

L'INTELLIGENCE ARTIFICIELLE

L'ORDINATEUR ÉCOUTE

(Suite voir n° 1433)

LES premiers travaux sur la reconnaissance automatique de la parole remontent à la dernière guerre, puisque en 1943, paraissait un article de L.L. Mjasnikov sur la « reconnaissance objective des sons de la parole ». Un an après, J. Dreyfus-Graf commençait ses recherches sur le sonographe à l'École des Arts et Métiers de Genève. Il devait venir les poursuivre, beaucoup plus tard, entre 1966 et 1971, à Lannion, par la construction du phonétographe IV et la conception du phonétographe V.

Entre temps, de nombreuses équipes s'étaient mises sur les pas des pionniers, d'abord en Grande-Bretagne et aux États-Unis (Smith, Fry, Denes), puis au Japon (Sakaie, Inoue) et, plus tard, en Europe continentale.

En France, on trouve, à l'origine des premières recherches, l'intérêt de spécialistes de disciplines différentes : S. Castan et G. Pérennou, mathématiciens de l'Université de Toulouse, J.-P. Tubach, informaticien du Centre d'Études pour la Traduction Automatique de Grenoble, G. Ferriou, J.-M. Person, J. Poncin et G. Roux, acousticiens et informaticiens au Centre National d'Études des Télécommunications (C.N.E.T.). Ces spécialistes, qui se sont retrouvés au sein de groupes tels que le G.A.L.F. (Groupement des Acousticiens de Langue Française) ou l'A.F.C.E.T. (Association Française pour le Cybernétique Économique et Technique) ont été rejoints rapidement par d'autres équipes : celle de M. Lamotte à Nancy, puis celle de C. Guéguen du nouveau laboratoire d'Automatisme à l'École Nationale Supérieure des Télécommunications. Actuellement une dizaine

d'équipes se sont constituées qui appartiennent, à deux exceptions près, au secteur public.

RECONNAITRE DES FORMES FLOUES

Aux États-Unis, après une mode qui a duré jusqu'aux premiers échecs sérieux, entre 1962 et 1968, les calculateurs numériques électroniques entraînent dans les mœurs, et où la reconnaissance des formes recevait un fondement mathématique, beaucoup de chercheurs se sont découragés. Certains ont abandonné; d'autres se sont tournés vers des études de type linguistique ou phonétique, dont le caractère fondamental est plus affirmé.

La reconnaissance des formes consiste à attribuer une classe d'appartenance, ou catégorie à des entités telles que lettres de l'alphabet, photographies ou phonèmes constitutifs de la parole.

La théorie de la Forme est née en Allemagne, à la fin du siècle dernier. Elle constitue une réaction contre la psychologie analyti-

que du XIX^e siècle, qui réduisait les faits de conscience et de conduite à des éléments (les sensations), et à leurs associations au sens où l'entendait Pavlov. Le comportement humain serait, ainsi, ramené à un ensemble d'automatismes simples, dont la plupart seraient acquis par apprentissage, depuis la naissance.

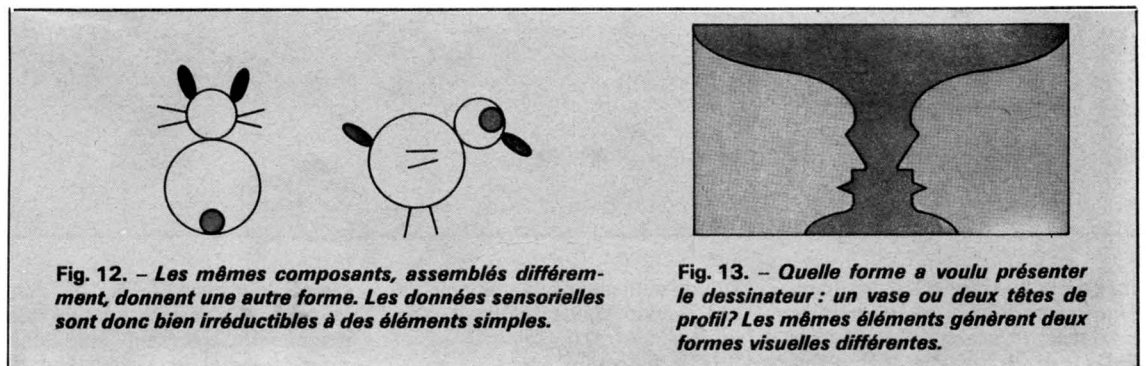
La psychologie de la forme, telle qu'elle a été énoncée par Von Ehrenfels, K'bler, Wertheimer, Koffka,... se place dans un tout autre plan : elle considère les données sensorielles comme irréductibles à des éléments simples; les faits psychiques s'organisent spontanément en formes, ou « totalités » perceptives dont les propriétés d'ensemble sont autre chose que la somme des propriétés des éléments constitutifs.

Ainsi, par exemple, la lettre « a » constitue-t-elle une forme. Qu'on l'écrive en majuscule ou en minuscule, en lettre d'imprimerie (avec les multiples variantes offertes dans les tailles et types de caractères d'imprimerie), ou manuscrite, le signe « a » aura toujours la même signification et sera toujours perçue comme étant la première lettre de l'alphabet.

Étudiant les groupements de points ou de taches élémentaires, les partisans de la théorie des formes ont mis en évidence le rôle de la ressemblance et de la proximité des éléments dans leur structuration perceptive. On trouve des illustrations de ce rôle dans le domaine auditif : il suffit de faire dire chaque mot d'une même phrase par une voix différente; la forme (ici : la phrase) perd de son évidence pour un auditeur non prévenu.

L'assemblage de formes, par juxtaposition ou superposition, peut donner lieu à l'apparition d'une forme nouvelle : c'est une superforme. Ainsi, une phrase écrite est une superforme par rapport au mot, qui, lui-même, est une superforme par rapport à l'alphabet; celui-ci, à son tour, se trouve être une superforme, comparée aux formes que sont les points et les traits.

La forme se trouve généralement dans un environnement : c'est le fond d'une image sur lequel se détache une figure, ou encore le bruit de fond dans lequel est noyé un signal électronique. La figure, le signal, possèdent leur organisation interne; ils ont des



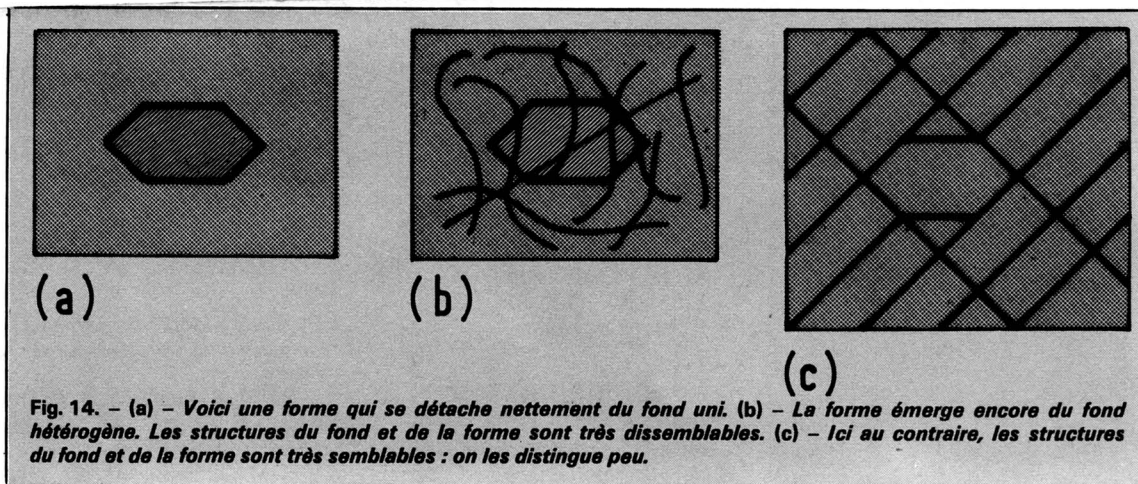


Fig. 14. - (a) - Voici une forme qui se détache nettement du fond uni. (b) - La forme émerge encore du fond hétérogène. Les structures du fond et de la forme sont très dissemblables. (c) - Ici au contraire, les structures du fond et de la forme sont très semblables : on les distingue peu.

contours, qui en sont partie intégrante. Souvent, le fond paraît informe et indéfini. Dans certains cas (Fig. 13), fond et forme peuvent être perçus alternativement.

L'émergence d'une forme sur un fond est essentiel pour sa perception. La distinction d'une forme est évidemment d'autant plus difficile que son organisation est voisine de celle du fond (Fig. 14).

Une même forme a souvent des contours différents, et même parfois des contours flous. Prenons le cas d'une forme acoustique : Un son, un mot ou une phrase parlés représentent en effet une forme, car ils transportent un message, comme les formes visuelles. Le problème de la reconnaissance des sons s'apparente en conséquence à celui de la reconnaissance des images.

Si l'on demande à un même locuteur de prononcer la même phrase, de manière identique, à quelques instants d'intervalle, on peut s'attendre à ce que le signal physique soit sensiblement identique dans les deux cas : une analyse spectrale révélera cependant, toujours, des différences qui peuvent être considérables, notamment dans la répartition des durées des sons composants, dans l'évolution du spectre, et dans l'intensité sonore; le locuteur peut adopter divers types de voix (voix parlée ordinaire, voix chuchotée, chantée, criée,...), divers timbres, divers rythmes, etc. Dans ce cas, il est très difficile d'extraire automatiquement du signal, les paramètres physiques qui caractérisent le sens du message.

D'un locuteur à l'autre, on observe *a fortiori* ces mêmes différences, auxquelles s'ajoutent les différences d'ordre linguistique (accents régionaux), morphologiques (dimensions du conduit vocal variant selon l'âge, le sexe,...) et physiologique (habitudes aléatoires).

Les formes acoustiques apparaissent ainsi extrêmement complexes, voire floues et leur

reconnaissance automatique (c'est-à-dire la recherche, par une machine, de leurs éléments caractéristiques) est d'une complexité inouïe.

Actuellement, une machine automatique n'est capable de reconnaître que quelques locuteurs; la reconnaissance de formes acoustiques produites par un grand nombre de locuteurs fait encore l'objet de recherches fondamentales.

Il est probable que les premières applications de la reconnaissance de la parole consisteront à poser des questions à un ordinateur. Jusqu'à présent, les langages de programmation, qui autorisent le dialogue homme-machine, sont des langages écrits. En vue d'améliorer les performances d'unités de reconnaissance de la parole, et de leur permettre de reconnaître plus de locuteurs, il est souhaitable d'étudier des langages de programmation qui soient des langages parlés, tenant compte de certaines contraintes liées aux erreurs de reconnaissance les plus fréquentes. Des recherches sont effectuées, au C.N.E.T., sur de tels langages.

CAPTER, TRAITER, DÉCIDER

Toute unité de reconnaissance se composera de trois classes d'outils : un capteur, fournissant une image des formes à reconnaître; un organe de mise en forme de cette image (cadrage, lissage, normalisation, extraction de paramètres), et enfin un organe de décision dans lequel sont stockés des critères de classification des formes.

Les capteurs que l'on utilise souvent sont :

- des capteurs optiques (pour les formes optiques) : caméra dont la sortie est discrétisée ou flying-spot;

- des capteurs acoustiques (pour les formes acoustiques) : microphone à charbon, par exemple, envoyant la parole à l'ordinateur par une ligne téléphonique. Signalons que le C.N.E.T. s'est livré à des expériences de reconnaissance de la parole « à distance », entre les centres de Paris et de Lannion : dans l'état actuel des méthodes de reconnaissance et des transmissions (bruits de commutation, diaphonie,...) lorsque l'appareil de mise en forme est situé à distance, les résultats sont médiocres.

Le niveau sonore de la parole est une grandeur extrêmement variable : un même locuteur peut parler plus ou moins fort, s'approcher du microphone, tourner la tête; les variations du niveau sonore atteignent rapidement plusieurs décibels, (voire 20 dB) et ces variations sont gênantes en reconnaissance automatique. Une régulation dynamique permet de maintenir sa constante, sur des intervalles de temps prédéterminés (50 à 100 millisecondes, par exemple), le niveau sonore moyen. Cette régulation peut, en outre, accentuer certaines modifications spectrales significatives du message (bruits d'impact rencontrés lors de la prononciation de plosives).

L'unité de mise en forme des signaux aura pour tâche d'élaborer des paramètres significatifs, destinés à représenter le contenu du signal vocal. Jusqu'alors, ce dernier était analysé soit par un banc de filtres, soit par un appareil équivalent comprenant un seul filtre, le signal d'entrée étant transposé en fréquence. Cet appareil, le sonographe, permet de disposer, sous une forme facile à interpréter, d'un grand nombre de renseignements, sur les évolutions des formants par exemple.

Les techniques d'analyse employées actuellement découlent des objectifs que l'on se fixe. On espère, par exemple, mettre en évidence tel ou tel phénomène; on

espère qu'un traitement particulier renforce telle ou telle caractéristique. On cherche ainsi à disposer d'informations sur la phase des composantes : il semble, en effet, que les variations de phase jouent un rôle important lors de la perception de la parole.

QUE DOIT-ON RECONNAÎTRE?

Le cerveau humain perçoit le message parlé, continu par essence, comme une suite discontinue de mots; lorsque le contexte permet de prévoir les mots, l'opération dite de « compréhension d'un mot » se réduit à sa comparaison avec quelques mots sélectionnés; si le mot est imprévu, sa compréhension nécessite une recherche en mémoire.

Les systèmes actuels sont incapables de prendre en compte le sens du message. Ils opèrent suivant l'une ou l'autre des deux approches suivantes :

- reconnaissance analytique; la machine reconnaît des éléments phonétiques du message : phonèmes, phonateurs, syllabes. Il n'y a pas de limitation de vocabulaire, mais l'emploi de cette technique suppose nécessairement une prononciation correcte. La suite phonétique ainsi obtenue est ensuite regroupée en mots, chaque mot du répertoire étant stocké en mémoire dans sa notation phonétique.

- Reconnaissance globale : on suppose le vocabulaire limité à un certain nombre de mots dont les configurations sont mémorisées. Chaque mot à reconnaître est traité dans son ensemble.

La reconnaissance globale de mots isolés a actuellement pris le pas sur la reconnaissance analytique, car elle semble devoir être plus rapidement opérationnelle dans l'environnement des calculateurs.

(à suivre)

Marc FERETTI

Nous avons déjà signalé (Haut-Parleur n° 1424) l'existence d'un enseignement visant à donner une formation commune aux électroniciens, acousticiens, informaticiens ayant à traiter le signal de la parole, et aux praticiens de la parole (phonéticiens, linguistes, orthophonistes,...).

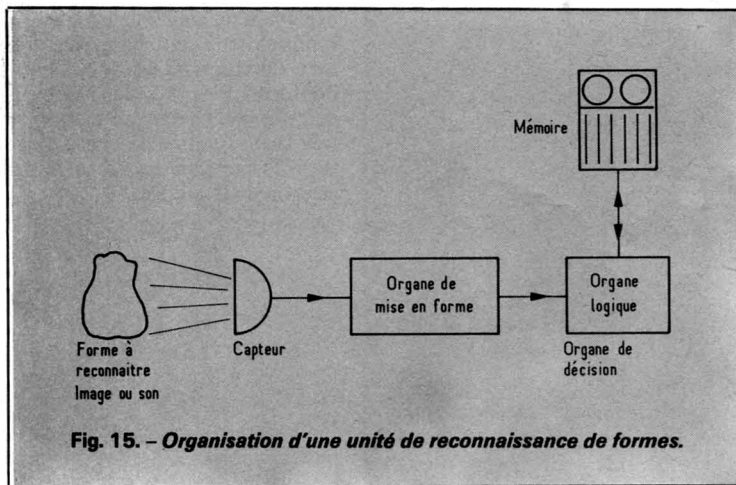


Fig. 15. - Organisation d'une unité de reconnaissance de formes.

Cet enregistrement a débuté le 21 novembre 1973 au laboratoire d'Acoustique de l'Université - Paris-VI, et s'achèvera en avril 1974. Les 40 heures du cycle d'enregistrement sont réparties en 24 heures de cours, exercices, discussions (les mercredis de 18 à 20 heures), 10 heures de conférences spécialisées, et 6 heures de démonstrations et travaux pratiques. Les conférences spécialisées seront données aux dates suivantes :

- 6 mars 1974 - Une nouvelle approche des problèmes d'intelligibilité de la parole, par E. LEIPP, Maître de Recherche au C.N.R.S.

- 18 mars 1974 - Analyse de la parole par modélisation récurrente, par C. GUEGUEN, professeur à l'École Nationale Supérieure des Télécommunications.

- 20 mars 1974 - Problèmes théoriques et expérimentaux concernant l'analyse intonative, par L.J. BOE, Ingénieur à l'Institut de Phonétique de Grenoble.

- 10 avril 1974 - Apprentissage de la parole : quelques aspects techniques de la méthode verbo-tonale par V. ARAMBASIN, Ingénieur électronicien à la Société S.E.D.I.

- 17 avril 1974 - Reconnaissance de la parole par ordinateur :

le projet E.A.R.S. de l'Université de Stanford (U.S.A.) par P. VICENS, Ingénieur à Compagnie Internationale pour l'Informatique.

CANDIDATURES ET RENSEIGNEMENTS COMPLÉMENTAIRES

S'adresser à :

M. J.-S. LIENARD
L.I.M.S.I. - B.P. 30
91406 - ORSAY
Tél. : 907-78-23, poste 33-95.

POUR EN SAVOIR D'AVANTAGE, ON LIRA...

« Reconnaissance de la parole et dialogue homme-machine » par J.-Y. Gresser.

L'ECHO DES RECHERCHES (C.N.E.T.), OCTOBRE 1973.

« Recherches actuelles sur l'extraction de caractéristiques et la reconnaissance de la voix parlée » par M.-J. Vigneron, M. Lamotte, J.-P. Haton, J. Bremont

AUTOMATISME, n° 12, déc. 1970.

« Quelques idées directrices en reconnaissance automatique de la parole »

par J.-S. Lienard, M. Castellingo, E. Leipp, M. Mlouka, G. Renard, J. Sapaly et D. Teil.
AUTOMATISME, n° 3, mars 1973.

DEUX GADGETS INSOLITES
DECRIIS DANS LE HAUT PARLEUR
SPECIAL GADGET

le **CASSE-TETE** et
la
BOUGIE ELECTRONIQUE

chaque modèle
* complet en kit 30,00 F
* en ordre de marche 40,00 F

R. PETTINI - JUJURIEUX 01450 PONCIN



devenez un VRAI CADRE

Le CIFRA met à votre portée trois préparations aux fonctions de cadres inédites et incomparables, adaptées aux principaux niveaux de responsabilités. Ces préparations (par correspondance) vous feront découvrir : l'état d'esprit, les facultés psychologiques, le sens de la réussite, les techniques, les principes, les outils, les objectifs à définir, les méthodes, les moyens; bref, tout le potentiel humain nécessaire pour accéder avec succès aux fonctions de cadre ou de direction.

Le temps de l'expérience personnelle est révolu; il faut profiter de suite de l'expérience des autres, sans quoi vous serez dépassé et écarté définitivement de la "compétition".

Le CIFRA a sélectionné parmi toutes les techniques de commandement et de gestion celles qui ont le mieux prouvé leur efficacité. Notre méthode de formation tient toujours compte de votre objectif et est bien adaptée aux souhaits des personnes engagées dans la vie professionnelle. Ces préparations vous permettront d'acquérir rapidement les connaissances et des moyens pratiques directement exploitables pour assurer votre promotion.

VOICI QUELQUES SUJETS TRAITES PAR NOS PREPARATIONS AUX FONCTIONS DE :

DIRECTION

Le management - La stratégie des affaires - La gestion prévisionnelle et contrôlée - L'informatique - Marketing et stratégie commerciale - Les prévisions à terme - Psychologie de la décision - La prospective - Les techniques de créativité - La communication - Conduite active des entretiens et réunions, etc...

CADRE

La gestion efficace du personnel - Logique et méthodologie - Organisation générale de l'entreprise - Le prix de revient - Marché Commun - Droit social - L'économie politique moderne - Commandement et autorité - Psychologie appliquée - Statistiques - Informatique - Stimulation des hommes - Commercialisation, publicité, etc...

AGENT DE MAITRISE

Organisation générale de la production - Les plannings - Relations humaines et psychologie du travail - Le prix de revient - Simplification et rationalisation des tâches - Les postes de travail - Rôle de l'agent de maîtrise - Facultés nécessaires pour diriger - Amélioration de la qualité - Législation du travail, etc...

LA PREPARATION AUX FONCTIONS DE CADRE EST UNE AFFAIRE DE SPECIALISTES :
Le CIFRA est un organisme privé, soumis au contrôle pédagogique de l'Etat, spécialisé dans la préparation aux fonctions de cadre et de direction. Former des hommes d'action volontaires et constructifs, c'est notre métier. Aussi, notre enseignement par correspondance moderne (avec compléments sur cassettes, études de cas, séminaires facultatifs) a-t-il été spécialement conçu pour mettre à votre portée la formation exacte qui fera de vous un vrai cadre.

Vous avez peut-être, vous aussi, tout ce qu'il faut pour réussir. Ne gaspillez pas vos chances ! Demandez de suite au CIFRA de vous expédier par retour, gratuitement et sans aucun engagement, la documentation qui vous intéresse.



Notre brochure contient les renseignements sur la gratuité possible de nos préparations (Loi sur la Formation Continue du 16.7.71)

BON POUR RECEVOIR GRATUITEMENT
la documentation complète sur la préparation aux Fonctions de DIRECTION
la documentation complète sur la préparation aux Fonctions de CADRE
la documentation complète sur la préparation aux Fonctions d'AGENT de MAITRISE

et sans aucun engagement de ma part. Préparation aux Fonctions de DIRECTION
 Préparation aux Fonctions de CADRE
 Préparation aux Fonctions d'AGENT de MAITRISE

(Faites une X) Préparation aux Fonctions de DIRECTION
 Préparation aux Fonctions de CADRE
 Préparation aux Fonctions d'AGENT de MAITRISE

NOM Prénom.....
ADRESSE
A renvoyer au **CIFRA**
71, Rue Saint-Lazare
75009 PARIS
Tél. : 874-91-68