : les hommes ont tendance à associer des traits de caractère purement humains à des entités qui ne le sont pas, comme les machines.

Les chatterbots peuvent aussi être utilisés à des fins malhonnêtes comme l'extorsion de mots de passe, de coordonnées bancaires, ou encore se faire passer pour un humain dans un site de rencontres.

Les bots informatiques sont donc une formidable avancée technologique pour l'être humain, permettant d'automatiser de nombreuses actions et de nous faciliter la vie. Dans le cadre de notre projet, ils seront très importants, et nous aurons lors de notre seconde année à programmer notre propre bot informatique. Cependant, le bot dont nous rêvons devra être capable de communiquer avec l'humain, et devra donc être doté d'un système de reconnaissance vocale, et d'un logiciel de synthèse de la voix pour formuler ses réponses.

Deuxième partie

LA VOIX ET L'INFORMATIQUE

Chapitre 3

Reconnaissance vocale

Si la machine veut un jour pouvoir ressembler à l'homme, il faudra qu'elle dispose de la capacité de parler, lire et écrire. Si la manipulation de l'écriture demande une somme conséquente de données, le langage nécessite encore plus de calculs et de complexité. De plus, l'écriture et la parole resteront toujours difficile à automatiser, car elles demandent une part de créativité importante, aligner des mots ne suffisant pas toujours pour créer un chef d'oeuvre.

Les premières avancées dans la reconnaissance vocale datent des années 1950, avec un système développé par Bell pouvant reconnaître les 10 chiffres. Il faudra attendre 1972 pour avoir un système pouvant reconnaître des mots, et les années 1990 pour véritablement pouvoir dialoguer avec la machine. Aujourd'hui, nous sommes capables de commander un téléphone portable à la voix, avec par exemple le logiciel Siri, développé par Apple, ou encore de rédiger un texte à l'ordinateur entièrement à la voix, grâce à Dragon NaturallySpeaking, par Nuance.



FIGURE 3.1 – Interface de Siri

3.1 Entendre : traitement du signal

La machine nous entend, capte notre voix grâce à un microphone, mais elle doit encore traiter cette voix pour en tirer des mots. Tout d'abord, la voix est amplifiée, filtrée et échantillonnée afin de faciliter son étude. Elle est ensuite analysée : pour cela, on peut étudier l'évolution de la fréquence du signal par rapport au temps, représentée à l'aide d'un audiogramme, avec différentes méthodes. [15] [16]

Transformée de Fourier Avec le carré du module de la transformée de Fourier, on peut obtenir un spectrogramme nous donnant l'énergie associée à une fréquence à un temps donné.

Prédiction linéaire On approche le signal enregistré par un modèle linéaire, plus facile à traiter. Il y a toutefois une marge d'erreur, qu'on tente de minimiser.

Evaluation des coefficients "cepstraux" Le "cepstre" est défini comme la transformée de Fourier inverse du module d'un spectre exprimé en échelle logarithmique. L'intérêt est de séparer la source du conduit vocal grâce aux propriétés du logarithme.



FIGURE 3.2 – Audiogramme : fréquence du signal par rapport au temps

3.2 Traduire : décodage phonétique

Pour obtenir des mots à partir du signal traité, la machine doit désormais décoder ce signal : elle le fait tout d'abord en découpant le signal en syllabes, demi-syllabes ou phonèmes en s'appuyant sur son évolution par rapport au temps, puis elle identifie chacun de ces segments en les comparant à des spectres de référence. [15] [16]

Une fois tous les segments identifiés, la machine peut désormais reconstruire la phrase qui lui a été adressée, à partir du décodage effectué.

La principale difficulté dans le décodage phonétique réside dans la diversité et la richesse linguistique : un mot ne sera pas forcément prononcé de la même manière par tout le monde.

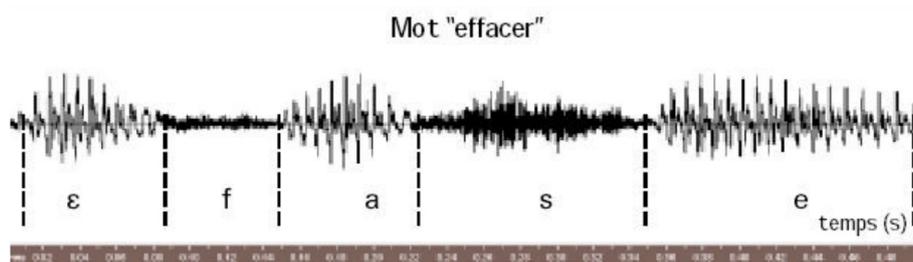


FIGURE 3.3 – Audiogramme du mot "effacer" décomposé

3.3 Comprendre : analyser la phrase

Pour être complète, une analyse doit passer par quatre étapes successives : [5]

Analyse morphologique

Cette étape consiste à identifier tous les mots présents dans la phrase à analyser, et à comprendre le forme de chacun de ses mots : le genre, le nombre, le type de chaque nom et adjectif, le temps des verbes, etc.

Pour que cette analyse soit complète, il faut que la machine possède la connaissance de l'ensemble de la langue concernée, avec toutes ses règles, mais aussi (et surtout) toutes ses exceptions.

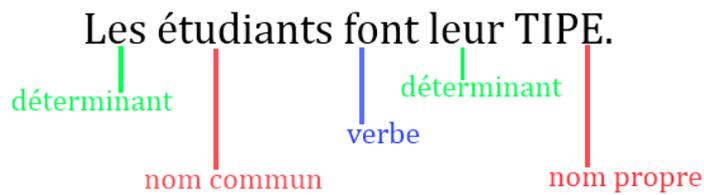


FIGURE 3.4 – Analyse morphologique

Analyse syntaxique

Cette seconde étape repose sur le fait d'ordonner les mots, identifiés dans la première étape, afin de créer une phrase. Il s'agit de mettre un ordre clair, compréhensible et grammaticalement correct pour la machine.

On crée donc des groupes nominaux, des groupes verbaux aux ensembles de mots, puis on associe à chaque mot le fait qu'il soit un article, un nom, un verbe ou autre.



FIGURE 3.5 – Analyse syntaxique

Analyse sémantique

Une analyse morphologique ne suffit pas toujours pour saisir le sens d'un mot. En effet, un mot peut avoir plusieurs sens : par exemple, un plateau peut très bien désigner un objet permettant de porter d'autre chose (comme un plateau de thé), mais ce mot peut tout aussi bien désigner un plateau montagneux, ou encore un plateau télévisé.

Analyse pragmatique

Toutes les analyses précédentes ne suffisent pas forcément à comprendre le sens d'une phrase, surtout si celle-ci a un sens abstrait. Il faut connaître un peu son interlocuteur, son niveau de langage, le sujet de la conversation. Une phrase comme "Le Grand Corbeau a puni tous les habitants." ne peut se comprendre que si l'on se place dans un contexte bien précis, en l'occurrence dans le contexte des jeux vidéos.

Donc, pour analyser une phrase, un ordinateur a besoin d'une quantité phénoménale de données. De plus, elle a besoin de d'informations supplémentaires, comme la langue de la conversation (L'analyse d'une phrase en français différera de l'analyse d'une phrase en allemand ou en anglais.).

Chapitre 4

Synthèse de la voix

Si le logiciel que nous cherchons à concevoir doit répondre à l'utilisateur, nous devons alors utiliser des méthodes de synthèse informatique de la voix [17]. Le procédé de la synthèse de la voix [18] existe depuis très longtemps à présent, car les premiers prototypes fonctionnels ont été créés dans les années 1960 (appareils de synthèse vocale par règles) et devenant de plus en plus évolués au fil du temps [19]. Cependant, l'idée même de créer des machines capables de reproduire la voix de l'homme remonte à la Renaissance, et nous avons retrouvé des croquis de telles machines écrits de l'inventeur Von Kempelen datés de 1791.

Nous aborderons dans cette partie les différentes étapes nécessaires à la synthèse de la voix sur ordinateur. [20]

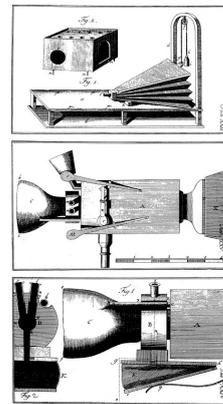


FIGURE 4.1 – Schémas de Von Kempelen

4.1 Synthèse de la parole à partir d'un texte

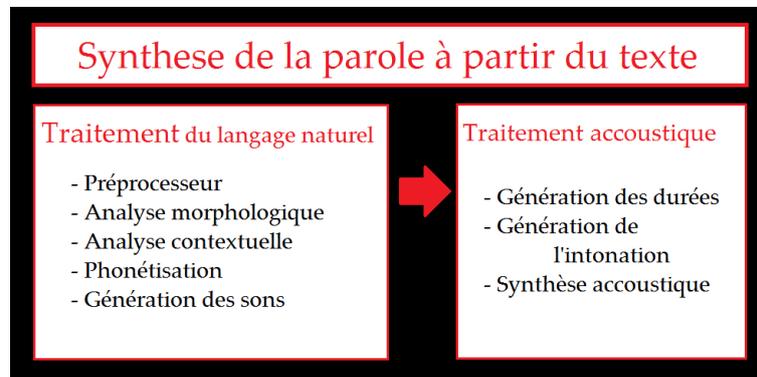


FIGURE 4.2 – Synthèse de la parole

4.1.1 Traitement du langage naturel

pré-processeur

Lors de cette partie, le logiciel va repérer toute partie du texte qui pourrait poser problème lors du traitement, comment les abréviations, les dates, les sommes d'argents ou encore les acronymes.

Exemple : 24/12/2011 se prononcera 24 décembre 2011

Cette partie identifiera également la ponctuation de la phrase.

Analyse morphologique et contextuelle

Le logiciel va ensuite faire une analyse morphologique de la phrase. A chaque mot, il va donner tous les types de groupes possibles auxquels le mot peut appartenir d'après la façon dont il est écrit.

Exemple : président

=> Nom (singulier)

=> Verbe (troisième personne du pluriel)

Ensuite vient l'analyse contextuelle, qui permet d'analyser la phrase grammaticalement et d'attacher à chaque mot une "étiquette" avec sa nature grammaticale, dans le but final de prononcer la phrase correctement.

Exemple : "Les présidents président" : les mot "présidents" et "président" ne se prononceront pas de la même manière, il faut savoir que le premier est un nom et le second est un verbe.

Phonétisation

D'après les informations recueillies par le logiciel dans les parties précédentes, le texte va être transcrit dans l'Alphabet Phonétique International [21]. C'est un alphabet constitué de 118 caractères différents, qui peuvent être agrémentés de diacritiques (des accents en général) permettant de marquer de subtiles nuances, et ainsi de couvrir les sons de toutes les langues. Cependant, cet alphabet est difficile à utiliser sur ordinateur, surtout s'il doit être tapé de la part d'un humain. Dans un soucis pratique, la majorité des entreprises utilisent l'alphabet SAMPA, qui n'est ni plus ni moins que l'Alphabet Phonétique International écrit avec des touches communes du clavier d'un ordinateur.

Voyelles :

/ a / patte (antérieur)

/ ɑ / pâte (postérieur)

/ e / pré (antérieur, fermé)

/ ɛ / sel (antérieur, ouvert)

/ i / lit (antérieur, fermé)

/ o / pot (postérieur, fermé, arrondi)

/ ɔ / port (postérieur, ouvert, arrondi)

/ ø / peu (antérieur, fermé, arrondi)

FIGURE 4.3 – Alphabet phonétique

Génération des tons

Avec les informations recueillies dans les parties précédentes (type grammatical, ponctuation), la logiciel va attribuer à chaque mot un "ton", qui déterminera plus tard la prononciation.

4.1.2 Traitement acoustique

Générations des durées et de la courbe intonative

Cette étape, aussi appelée synthèse de la prosodie permettra de définir un rythme et une intonation à la phrase, grâce aux "tons" et aux autres données des étapes précédentes, afin que la prononciation soit la plus naturelle et la plus fluide possible.

Synthèse acoustique

Finalement, à chaque caractère de l'alphabet phonétique international [21] (ou de l'alphabet SAMPA si celui-ci est utilisé), le logiciel va associer un court extrait audio correspondant au son recherché. Toutes les informations des étapes précédentes permettent de choisir avec exactitude quelle séquence est la plus adaptée pour une prononciation la plus naturelle possible. Les différents morceaux audio sont ensuite concaténer les uns aux autres, en prenant compte de la ponctuation.

Il existe des bibliothèques de morceaux sonores pour chaque langue. Ces bibliothèques vont de 40000 à 100000 morceaux.

CONCLUSION

Ce second semestre de travail nous a vu mener des recherches plus approfondies sur des parties que nous n'avions pu qu'effleurer auparavant, à savoir la reconnaissance vocale et la synthèse de la voix. Ce sont des sujets très intéressants à étudier, malgré le manque d'informations précises sur les sites et les livres réservés au grand public.

Lors de nos recherches, nous avons eu la chance de pouvoir discuter avec Fabrice Malfrère, chef technique dans la section Recherche et Développement de l'entreprise Acapela, spécialisée dans la synthèse de la voix. Une telle interview nous a beaucoup appris, et nous avons ainsi pu poser toutes les questions que nous souhaitions. Les informations que nous avons récupérées sont plus précises que ce que nous aurions pu glaner sur des livres ou des sites internet non spécialisés, nous permettant ainsi de pousser plus loin notre réflexion.

Nous pouvons à présent porter notre regard vers la seconde année de notre projet, à savoir la phase de réalisation. Forts des informations que nous avons récoltées tout au cours de l'année, nous sommes impatients de concrétiser notre idée initiale, et peut être que nous pourrons bientôt discuter avec nos ordinateurs.



Bibliographie

- [1] *Silicon.fr*.
www.silicon.fr/synapse-la-puce-dibm-qui-se-rapproche-du-cerveau-humain-58524.html.
Description de la nouvelle puce SyNAPSE, grand pas dans le monde de l'I.A.
- [2] *Wikipédia*.
fr.wikipedia.org/wiki/deep_blue.
Article sur Deep Blue.
- [3] *Wikipédia*.
fr.wikipedia.org/wiki/intelligence_artificielle.
Article sur l'intelligence artificielle.
- [4] *Elektronique.fr*.
www.elektronique.fr/news-9-intelligence-artificielle.html.
Article sur l'intelligence artificielle.
- [5] *Joanna Pomian*. L'Intelligence Artificielle. mai 1993.
- [6] *Université de Genève*.
tecfa.unige.ch/tecfa/publicat/schneider/these-daniel/wmwork/www/phd_93.html.
Extrait d'une thèse sur l'I.A.
- [7] *Wikipédia*.
fr.wikipedia.org/wiki/test_de_turing.
Article sur le test de Turing.
- [8] *Learn Intelligence*.
www.learn-intelligence.com/fr/ia/lintelligence-artificielle.
Test de Turing : pièce chinoise.
- [9] *Journal de la Science*.
www.journaldelascience.fr/technologie/articles/un-logiciel-dintelligence-artificielle-aurait-reussi-le-test-de-turing-2223.
Article sur l'exploit de Cleverbot.
- [10] *Wikipédia*.
fr.wikipedia.org/wiki/bot_informatique.
Article sur le bot informatique.

- [11] *Wikipédia*.
en.wikipedia.org/wiki/internet_bot.
 Article (anglais) sur le bot informatique.
- [12] *Wikipédia*.
fr.wikipedia.org/wiki/agent_conversationnel.
 Article sur le chatterbot.
- [13] *Wikipédia*.
en.wikipedia.org/wiki/chatterbot.
 Article (anglais) sur le chatterbot.
- [14] *CleverBot*.
www.cleverbot.com.
 Le chatterbot CleverBot.
- [15] *Université de Nice*.
deptinfo.unice.fr/twiki/pub/linfo/planningdessoutenances20032004/benguiguismais-hamdan.pdf.
 Reconnaissance vocale.
- [16] *Conservatoire National des Arts et Métiers*.
r.batault.free.fr/probatoire/probatoire.html.
 Reconnaissance vocale.
- [17] *gralon.net*.
www.gralon.net/articles/materiel-et-consommables/materiel-informatique-et-consommable-informatique/article-la-synthese-vocale-presentation-et-fonctionnement-1910.htm.
 Principe de fonctionnement simplifié de la synthèse vocale.
- [18] *Wikipédia*.
fr.wikipedia.org/wiki/synthese_vocale.
 Article sur la synthèse vocale.
- [19] *tcts*.
tcts.fpms.ac.be/synthesis/.
 Projet d'un nouveau logiciel de synthèse vocale multilinguiste.
- [20] *Article sur la synthèse de la voix*.
tcts.fpms.ac.be/publications/regpapers/2001/tal2001_all.pdf.
 Article écrit par Fabrice Malfrère, ingénieur que nous avons interviewé.
- [21] *Wikipédia*.
fr.wikipedia.org/wiki/alphabet_phonetique_international.
 Article sur l'alphabet phonétique internationale.