

Actions de sensibilisation scientifique

"Utilisons nos facultés!"

THEME N°2

**Ces merveilleux fous parlant avec leurs drôles de machines
Introduction à la synthèse et à la reconnaissance vocales**

Handicapé moteur suite à une maladie dégénérative de son système nerveux, l'astrophysicien Stephen Hawking donne ses conférences grâce à un ordinateur qui parle à sa place. Hawking tape son texte sur le clavier et la machine le transforme en sons. Elle se trompe très rarement, même quand elle est confrontée à un mot dont la signification change en fonction de la prononciation, comme dans "les poules du couvent couvent"! ¹

Evidemment, c'est Hawking qui pense, pas la machine. Une machine ne peut pas penser. Elle ne produit ni idées, ni concepts. Par conséquent, elle ne peut pas **générer** de parole, car la parole est plus qu'une juxtaposition de mots: c'est **un flux continu de mots qui traduisent une idée**. Par contre, si on lui fournit du texte, la machine peut en déduire des sons possibles, choisir la succession des sons qui ont la plus grande probabilité d'avoir un sens, et produire ces sons en imitant la voix humaine.²

Les machines qui remplissent ces fonctions sont appelées synthétiseurs vocaux, ou synthétiseurs de parole. Attention: il ne faut pas les confondre avec les systèmes d'annonces utilisés dans les gares ou dans certaines grandes surfaces. Quand le haut-parleur de la Gare du Midi vous dit, sur un ton impersonnel, que "*le train IC à destination de Mons, départ prévu 19H03, entre en gare voie 20*", la machine ne fait qu'assembler des éléments pré-enregistrés qu'elle a en mémoire. Son vocabulaire est limité (quelques centaines de mots) et le locuteur doit être toujours le même. Ce n'est pas de la **synthèse** vocale mais de la mise bout à bout ("concaténation") de mots.

¹ De tels mots sont appelés "homographes hétérophones". L'anglais en comporte énormément.

² C'est pourquoi, lorsque le handicap de la parole est le résultat d'un désordre moteur ou d'un désordre de la perception sensorielle, la machine peut venir en aide à la personne handicapée et lui permettre de communiquer avec son entourage. L'exemple de Hawking montre qu'elle peut même lui permettre de donner des conférences sur des sujets compliqués. On peut lire l'émouvant témoignage de Hawking (en anglais) à l'URL suivante: <http://www.hawking.org.uk/text/disable.html>

On travaille beaucoup sur la synthèse vocale dans nos régions – et sur le domaine connexe, mais néanmoins distinct, de la reconnaissance vocale³. Des cours sont donnés et des travaux de recherche sont menés, tant à Mons qu'à l'Université de Valenciennes. **Multitel**, le laboratoire mis sur pied par la Faculté Polytechnique de Mons, est à la pointe de la recherche internationale en ces matières. En 1997, des ingénieurs de Multitel ont fondé une "spin off": **Babel Technologies**. Cinq ans plus tard, l'entreprise réalisait 3,5 millions d'Euros de chiffre d'affaire. Elle emploie aujourd'hui une quarantaine de personnes, dont une vingtaine dans notre région, et occupe une position de choix sur ce marché relativement nouveau.⁴

MYSTERES DE LA PAROLE

Complicquée, la synthèse vocale? A première vue, non. La synthèse musicale existe depuis belle lurette. Pourquoi serait-il plus difficile de synthétiser la parole que la musique? Les sons produits par la voix humaine sont tous bien identifiés et on connaît leurs caractéristiques physiques: fréquence et longueur d'onde. Les dictionnaires nous donnent une indication de la prononciation des mots, par le biais des symboles phonétiques. Il semble donc qu'il suffise de mettre un bon dictionnaire dans la mémoire de l'ordinateur, et de le relier à un synthétiseur, pour résoudre le problème. Pourtant, les choses sont infiniment plus complexes...

Premièrement, les dictionnaires nous donnent seulement la prononciation de la **racine** des mots. Les variations résultant des pluriels (un cheval... des chevaux), des féminins (beau... belle), des conjugaisons (je mange, nous mangeons, nous mangeâmes,... que nous mangeassions!) ne sont généralement pas explicitées.⁵

Deuxièmement, il y a ce problème des **homographes hétérophones**: non seulement "les poules du couvent couvent", mais "tu les hélas, hélas!", "nous portions des portions", "le président et les autres membres du présidium président", "tu as un as dans ton jeu"...

Du fait que nous avons acquis une connaissance profonde de notre langue maternelle, nous trouvons spontanément la prononciation correcte de ces propositions. Nous pouvons même, à première lecture, prononcer correctement un texte dont nous ne comprenons pas le sens – ou qui **n'a pas** de sens. C'est ainsi que nous pouvons lire, avec l'intonation "correcte", un texte comme le *Jabberwocky*, de Lewis Carrol:

³ Synthèse vocale: la machine transforme le texte en parole (Text-to-Speech, ou TTS); reconnaissance vocale: elle transforme les sons en texte. Synthèse et reconnaissance utilisent des outils mathématiques et statistiques analogues, mais d'une manière différente, d'où la séparation de la recherche sur les deux domaines. Pour ne pas alourdir cette présentation, on se concentrera ici sur la synthèse vocale.

⁴ Babel Technologies est spécialisée sur quatre "niches": multimédia, télécoms, industrie et applications mobiles. Un de ses points forts est de travailler à la fois en synthèse et en reconnaissance vocales. Certains produits de Babeltech peuvent être testés sur l'URL de l'entreprise: www.babeltech.com

⁵ Difficulté plus marquée dans le cas des langues fortement infléchies, comme le français ou le néerlandais, que dans le cas de l'anglais, où les adjectifs ne varient pas en fonction du genre, par exemple.

*Il était grilheure: les slictueux toves
Gyraient sur l'alloinde et vriblaient
Tout frivoleux allaient les borogoves
Les verchons fourgus bourniflaient.⁶*

Une instance de notre cerveau fait spontanément la différence entre les verbes et les non-verbes, les noms, les articles et les adjectifs, etc. en fonction de règles syntaxiques que nous maîtrisons... même si nous sommes incapables de les formuler. Mais la machine, elle, comment pourrait-elle s'y retrouver?

Troisièmement, les symboles phonétiques du dictionnaire ne nous aident que si nous savons... **comment ils sonnent**. En soi, un mot formé de symboles phonétiques, comme ʒθhθl , ne signifie rien pour nous. Ce problème est bien connu des lycéens qui étudient l'anglais et qui tentent désespérément, sur base des symboles, de saisir la différence entre "though", "through", "cough" ou "tough"!... Bonjour les postillons!

Quatrièmement, et surtout, les mots contenus dans une phrase **ne sont pas prononcés comme s'ils étaient isolés**. Il s'altèrent à leurs frontières avec les mots environnants. Il y a des liaisons ("les-z-yeux") ou des élisions ("com' tu veux"). Les spécialistes nous apprennent qu'il y a même des altérations de prononciation en fonction de l'organisation de la phrase en groupes de mots...

Enfin, tous les mots ne se trouvent pas dans les dictionnaires phonétiques. De **nouveaux mots** apparaissent tous les jours. Leur prononciation doit être déduite de celle de noms connus. Sans compter le problème des noms propres, et celui des mots empruntés à d'autres langues (mais prononcés autrement que dans la langue d'origine: le "steak" qui se dit "stèck" en français se prononce "stééck" en anglais, par exemple).

GYMNASTIQUE ARTICULATOIRE

Reprenons la comparaison avec la synthèse musicale. A toute note du piano correspond un son précis. Il n'en va pas de même avec les lettres de l'alphabet, une lettre pouvant produire plusieurs sons: le "t", par exemple, a deux prononciations différentes dans le mot "attention" (sans compter qu'il peut aussi être muet). Même chose avec le "s", le "c", etc.⁷ C'est pourquoi la synthèse vocale ne travaille pas à partir des lettres, mais à partir des **phonèmes**, représentés par les symboles phonétiques.

Cependant, les phonèmes ne suffisent pas à nous tirer d'embarras. En effet, alors qu'un "do" sera toujours un "do" et gardera exactement la même fréquence, qu'il vienne

⁶ Traduit par H. Parisot, cité dans "Synthèse Vocale et Reconnaissance de la Parole: Droites Gauches et Monde Parallèles", par T. Dutoit, L. Couvreur, F. Malfrère, V. Pagel, C. Ris in "Actes du 6^e congrès français d'acoustique".

⁷ On connaît une exception: le turc. Dans cette langue, à chaque graphème correspond un et un seul phonème.

après un "ré" ou après un "si", il n'en va pas de même avec nos phonèmes: leur réalisation acoustique **change en fonction des autres phonèmes qui les entourent!** C'est ainsi que le son "a" (symb. phon.?) n'est pas exactement le même dans "bar", dans "cap", dans "mat" et dans "sas".

Cause de ces nuances? Chaque son est le résultat d'une position bien précise des lèvres, de la langue, du voile du palais, de la glotte... Or, les muscles articulatoires qui positionnent ces organes travaillent, pour ainsi dire, à l'économie: ils évitent les gymnastiques trop périlleuses dans le passage d'un phonème à un autre. Ils préfèrent enchaîner les mouvements doucement, en un mécanisme appelé **coarticulation**.⁸

C'est pourquoi, en synthèse vocale, on ne travaille pas seulement à partir des phonèmes, mais aussi à partir de ce qu'on appelle les **diphones**, qui sont, pour ainsi dire, les "jointures" entre phonèmes. Dans l'exemple cité, "ba", "ca", "ma" et "sa" sont quatre diphones différents, partant chacun du milieu du premier phonème et se terminant au milieu du phonème "a".

MELODIE DU DISCOURS

Enfin, de même que la musique connaît des "silences", "soupirs", "demi-soupirs", etc., et toute une série de nuances (depuis le *pianissimo* jusqu'au *fortissimo*), de même **la parole ne se débite pas sur un ton monocorde et de façon ininterrompue**. Il y a des pauses et des accélérations dans le discours. Le ton (la "hauteur" de la voix), l'accentuation (le volume de la voix) et la longueur des syllabes peuvent varier, produisant une certaine mélodie de la parole. C'est ce qu'on appelle l'intonation, ou **prosodie**.

Or - et c'est là que ça se corse - ces "variations prosodiques" **codéterminent le sens du message**⁹. Nous ne disons pas exactement la même chose lorsque nous disons "*Je le lui ai dit hier*" et lorsque nous disons "*Je le lui ai dit hier*". La prosodie met des mots ou des parties de mots en évidence (information sémantique) et établit des relations entre différents éléments de la phrase (information syntaxique). De plus, la dernière syllabe délivre une information spécifique (fin ou continuation du message).

Chacun peut tester la richesse des nuances prosodiques en s'exerçant à prononcer de différentes manières une phrase simple, telle que: "*je le lui ai dit hier*". Le tableau ci-

⁸ Les chanteurs connaissent bien cette difficulté, qui apprennent à "placer" chaque voyelle et font des vocalises d'enchaînement des voyelles dans un certain ordre, pour éviter des heurts dans le flux de la voix. La coarticulation renvoie en fait à la nature de la parole en tant que processus continu: le langage n'est pas plus une simple juxtaposition de phonèmes qu'il n'est une simple juxtaposition de mots isolés.

⁹ Dans certaines langues, dites tonales, la tonalité fait plus que codéterminer le sens: elle **est** sens en soi. En chinois, par exemple, certains mots ont des significations différentes en fonction du ton sur lequel ils sont prononcés.

dessous reprend une petite partie des possibilités; les mots en **grasses** sont accentués, les mots en ^{exposant} sont dits un ton plus haut:

Information sémantique	Information syntaxique	Info de la dernière syllabe
Je le lui dit hier	Je le lui ai dit hier	Je le lui ai dit hier
Je le lui ai dit hier	Je le lui ai dit hier	Je le lui ai dit hier
Je le lui ai dit hier	Je le lui ai dit hier	

Conclusion: la prosodie est, au même titre que l'enchaînement des mots, **une expression des idées** du locuteur. Or, on le sait, la machine ne comprend pas les idées - elle ne fait que traiter des signaux. Pour qu'elle produise une intonation totalement conforme à l'intention ou à l'émotion de la personne qui parle par son intermédiaire, il faudrait que la machine reçoive et déchiffre, en plus du texte et en parallèle à celui-ci, des signaux prosodiques complets.

A l'étape actuelle, cela est impossible. Les systèmes de synthèse sont capables de faire la distinction entre une dernière syllabe de fin ou de continuation (la troisième colonne du tableau dans l'exemple "*Je le lui ai dit hier*"), mais ils ne sont pas capables de reproduire les finesses de sens incluses dans les intonations des deux premières colonnes.

Comme il est très désagréable d'entendre une voix monocorde, les chercheurs résolvent le problème en donnant à la machine les informations simplifiées nécessaires pour produire ce qu'on appelle une "prosodie acceptablement neutre". Cette prosodie correspond à l'intonation lors de la lecture d'une phrase citée hors de son contexte. Le résultat n'est pas aussi riche que la voix humaine, mais il peut néanmoins être d'une qualité tout à fait surprenante. Ainsi, lorsque le logiciel *Bright speech* de Babel Technologies "lit" un texte, on a vraiment l'impression, pendant quelques minutes, que c'est une personne en chair et en os qui s'exprime.¹⁰

UN DEFI COMPLEXE

On le voit: la synthèse de la parole est un défi extrêmement complexe. Pour s'y retrouver et progresser, les chercheurs tentent de "séquencer" méthodiquement le processus. On distingue deux grandes phases:

¹⁰ En français et en anglais, le caractère synthétique de la voix n'apparaît qu'après un certain temps de "lecture". Bright speech est un peu moins convaincant (pour le moment) en néerlandais et en espagnol, tout en restant d'une très grande qualité. Le système fonctionne en dix-huit langues: allemand, anglais (US et UK), arabe, danois, espagnol, français, finois, islandais, italien, néerlandais (B et NL), norvégien, portugais (P et Br), suédois, tchèque et turc.

- *Natural Language Processing* (NLP). La fonction de cette phase est de transformer le texte de départ en informations sur la succession de phonèmes et sur la prosodie;
- *Digital Signal Processing* (DSP). Ici, il s'agit de transformer l'information sur les phonèmes et la prosodie en sons, en parole de synthèse.

La NLP est divisée en plusieurs étapes:

- un **préprocesseur** découpe le texte en unités. Une définition de base de l'unité peut être: toute suite de lettres contenues entre deux séparateurs, un séparateur pouvant être un espace, un signe de ponctuation ou un signe diacritique (les accents et les apostrophes, par exemple). Néanmoins, le préprocesseur pousse la notion d'unité un peu plus loin. Des expressions régulières ¹¹ permettent de définir des unités plus complexes, comme les adresses URL, les numéros de téléphone ou les symboles monétaires (FS = "franc suisse", USD = "dollar américain"). Au sortir du préprocesseur, toutes les unités ont été séparées. Il n'empêche cependant que certaines unités seront par la suite reconnues comme faisant partie d'un tout qui n'aurait pas dû être divisé. C'est le cas lorsqu'une apostrophe fait partie d'un mot, comme dans "aujourd'hui". Cette unité sera donc reconstituée après le préprocesseur. Lorsque ce travail est achevé, on estime que l'on a correctement repéré des unités qualifiées de "linguistiques" parce qu'elles forment un tout au contenu cohérent;
- un rôle-clé est joué par l'**analyseur morphosyntaxique**. Au cours de cette étape, la machine propose tous les usages et fonctions possibles de chaque mot pris individuellement et décompose chaque mot en ses composantes possibles. Pour "*les poules du couvent couvent*", par exemple, la machine conclura que chacun des deux mots "*couvent*" peut être soit un nom, soit un verbe conjugué (avec "*couv-*" comme radical et "*-ent*" comme terminaison). Ensuite, l'analyseur réduira la liste des usages possibles en examinant chaque unité linguistique dans son contexte (les unités qui l'entourent). Sur base des règles syntaxiques qu'on lui a mises en mémoire, il conclura que seul le second "*couvent*" peut être la troisième personne du pluriel du verbe *couver*, au présent.
- La première phase se termine par un module de **phonétisation** (le texte est traduit automatiquement en phonèmes) suivi éventuellement d'un module **prosodique**. La phonétisation peut se faire notamment par modèle statistique. Dans ce cas, pour toute lettre, la machine parcourt un "arbre de décision". Chaque embranchement de l'arbre correspond à un choix. La machine calcule le nombre de combinaisons possibles (en ne retenant, à chaque embranchement, que les deux possibilités les plus probables), puis procède à une optimisation

¹¹ Expression régulière: notation faite de symboles et d'opérateurs qui permet de décrire un ensemble de mots qui forment un tout, un langage. Par exemple, l'expression " $^a.+$ " signifie: tous les mots qui commencent (^) par un 'a' (a) suivi de n'importe quelle lettre (.) au moins une fois (+).

des possibilités et décide si la lettre doit être prononcée ou pas, et comment, en fonction du contexte environnant.¹²

La seconde phase, *Digital Signal Processing*, peut se concevoir de deux manières différentes: synthèse par règles ou synthèse par concaténation.

- La **synthèse par règles** signifie qu'on met dans la mémoire de la machine toute une série de règles qui lui permettront de produire une approximation de la parole. Le problème de cette méthode est que, pour obtenir une voix d'un certain "naturel", il faut introduire un très grand nombre de règles, plus une série de paramètres de contrôle pour gérer l'intonation et la coarticulation. A un certain niveau, la masse de paramètres à gérer devient si importante que la coordination pose problème, ce qui se répercute négativement sur la qualité.
- La **synthèse par concaténation** tourne cette difficulté en mettant à disposition de la machine un vaste stock d'unités acoustiques (des diphones, voire des triphones). Comment ce stock est-il constitué? Tout simplement à partir d'enregistrements de voix humaines, qu'on découpe en petits bouts et qu'on sauvegarde dans des bases de données. Incroyable mais vrai: on peut aujourd'hui, sur l'internet, s'acheter une collection de diphones pour telle ou telle langue auprès d'une base de données. Des sites vous proposent de faire votre choix dans leur catalogue: français de Belgique ou de France? Voix d'homme ou de femme? Pour avoir une idée de la masse de travail nécessaire, il faut savoir que la synthèse d'un anglais correct nécessite au minimum 1.000 diphones. Mais il est plus facile de mettre en mémoire un bon millier de phonèmes que les 20.000 à 30.000 mots d'un dictionnaire. De plus, on ne dépend plus du locuteur. Enfin, la coarticulation est incluse dans les signaux utilisés...

"On le constate aujourd'hui: plus on réduit le traitement du signal, plus on augmente la qualité, le naturel du signal produit", nous dit Richard BEAUFORT. "C'est pourquoi, chez Multitel, nous travaillons sur la synthèse par concaténation. Dans ce cadre, nous essayons de nous passer du module prosodique de la NLP: c'est possible, dans la mesure où les diphones de la base de données incluent des paramètres prosodiques". Linguiste de formation, Richard BEAUFORT est chercheur chez Multitel, où il travaille en duo avec son collègue Laurent COUVREUR, polytechnicien spécialiste en reconnaissance vocale. Tous deux sont des collaborateurs du Professeur Thierry DUTOIT.

LE 22, A ASNIERES?

¹² Pour les matheux: l'optimisation combinatoire se fait à l'aide d'un algorithme basé sur la théorie de Belman, qui dit que la solution optimale d'un problème combinatoire est la somme des solutions optimales des problèmes partiels.

Que ce soit en synthèse ou en reconnaissance vocales, et en dépit des progrès réalisés, les machines sont **très loin d'égaliser les performances du cerveau**. Les chercheurs seront encore longtemps de merveilleux fous parlant avec leurs drôles de machines. Mais leurs progrès sont rapides et, dès maintenant, leurs systèmes ont des applications extrêmement utiles.

- Stephen HAWKING n'est pas un cas isolé. Avec l'aide d'un clavier adapté et d'un programme de "prédiction" des mots¹³, des **personnes handicapées** peuvent parler à leur entourage. La technologie existe. Grâce au projet européen FASTY, elle sera bientôt opérationnelle non seulement en anglais mais aussi en allemand, français, suédois et néerlandais. Jeune diplômé de l'Université de Valenciennes, travaillant actuellement dans le cadre de Multitel, Xavier RICCO parle avec passion de la façon dont FASTY cherche à coller au plus près aux besoins spécifiques de chaque personne handicapée: *"Nous sommes neuf partenaires européens. Multitel est 'work package leader' pour les aspects vérification et validation. Nous travaillons avec un département universitaire de psychologie, qui teste le matériel auprès de personnes handicapées. Cette démarche interdisciplinaire est intéressante. Et, humainement, le projet est très motivant"*.¹⁴ FASTY est dans sa phase terminale. Le système sera mis sur le marché prochainement, pour un prix ne dépassant pas celui d'une machine à laver bon marché.¹⁵
- Les systèmes de reconnaissance et de synthèse se révèlent d'ores et déjà très précieux dans **l'apprentissage des langues**. Certes, rien ne remplacera jamais le dialogue avec un bon prof de langues, mais les machines peuvent donner une aide appréciable. Des "dictionnaires parlants" nous disent comment prononcer les mots, et même les phrases. Des logiciels de reconnaissance vocale peuvent nous aider à tester et à perfectionner notre prononciation des *"though"*, *"through"*, *"cough"*, *"tough"* et autres *"enough"*. Des logiciels de synthèse peuvent prononcer ces mots au ralenti, voire synchroniser la prononciation avec l'image d'un enseignant virtuel, qui nous montre comment articuler chaque mot;
- Dans le secteur des **télécommunications**, une nouvelle petite révolution est à l'ordre du jour. Dans un avenir proche, il nous sera possible d'appeler n'importe quelle base de données par téléphone et d'interroger des ordinateurs qui, d'une belle voix synthétique, répondront infailliblement à toutes nos questions. Si Fernand Reynaud créait aujourd'hui son mémorable sketch, il ne demanderait pas à la standardiste de lui passer "le 22 à Asnières"; par contre une machine pourrait lui dire le nom de l'abonné qui possède le numéro de téléphone 22.22.222 à Tokyo, ou n'importe où sur la planète... L'histoire ne dit pas ce que

¹³ "Prédiction des mots": sur base des trois premières lettres tapées, la machine propose le choix entre plusieurs mots complets, afin d'accélérer la formation des phrases. L'économie de frappe peut atteindre 60%.

¹⁴ Le travail avec les handicapés vise à la validation optimale des systèmes. Il est réalisé au département de psychologie des Facultés Universitaires Notre-Dame de la Paix, à Namur. Les demandes des utilisateurs portent surtout sur l'économie de mouvements et la vitesse de prise de notes. Pour plus de renseignements sur FASTY: <http://fortec.tuwien.ac.at/fasty>

¹⁵ Les aveugles aussi peuvent tirer parti de la synthèse vocale. Babel Technologies a mis au point un logiciel qui permet aux mal-voyants de lire le contenu des sites web. Le prix du matériel varie en fonction de la langue. En Suède, les frais sont pris en charge par la sécurité sociale...

Reynaud et la standardiste penseraient de cette évolution qui, reconnaissons-le, réduit considérablement la part du comique dans les télécommunications...

Alain Tondeur
Le 02/07/03

(Encadré)

Les performances de notre cerveau

Si le processus de gestion de la parole par le cerveau reste mystérieux, par contre, on connaît les parties du cerveau qui interviennent dans le processus. En effet, on constate que:

- **des personnes ont la capacité de parler mais pas de comprendre, ni de la parole, ni du texte** (leur parole est vide de sens). Cela correspond à une lésion d'une certaine zone du cerveau, appelée aire de Wernicke et située dans l'hémisphère gauche;
- **des personnes ont la capacité de comprendre de la parole et du texte, mais sont handicapées gravement dans la construction de phrases complexes** (elles "parlent" par mots clés, sans conjugaison, etc.). Ce type de trouble coïncide avec une lésion d'une autre partie du cerveau, laire de Broca, également située dans l'hémisphère gauche mais plus en avant, vers le front;
- **des personnes ont la capacité de comprendre et de parler spontanément, mais pas la capacité de lire un texte à haute voix**. Ce handicap est appelé "aphasie de conduction" car il résulte de la lésion du faisceau de neurones qui unit les aires de Wernicke et de Broca;
- **des personnes ont la capacité de comprendre mais sont incapables de produire une parole articulée (lue ou spontanée)**. Dans ce cas, on est en présence d'une lésion dans les régions du cortex moteur, qui commande les muscles intervenant dans l'articulation.

Dans les machines, les tâches de reconnaissance et de synthèse vocale sont séparées. Dans le cerveau, tout en étant distinctes, ces deux fonctions sont liées physiologiquement. Selon les spécialistes, voici très schématiquement comment les choses semblent se passer:

1. les stimuli visuels (lecture) ou auditifs (écoute) sont intégrés, puis transmis à l'aire de Wernicke;
2. là, ils sont interprétés une première fois et transformés pour ainsi dire en "images". A noter que ces "images" seraient assez peu liées à la langue;
3. le faisceau de neurones transmet ces "images" à l'aire de Broca;
4. dans l'aire de Broca, une deuxième interprétation a lieu, plus analytique, qui vise à énoncer une phrase;

5. le cortex moteur active les muscles articulatoires: je parle, tu parles, nous parlons!

Quel que soit le mécanisme exact, une chose est certaine: notre cerveau est extrêmement performant. Nous pouvons comprendre des gens qui parlent avec un accent épouvantable, nous pouvons suivre une conversation lorsque plusieurs personnes parlent, ou au milieu d'un bruit ambiant important, nous pouvons déchiffrer l'écriture du médecin et même lire un message parmi plusieurs écritures superposées. De plus, quand nous parlons, nous écoutons ce que nous disons et nous ajustons notre parole en fonction de ce retour d'écoute. Tout cela, les machines sont – et resteront ?- incapables de le faire.