

L'INTELLIGENCE ARTIFICIELLE :

L'ORDINATEUR PARLE suite/
voir n°1414

JUSQU'À la fin du XVIII^e siècle — écrit Jean-Sylvain Liénard dans sa thèse sur l'analyse, la synthèse et la reconnaissance de la parole* — les machines parlantes furent un prétexte à légendes et mystifications ».

Aujourd'hui, les ordinateurs, instruments de calcul et de manipulation des informations, doivent communiquer avec le monde extérieur, et en particulier avec l'homme, tant pour recueillir des données, que pour faire connaître les résultats de leurs travaux. Les supports physiques du dialogue peuvent maintenant simuler les moyens mêmes dont l'homme dispose pour communiquer. C'est ainsi que l'ordinateur a acquis le sens de la vision, grâce auquel il peut lire les textes imprimés. Il a aussi appris à parler.

Si l'on en croit les experts américains de l'Organisation Diebold, spécialistes en étude de marchés pour des applications nouvelles, les unités de réponse vocale sont promises à un très brillant avenir. D'ici peu, cette forme de « conversation » peut être le complément de petits terminaux de très grande diffusion, simplement composés d'un clavier* connectable acoustiquement à un poste téléphonique, et où elle constitue la voie de retour du dialogue et ce, de façon beaucoup plus économique que toute visualisation ou impression.

Dans un avenir beaucoup plus lointain, lorsque les unités de reconnaissance de la parole seront devenues opérationnelles, il est évident que la véritable conversation vocale constituera pour les utilisateurs, un moyen idéal de dialogue. Il est peut être aussi évident que ceci deviendrait le cauchemar des concepteurs de software, car après avoir eu bien des difficultés à reconnaître

des mots, il faudra interpréter la sémantique, et, ce au prix d'une syntaxe que les utilisateurs voudront de moins en moins rigide.

UN PEU D'HISTOIRE

Les premiers essais scientifiques de machines parlantes peuvent être attribués à trois inventeurs : l'abbé Mical, Kratzenstein et W. Von Kempelen.

En France, l'abbé Mical construisit une machine dont les « deux têtes parlantes » échangeaient, en manière de dialogue, les phrases suivantes :

— Le roi a donné la paix à l'Europe.

— La paix fait le bonheur des peuples.

Au Danemark, Kratzenstein présenta une série de cinq résonateurs imitant les voyelles *a, æ, i, o, u*. Ces résonateurs, de formes variées, avaient à peu près le volume de la cavité buccale.

La machine de Kempelen, gentilhomme de la cour d'Autriche-Hongrie, a le grand mérite d'exister encore ; J.S. Liénard a pu la reconstituer à partir de ses descriptions : l'opérateur appuie son coude droit sur le soufflet de la machine et sa main droite commande divers leviers et ouvertures se trouvant sur le dessus de la boîte à vent. A l'intérieur de celle-ci se trouve une anche qui vibre quand la pression d'air est suffisante et des soupapes commandées par des leviers qui permettent à l'air de passer à travers des sifflets [son : « s »]. Le son « r » est produit par un levier qui amène une aiguille de laiton au contact de l'anche, introduisant des perturbations dans son fonctionnement, et il en résulte une sorte de raclement. Les voyelles s'obtiennent en faisant varier, avec la main gauche, les caractéristiques du résonateur en caoutchouc qui constitue « la

bouche » : complètement ouverte, elle donne un son voisin de « a » ; à moitié fermée, on peut entendre « œ » ; presque complètement fermée, on entend « u ».

Cette machine est un véritable « instrument de parole » dont il faut apprendre à jouer.

« Je lui fais prononcer, écrivait Kempelen, plusieurs centaines de mots clairement et distinctement. Par exemple : papa, maman, marianna, roma, maladie, santé, astronomie, anatomie, chapeau, racine, supé, charmante, opéra, comédie, pantomime, et aussi des mots longs et difficiles, tels que Constantinopolis, Monomotapa, Mississippi, Astrakan, Anastasius, etc. ».

Joseph Faber présenta, en 1835, à Vienne, une machine plus élaborée, qui parlait à voix haute et à voix chuchotée, et chantait « God Save the Queen » ! Après avoir été montrée pendant un demi-siècle dans les capitales du Vieux et du Nouveau Monde, la machine de Faber, a, selon toute vraisemblance, terminé sa carrière dans les caves de l'Ecole de médecine de Paris.

Après Faber, les physiiciens abandonnèrent la méthode empirique et globale, qui avait fait la fortune de Kempelen et de Faber. Pour imiter les voyelles Helmholtz utilisa des combinaisons de diapason, Koenig des sirènes, Millet et Stumpf des tuyaux d'orgue, Marage et Paget des résonateurs.

Ce n'est qu'en 1937 que Riesz présenta une machine parlante, dont les grandes lignes étaient semblables à celle de Faber, et qui produisait une parole courante et compréhensible.

En 1939, Dudley, un ingénieur de la Compagnie Bell Telephone, présenta le VODER (VOICE DEMONSTRATOR), banc de 10 filtres passe-bande répartis entre

300 et 3 000 Hz, et connectés en parallèle. Depuis, de nombreux travaux ont été entrepris tant en France qu'aux U.S.A., sur des montages électroniques similaires à celui de Dudley.

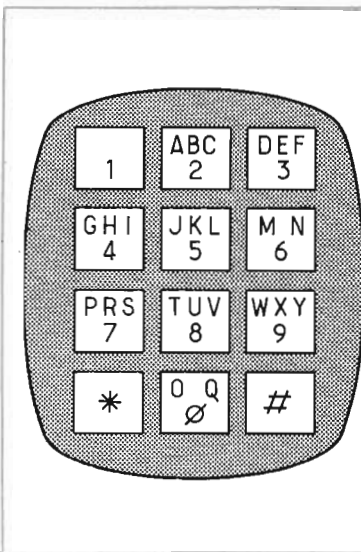
PHONEMES
ET PHONATOMES

La parole est très redondante sur le plan du signal proprement dit et sur le plan linguistique : 50 000 unités d'information par seconde environ sont transmises par une ligne téléphonique classique au cours d'une conversation alors que notre cerveau, selon certaines hypothèses, ne pourrait décoder, en fin de traitement, qu'un maximum de 50 unités d'information par seconde, ces 50 unités étant porteuses de données concernant aussi bien le message que la qualité et les caractéristiques individuelles de la parole.

On cherche donc à détecter ces 50 unités d'information fondamentales qui permettent la reconnaissance, par le cerveau, et on cherche à reconstituer la parole à partir d'un même nombre d'unités d'information.

Cette réduction considérable des informations doit grandement faciliter la mise en œuvre de systèmes de reconnaissance de la parole, et la construction d'unités à réponse vocale. On pourra ainsi multiplier par un facteur de l'ordre de 1 000 le nombre des communications parlées, tout en conservant le même nombre de lignes téléphoniques.

La phonétique considère depuis longtemps l'existence de sons élémentaires appelés « phonèmes » : en français, on dénombre environ 36 phonèmes comprenant 16 voyelles et 20 consonnes ou semi-voyelles. Dans la prononciation courante, certaines différences sont peu significatives. Ainsi, IN et UN, E et EU..., sont



Touches	*	*	
1	- Soustraction	SIN sinus	Appel du S.C.T.
2	* Multiplication	COS cosinus	Demande de répétition
3	Division	ATN arc tangente	Relance du S.C.T.
4	Elevé à la puissance	LGN logarithme népérien	X
5	(Facteur de	EXP exponentielle	Y
6) Fin de parenthèses	TGH tangente hyperbolique	Z
7	Point	RAC racine carrée	Effacement d'un nombre
8	E Dix à la puissance	ABS valeur absolue	Rappel du registre T
9	+ Addition	NEG opposé	M
0	;	FOR format	Abandon du S.C.T.

(b)

Le codage des instructions à partir du clavier S.C.T. :
 La touche correspond à une commande émise au superviseur ;
 La touche correspond à l'émission d'opérateurs ;
 L'emploi séquentiel de ces deux touches correspond à l'émission de fonctions plus complexes.

ils souvent confondus, sans aucun dommage pour l'intelligibilité de la parole. Il semble donc qu'un ensemble de 30 phonèmes, ou peut-être moins, soit suffisant pour décrire, sans ambiguïté, la prononciation française.

Pour synthétiser le mot « chat » par exemple, enregistrons, sur une bande magnétique un « ch... » prolongé, puis un « aaa... ». Par un montage simple, il est possible de juxtaposer deux petits morceaux de bande magnétique, contenant chacun, l'un des deux sons précédents. Le résultat n'est pas le mot « chat » : on entend simplement les deux sons consécutifs.

L'information se trouve à la transition d'un phonème à l'autre, et non dans les phonèmes eux-mêmes. L'association de deux phonèmes est désignée par le vocable « phonatome ». Le mot « chat » précédent est constitué du seul phonatome « Cha ».

Un phonatome est également appelé un « diphone » ou un « di-phonème ».

Il suffit d'établir une fois pour toutes un dictionnaire des caractéristiques acoustiques des phonatomes, pour réaliser la synthèse efficace de la parole : il suffit alors d'assembler ces éléments par leur extrémité commune, exactement comme on assemble des dominos.

L'association de 30 phonèmes deux à deux donne lieu à 900 phonatomes possibles. Une étude statistique effectuée sur des textes a montré que 200 d'entre eux n'étaient pas utilisés. En outre, certains phonatomes peuvent être considérés comme réversibles (ME et EM). Il apparaît pratiquement qu'un répertoire de 400 phonatomes est suffisant pour reconstituer correctement la prononciation du français.

Un ingénieur du C.N.R.S., Daniel Teil a écrit un programme,

amélioré ensuite par H. Lucot, fabriquant des mots aléatoires, en juxtaposant des phonèmes. Les mots nouveaux, ainsi créés par l'ordinateur, présentent la caractéristique de « sonner français ». Voici quelques échantillons de ces mots générés par ordinateur, retranscrits dans une orthographe arbitraire :

- Mots de 3 phonèmes : déa, nance, laule, vran, nure ;

- Mots de 4 phonèmes : anure coufan, reté, vobi, boil, zizan, lora, luzan, fouaque ;

- Mots de 5 phonèmes : la-rare, bousque, omance, massire, vesti, filure, aillard, fonelle, laque-fan...

POURQUOI SYNTHETISER LA PAROLE ?

L'intérêt porté à la possibilité d'un dialogue acoustique entre l'homme et la machine n'est pas purement spéculatif. Des applications immédiates sont au contraire possibles et trouveraient des débouchés importants, même en s'en tenant à des objectifs de portée réduite.

Imaginons un ingénieur, assis devant une console à écran cathodique, unité périphérique d'un ordinateur. Il dessine un projet sur l'écran, l'ordinateur en analyse les caractéristiques structurales, calcule les contraintes... Le dialogue serait certes facilité s'il suffisait à l'ingénieur de demander, oralement « moment fléchissant ? » et l'ordinateur de répondre : « 50 mètres x kilogrammes ».

Imaginons encore une entreprise de distribution ayant pour tâche de desservir un grand nombre de détaillants, à partir d'un ensemble d'entrepôts régionaux. Un ordinateur assure la gestion du stock ; il connaît, à tout moment, l'état des disponibilités de chaque magasin. Le détaillant, de son côté, a besoin de se procurer l'article dont il

peut assurer la vente à un moment donné. Il aura accès à l'information, en téléphonant à l'ordinateur : ayant posé sa question sous forme codée, soit à partir d'un clavier associé au poste, soit à partir du cadran, l'utilisateur reçoit immédiatement une réponse orale.

Le Groupement informatique et transmission de données du Centre national d'études des télécommunications a étudié et réalisé, pour sa part, un système composé d'un ordinateur central et de terminaux (dont une unité de réponse vocale), destiné à assurer un service de calcul de bureau **. Le terminal d'entrée-sortie est un simple poste téléphonique à clavier. La base du système est composée d'un calculateur T2000 de la Télémécanique qui permet de gérer 8 lignes entrantes en temps partagé. Le système autorise trois types de calculs :

- En mode « calcul direct », le système permet d'effectuer des opérations usuelles telles que l'addition, la soustraction, la multiplication, la division, l'élevation à une puissance, la valeur absolue, l'opposé d'un nombre, ainsi que des opérations plus complexes, telles que sinus, cosinus, arc tangente, logarithme népérien, exponentielle, tangente hyperbolique, extraction de racine. Les expressions de calcul sont émises dans un langage dont la syntaxe est très proche de l'écriture mathématique usuelle.

- En mode « calcul paramétré », trois paramètres (x, y, z) sont disponibles.

- En mode « calcul itératif », le système demande le nombre de répétitions que l'on veut effectuer avant d'avoir le résultat final.

Les réponses de l'ordinateur apparaissent sous forme orale : l'unité de réponse vocale utilise un vocabulaire de base de 33 mots (chaque mot correspon-

dant à un son de base : zéro, un, ...dix, vingt...), ce qui permet de reconstituer un nombre inférieur au million.

D'autres applications des unités de réponse vocale concernent les machines à enseigner. Là, l'élève, assis devant un pupitre, communique personnellement avec la machine. Certains messages imprimés sont complétés par des commentaires oraux. Le texte parlé, plus souple dans ses intonations, permet de mieux nuancer les observations qu'une simple phrase dactylographiée. C'est là un aspect psychologique qui peut être précieux dans l'enseignement programmé.

Dans un autre domaine, considérons les problèmes de la surveillance des avions en vol, et du guidage des cosmonautes. Le contrôleur au sol n'agira bientôt plus que comme traducteur. Il lui sera seulement demandé de lire des messages écrits, reçus d'un ordinateur, pour les communiquer, par radio, au pilote : verra-t-on un jour l'ordinateur au sol parler directement au pilote ?

Un exemple pratique de système fonctionnant dans le domaine précédent est fourni par le système DECLAM de la C.I.T.-Alcatel. Il s'agit d'un dispositif destiné à émettre par radio des informations météorologiques à l'usage des avions commerciaux. Ces informations arrivent de divers points du territoire, et sont acheminées vers un ordinateur central par liaisons télégraphiques. A partir des données qu'elles contiennent (température, pression...) l'ordinateur compose un message vocal.

* Cette thèse a été soutenue le 21 avril 1972 à l'Université Paris VI.

** On lira avec intérêt l'article de C. Carrouge : « Le service de calcul par téléphone : le système SCT » paru dans l'Echo des recherches, en janvier 1973.

166

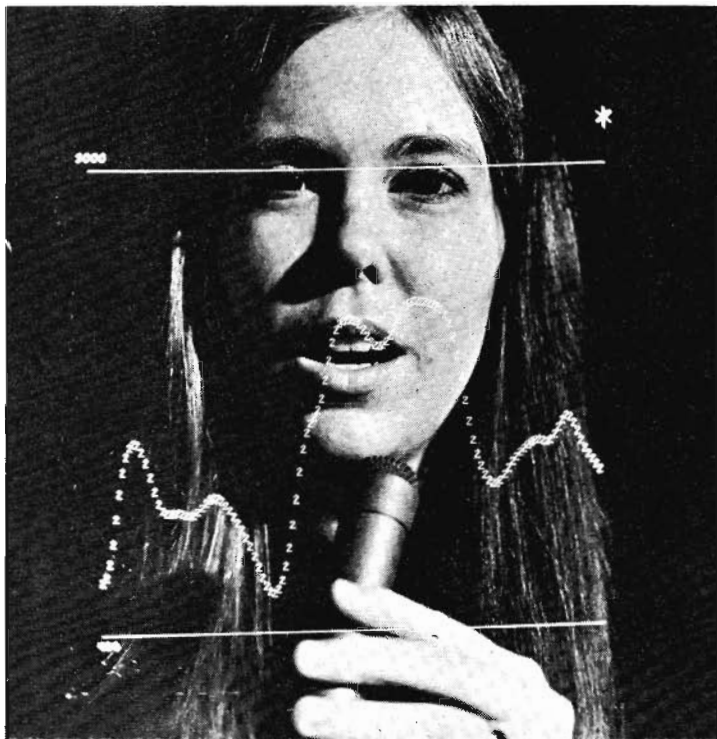


Photo 19. — Les mots prononcés par Judy Johnson, de Bell Laboratories, sont analysés, convertis en informations numériques et stockés dans la mémoire d'un ordinateur. Les techniques permettent de sélectionner celles des informations numériques absolument nécessaires à la compréhension de la parole, afin d'occuper le minimum de place en mémoire.

BIEN D'AUTRES APPLICATIONS ENCORE !

La synthèse de messages vocaux élaborés par ordinateur peut être utilisée pour la gestion des stocks, la réservation des places, dans les gares, les centres de calculs, les agences de presse... L'abonné au téléphone aux prises avec des difficultés de trafic ou de numérotation peut entendre des conseils enregistrés et l'horloge parlante est un exemple de composition de message à partir d'un nombre de mots restreints. D'autres services ont été étudiés ; ainsi une unité de réponse vocale doit permettre de

fournir aux abonnés, sous forme vocale, et de façon entièrement automatique, le coût de leur dernière communication téléphonique, et le contenu de leur compteur ; ce service est le premier d'une liste qui pourra être complétée (service du réveil, changement de numéro, par exemple).

Tout système automatisé dont le fonctionnement doit être suivi par plusieurs personnes ayant une activité annexe, bénéficierait des services d'unités de réponse vocale :

— Les personnes concernées ne seraient plus rivées à l'écran cathodique ou à la machine télé-



Photo 20. — Au service de l'enseignement programmé : le texte parlé, plus souple dans ses intonations, permet de mieux nuancer les observations qu'une simple phrase dactylographiée. (Cliché G.E.).

— Le coût d'une extension géographique est minime (coût d'un haut-parleur : quelques dizaines de francs ; coût d'un télétype : 10 000 francs ; coût d'une console à écran cathodique : au moins 20 000 francs !).

Un aspect non négligeable de cette application serait la possibilité offerte aux aveugles de s'adapter à toutes les activités utilisant l'informatique.

Miniaturisé, un synthétiseur de parole pourrait également servir de prothèse pour muets.

(à suivre)

Marc FERRETTI

UNE REVUE SE CONSACRE AUX ROBOTS INDUSTRIELS

Dans le cadre de cette série d'articles sur l'intelligence artificielle, nous devons de signaler la parution, dès le 14 septembre 1973, de la première revue consacrée exclusivement aux robots industriels.

« The Industrial Robots » sera une revue trimestrielle, et le rédacteur en chef sera un universitaire : le professeur W.B. Heginbotham, spécialiste des robots industriels à l'Université de Nottingham (Grande-Bretagne).

L'abonnement aux six premiers numéros de cette revue coûte 22 livres britanniques. Pour s'abonner, écrire à :

The Industrial Robot
International Fluidics Services Ltd
71 The Moor
Carlton, Bedford, MK43 7 JS
Angleterre.

Pour votre collection, procurez-vous

- LA RELIURE « HAUT-PARLEUR » (Marron)
- LA RELIURE « HI-FI STÉRÉO » (Bleu)
- LA RELIURE « ÉLECTRONIQUE PROFESSIONNELLE » (Rouge)

Au prix de **10 F** l'une + 2,50 F de port

Adressez commande au :

LE HAUT-PARLEUR

2 A 12, RUE DE BELLEVUE - 75019 PARIS
TÉL. : 202-58-30 C.C.P. 424-19 PARIS

A NICE JEAN COUDERT

vous présente
le plus grand choix
aux meilleurs prix...

TOUS LES MATÉRIELS
HI-FI

ainsi que les KITS
accessoires, haut-
parleurs, etc.

Service après-vente

INSTALLATION GRATUITE - CRÉDIT

JEAN COUDERT

85, bd de la Madeleine
06-NICE - Tél. ; 87-58-39

145