RECONNAISSANCE DES FORMES ACOUSTIQUES

par

Gilbert FERRIEU

Ingénieur des télécommunications *

Comment construire une machine qui soit capable d'écouter et de comprendre le langage humain, quels sont les principes généraux de reconnaissance des formes que nous avons pu mettre en application, c'est ce que nous allons nous efforcer de préciser maintenant.

— Le schéma général d'une chaîne de reconnaissance des formes peut être réduit au bloc diagramme de la figure 1.

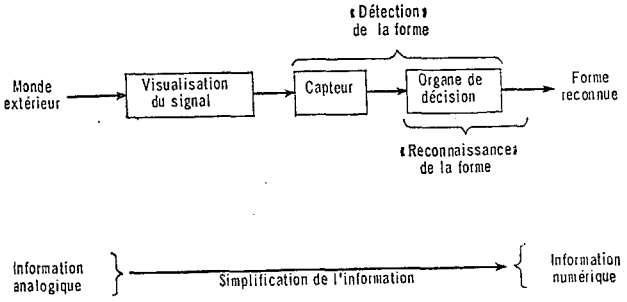

FIG. 1. — Schéma d'une chaîne de reconnaissance de formes.

Le monde extérieur, l'environnement de l'homme est *visualisé* sous forme d'un « signal ». Ce signal, complexe est analysé par un « *capteur* » qui le transforme en « *forme* ». Ensuite un *organe de décision* est chargé de reconnaître cette forme parmi un ensemble de formes possibles.

Faisons tout de suite plusieurs remarques.

— L'information qui transite dans la chaîne de reconnaissance est au départ « analogique » pour acquérir à la sortie de l'organe de décision un caractère quantifié, donc « numérique ».

— L'ensemble du capteur et de l'organe de décision effectue une opération de détection de la forme, l'organe de décision effectue l'opération proprement dite de reconnaissance de la forme.

Dans notre problème, le monde extérieur est *l'environnement acoustique*, l'organe de visualisation du signal est le *microphone*, le capteur est un *analyseur de spectre* (le vocoder), l'organe de décision un *calculateur*, la forme reconnue le son, la syllabe, ou le mot.

Avant de préciser un peu plus le rôle et l'organisation de chacun des carrés du bloc diagramme précédent, nous voudrions tout de suite signaler qu'il nous a paru judicieux d'effectuer la numérisation de l'information le plus tôt possible, c'est-à-dire au niveau du capteur : principalement parce que notre organe de décision est un calculateur numérique qui aime particulièrement l'information sous forme binaire !

— Nous ne nous étendrons pas sur l'organe de visualisation du signal. C'est dans notre cas un simple microphone, qui fournit donc un signal électrique traduisant des variations de pression acoustique.

Le rôle du capteur est de transformer ce signal électrique en une forme numérique qui soit représentative de ce que l'on veut reconnaître, à savoir la parole et plus principalement les différents sons ou phonèmes constitutifs du langage parlé. Il est bien connu que le spectre instantané de fréquences du signal de parole contient, sous une forme beaucoup moins redondante que le signal lui-même, l'essentiel de l'information véhiculée par ce signal. L'oscillogramme d'un « A » n'a rien de bien caractéristique, par contre l'évolution au cours du temps du spectre de fréquence de ce « A » est assez caractéristique. Notre capteur sera donc un analyseur de spectre fournissant un certain nombre d'informations quantifiées. Ce capteur, ce « vocoder » puisque tel est son nom, fournit en réalité 50 fois par seconde, un ensemble de 12 grandeurs codées numériquement sur une échelle à 16 niveaux. Ces 12 nombres caractérisent de façon suffisante le spectre de fréquence du signal de parole, ils contiennent bien l'essentiel de l'information linguistique véhiculée par la parole, puisque en effectuant la synthèse d'un son ayant un spectre défini par ces 12 paramètres, on obtient une véritable parole synthétique parfaitement intelligible.

A la lumière de ces trop brèves précisions sur les vocoders, on comprendra mieux le rôle du capteur

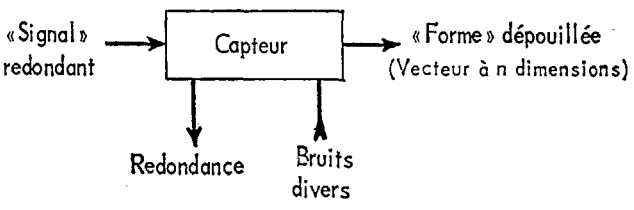

FIG. 2. — Rôle du capteur.

* Au CNET. Lannion. Département Études et techniques d'acoustique.

dans notre chaîne de reconnaissance, rôle qui est illustré à la figure 2.

Le capteur reçoit un signal analogique redondant, il élimine le maximum de redondance, il fournit une forme qui est un « vecteur » dans un espace à n dimensions. Ce vecteur à n dimensions doit conserver l'information *utile* portée par le signal. Le capteur est ainsi un véritable *organe périphérique* du calculateur constituant l'organe de décision. On remarque enfin que le capteur superposé à la forme des bruits qui n'existaient pas dans le signal (bruit de codage, bruit dû à une mauvaise adaptation du capteur aux formes à reconnaître).

— Quel va être maintenant le rôle de l'organe de décision ?

Une forme, pour lui, est un vecteur à n dimensions. Pour effectuer une reconnaissance du vecteur $\vec{x}$ inconnu, il est nécessaire d'avoir défini et catalogué l'ensemble des vecteurs $\vec{E}$ qui sont les formes types à reconnaître.

On arrive ainsi au schéma d'organisation de la figure 3.

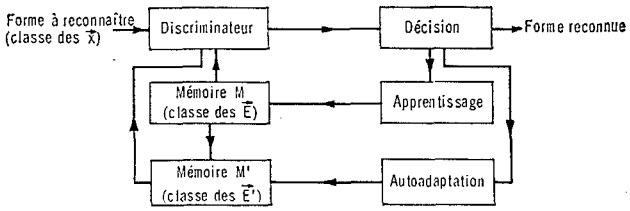

Fig. 3. — Rôle de l'organe de décision.

— Le discriminateur effectue l'application de l'ensemble des $\vec{x}$ sur l'ensemble des $\vec{E}$, mis en mémoire dans la mémoire permanente. Cette application constitue la « fonction » de reconnaissance (cette fonction pourra être un produit scalaire, une distance...). L'organe de décision se borne à prendre une décision logique sur les résultats de la fonction de reconnaissance. Le rôle de l'organe, l'apprentissage est de calculer les coefficients corrects de la fonction de reconnaissance.

Cette opération d'apprentissage est faite une fois pour toutes, *avant* la phase de reconnaissance : on « apprend » alors au discriminateur à reconnaître de façon optimale des vecteurs $\vec{x}$ *connus*. Parallèlement, existe un organe d'autoadaptation *en temps réel*, établissant comme l'organe d'apprentissage une boucle de réaction sur le discriminateur. Son rôle est de corriger légèrement la classe des $\vec{E}$, donnant une classe de vecteurs types $\vec{E'}$, plus proches des formes à reconnaître, à cet instant, que les formes types $\vec{E}$.

Ces nouvelles formes types temporaires sont classées dans la mémoire M'.

Il n'est évidemment pas possible de rentrer dans le détail des opérations qui sont effectuées dans les organes que nous venons de décrire sommairement. Nous voudrions surtout, avant de conclure ce bref exposé, insister sur les particularités de la reconnaissance automatique de la parole, particularités qui en fait constituent principalement des difficultés !

— Tout d'abord les formes acoustiques que nous avons à reconnaître ne sont pas uniques et bien définies. Il n'existe pas de véritable spectre type de tel ou tel phonème. Le spectre d'un même phonème variera d'un individu à l'autre et même pour un individu, d'un moment à l'autre. Bien plus, des phonèmes différents, et effectivement perçus comme différents par tout auditeur peuvent avoir des spectres très semblables. Il est donc nécessaire que l'organe de reconnaissance de formes simule en fait le fonctionnement des organes d'audition humaine, il faut qu'il complète l'information purement acoustique par une mémoire du *vocabulaire*, de la *linguistique*, de la *grammaire*. Le cerveau humain reconnaît difficilement des sons isolés, comme lui l'organe de reconnaissance des formes doit s'efforcer de reconnaître les sons à l'intérieur d'un *contexte sonore*.

— D'autre part, les machines de reconnaissance automatique de parole doivent fonctionner en temps réel, sous peine de perdre une très grande partie de leur intérêt. Il n'est donc pas possible de mettre en jeu des processus de reconnaissance trop compliqués qui dépasseraient les capacités de travail en temps réel du calculateur.

— Enfin, on ne connaît pas encore suffisamment le mécanisme de perception auditive pour arriver à simuler sur le calculateur la faculté du cerveau humain de reconnaître un son ou une voix déterminée dans une ambiance sonore complexe comme on en rencontre dans une cocktail party !

Il reste donc beaucoup de chemin à parcourir avant d'avoir mis au point une véritable machine capable de percevoir la parole comme n'importe quel être humain, mais l'intérêt évident d'une telle machine, les applications immédiates qu'elles permettraient de développer amènent un nombre croissant de chercheurs à s'intéresser à l'étude de ce problème, dont la solution est peut-être pour demain.