

La reconnaissance de la parole au Cnet, entre 1965 et 1980

Michel Cartier, Christian Gagnoulet, Guy Mercier*
Patrice Quinton†

20 juillet 1995

Résumé

Les pionniers de la reconnaissance de la parole (Myasnikov en 1943 ; Dreyfus-Graf en 1952 ; Davis en 1952) utilisaient des techniques analogiques. Les recherches en France ont débuté en 1965, en s'appuyant sur des moyens informatiques. A partir de 1970, des programmes de recherche plus ambitieux ont permis la mise en place de véritables équipes de chercheurs et d'ingénieurs. Les premières réalisations industrielles se situent aux alentours de 1980.

En France, le Cnet à Lannion a été l'un des acteurs principaux de cette évolution. A travers son histoire, et à partir de l'expérience des auteurs, cet article évoque les recherches en reconnaissance de la parole entre 1965 et 1980, période qui va des premières expériences à l'émergence des premiers systèmes commercialisés.

1 Introduction

Les pionniers de la reconnaissance de la parole (Myasnikov en 1943 [52]; Dreyfus-Graf en 1952 [24]; Davis en 1952 [20]) utilisaient des techniques analogiques. Les recherches en France ont débuté en 1965, en s'appuyant sur des moyens informatiques. À partir de 1970, des programmes de recherche plus ambitieux ont permis la mise en place de véritables équipes de chercheurs et d'ingénieurs. Les premières réalisations industrielles se situent aux alentours de 1980.

En France, le Cnet à Lannion a été l'un des acteurs principaux de cette évolution. À travers son histoire, l'article que nous proposons a pour but d'évoquer les recherches de cette période, auxquelles ont participé les auteurs, à partir de documents d'archives (articles scientifiques et techniques, comptes-rendus de débats, interviews, brochures, etc.).

L'article comporte trois parties. Dans la première partie, nous présentons quatre maquettes réalisées au Cnet entre 1965 et 1977 : le phonétographe, Charles, Dynamo et Keal; nous exposons à travers leur description les principes des méthodes de reconnaissance utilisées à cette époque. Le phonétographe et Charles sont des dispositifs d'inspiration analogique, réalisés entre 1965 et 1972. Dynamo est une machine à reconnaître les mots isolés, fondée sur l'utilisation de la programmation dynamique. Keal est un système de reconnaissance de la parole continue, combinant l'utilisation de sources de connaissances phonétiques, lexicales, prosodiques, syntaxiques et pragmatiques, mis au point vers la fin des années 1970.

La seconde partie est consacrée au contexte national et international. Nous décrivons brièvement les travaux des principaux laboratoires français et étrangers menant des recherches sur la parole. Nous insistons sur les modes d'échanges entre ces laboratoires, en montrant le rôle des institutions et des sociétés savantes (Galf, Afcet, IEEE, Asa).

Dans la troisième partie, nous revenons sur les facteurs qui, selon nous, ont marqué ces recherches de façon profonde :

- les connaissances linguistiques et le développement des recherches fondamentales sur la production, la perception et la compréhension de la parole;

*. Cnet Lannion, Route de Trégastel, 22301, Lannion.

†. Irisa-CNRS, Campus de Beaulieu, 35042 Rennes Cedex

- les modèles mathématiques qui ont permis de mettre en œuvre ces connaissances pour reconnaître la parole (programmation dynamique, réseaux de Markov, grammaires, réseaux sémantiques, etc.);
- la technologie des circuits et des ordinateurs, dont les progrès très rapides ont décuplé les possibilités de recherche et de réalisation de maquettes.

Nous montrons comment les rapports entre ces différents facteurs se sont établis, à partir du débat scientifique et technique tel qu’il apparaît dans les publications.

En conclusion, nous analysons l’influence de cette période sur les recherches ultérieures.

2 Quatre réalisations

Entre 1965 et 1980, le Cnet Lannion a joué un rôle important dans les recherches en reconnaissance de la parole. Celles-ci se sont concrétisées par la réalisation de plusieurs maquettes assez représentatives des principales techniques utilisées à cette époque¹.

2.1 Le Phonétographe et Charles

Les recherches sur la reconnaissance de la parole au Cnet-Lannion ont une double filiation : informatique et acoustique [9]. Ainsi que l’indique un rapport interne de 1967 [62]:

« Le but immédiat était de prendre contact avec les techniques de la reconnaissance des formes et de les appliquer dans le domaine acoustique en utilisant un vocodeur pour capter les informations relatives à la parole et un calculateur électronique universel pour traiter les informations. »

Bien que les deux équipes concernées (celle du centre de calcul, CEI, et celle du département d’acoustique, ETA) aient poursuivi leur coopération dans l’utilisation des vocodeurs, leurs programmes de travail respectifs ont été conduits de façon relativement indépendante. Ainsi, à CEI, dans la continuité de la citation [62], tous les efforts étaient consacrés au traitement des données dans le calculateur (purements statistiques au départ, dans l’optique reconnaissance des formes, puis lexicaux, syntaxiques et sémantiques dans l’optique intelligence artificielle). À ETA, où l’inspiration était analogique, on considérait que le matériel contenait la clé des problèmes. Est-ce la cause ou la conséquence du recrutement dans ce département, fin 1966, de Jean Dreyfus-Graf? Cet inventeur était connu pour ses travaux sur la représentation des sons [23] et la détection analogique de formants [26], mais surtout pour la réalisation d’une machine à écrire à commande vocale [24, 25].

Le Phonétographe fut installé à Lannion et raccordé à un calculateur PB-250 récupéré des études de commutation électronique. Constitué de compresseurs d’amplitude sélectifs et de filtres, le Phonétographe réalisait un prétraitement très élaboré du signal et fournissait toutes les 10 millisecondes un jeu de 15 paramètres binaires représentant 11 canaux d’analyse spectrale, 2 canaux d’énergie globale, le voisement et la détection de plosives. Un algorithme simple programmé dans le calculateur transformait les données relatives à un mot en une séquence de segments dont chacun représentait un son détecté (plus précisément, une classe de sons parmi 10). Une expérience fut menée sur la reconnaissance de 18 mots (“zéro” à “dix”, plus, moins...) prononcés par 14 locuteurs masculins. Le pourcentage de reconnaissance pouvait atteindre 97% pour des locuteurs « suffisamment coopératifs » [16].

Cette expérience montrait les limites de l’approche analogique pour des locuteurs tout-venants. Cependant l’idée de commander des machines à la voix avec des moyens informatiques sommaires conduisit J. Dreyfus-Graf à imaginer des « phonocodes », commandes vocales artificielles résultant de l’association de sons peu nombreux et faciles à séparer indépendamment du locuteur. A partir de Sotina [27] – nom donné à la première

1. Pour une introduction aux techniques de traitement de la parole voir [14].

version de ce code, un appareil fut réalisé à Lannion avec le soutien du Comité de Recherche en Informatique (contrat CRI 73.034) :

« L'intérêt que nous trouvons à ce projet est d'avoir valeur d'exemple dans la réalisation d'applications rustiques qui pourraient servir par exemple à la commande de machines. »²

L'étude se prolongea au-delà des délais initiaux (compte rendu final fourni à l'Iria le 30 novembre 1976) et aboutit à la réalisation de la machine « Charles » (Capteur Hybride Adapté à la Reconnaissance d'un Langage d'Éléments Spéciaux). Cette machine était une sorte de phonétographe où le PB-250 avait laissé la place à l'un des premiers microprocesseurs (un IMP-16P acquis en juillet 1975 comportant une mémoire de 4 K-mots de 16 bits). Avec un « vocabulaire » de 36 mots artificiels de structure consonne-voyelle-consonne-voyelle et 15 locuteurs (hommes et femmes) on obtenait un taux de reconnaissance de 89% et un taux d'erreurs de 2% (9% de rejets) [19].

2.2 Dynamo

En 1972, à la suite d'un voyage de J.Y. Gresser, M. Cartier et R. Carré en Union soviétique [33], J.F. Bars [3] développe le premier programme basé sur une méthode dite de « comparaison dynamique » entre le mot à reconnaître et les références acoustiques. Ce système monolocuteur peut détecter ou reconnaître les mots dans la parole continue. Le taux de reconnaissance pour un vocabulaire de 21 mots est de 84%. En février 1974, ce système est utilisé pour la reconnaissance de mots isolés monolocuteur [57]. Le taux de reconnaissance pour un vocabulaire de 100 mots est de 96%. Ce taux descend à 73% pour un vocabulaire de 1000 mots (83% si on prend les deux meilleures réponses).

La première maquette Dynamo (première apparition du nom) est conçue entre 1974 et 1975. Elle utilise un vocodeur à canaux analogique, un minicalculateur CII-10020, puis plus tard un Mitra 115. L'entrée est soit un microphone, soit un poste téléphonique. Les dictionnaires utilisés sont au choix: oui/non en multilocuteurs, les chiffres, les villes plus les nombres (en vue de la mécanisation du tri postal), un mini-répertoire téléphonique ou encore le dictionnaire artificiel Chokina. La sortie se fait sur le terminal ou en synthèse vocale. En 1975 est réalisée une première expérience à travers le réseau téléphonique entre le Cnet/Issy-les-Moulineaux et le Cnet Lannion, lors du congrès européen Fase [12]. Un module de dialogue complète le système. De 1978 à 1980, une première maquette d'un autocommutateur commandé à la voix et une maquette de composeur vocal sont réalisées. Le composeur téléphonique à commandes vocales est une application de Dynamo: l'abonné appelle un correspondant par son nom au lieu de composer son numéro.

« Le principal but de cette maquette est d'établir des communications ordinaires entre abonnés. [...] L'utilisateur choisit le service qu'il désire en prononçant un mot clé lorsque la machine le lui demande (communication, absence, présence, transfert, rappel automatique ...) »³.

Cette maquette aboutira au premier transfert industriel du Cnet dans le domaine de la reconnaissance de la parole.

2.3 Keal

Keal (du breton *Kenreizadur Emfreeg evit Anavezout Lavar*) est un ensemble de programmes informatiques destiné au dialogue vocal [50]. La décision d'explorer la voie intelligence artificielle de la reconnaissance de la parole est prise en 1973 au département CEI, au moment où démarre le programme *Speech Understanding and Recognition* de l'agence Arpa (*Advanced Research Projects Agency*) aux États-Unis. L'application visée est le renseignement téléphonique. L'idée est de compléter les programmes de reconnaissance acoustico-phonétique existants avec des programmes d'analyse syntaxique et sémantique, afin de permettre à un ordinateur de « simuler » le comportement d'une opératrice de renseignement. Des « expériences » de simulation de dialogue [59] montrent que les usagers utilisent des formulations très standard, ce qui laisse espérer que

2. Lettre adressée le 1er août 1972 par le Secrétaire du CRI au Directeur du Cnet de Lannion.

3. "Recherches Acoustiques", rapport annuel sur les recherches en acoustique au Cnet, Volume V, 1978.

la tâche de reconnaissance n'est pas insurmontable. Keal utilise une technique dite « ascendante », allant du signal de parole à la représentation sémantique : vocodeur à canaux, reconnaissance des « phonèmes » par des techniques de discrimination statistiques [49], assemblage des caractères (analyse lexicale [76]), reconnaissance des phrases (analyse syntaxique [58]), puis intégration des informations reçues pour préparer la réponse du système par synthèse vocale). L'ensemble du dispositif comporte donc un vocodeur piloté par un microprocesseur, relié à un ordinateur CII-10070 (comportant 30K mots de mémoire) sur lequel sont mis en œuvre les logiciels programmés en Fortran. La reconnaissance d'une phrase prend environ une minute. Keal est essayé en 1978 sur deux applications. Dans une application d'annuaire vocal (37 personnes du département CEI), on arrive à satisfaire 98% des demandes après épellation éventuelle. Sur une autre application destinée à l'entrée vocale de circuits imprimés, Keal atteint 82% de reconnaissance des phrases.

3 Le contexte national et international

Les recherches du Cnet sont loin d'être isolées : la reconnaissance de la parole est en effet un sujet de recherche auquel de nombreux pays s'intéressent. Nous résumons dans la présente section l'activité du Japon, de l'Union soviétique, des Etats-Unis, et de l'Europe.

3.1 Le Japon

Dès 1950, le Japon a été présent dans les recherches sur la reconnaissance de la parole [78]. Des systèmes expérimentaux de reconnaissance des phonèmes et des voyelles sont réalisés dans les universités de Kyoto et de Sendai, au début des années 60 [63, 67, 68]. En 1964, Nec présente un système de reconnaissance des chiffres isolés [43] atteignant 93 à 98% de reconnaissance. Stakushi et al. [66] étudie la reconnaissance de mots à partir de chaînes de phonèmes. Sakoe et Chiba [64] sont parmi les premiers à utiliser la programmation dynamique. Sur ces bases, les années 70 voient plusieurs réalisations de systèmes de reconnaissance de la parole continue. Le système de l'université Tohoku à Sendai [68] reconnaît des noms de villes. Le système Lithan de l'université de Kyoto [53] est un système de reconnaissance et de compréhension de la parole. Le système de l'université Yamanashi [65] vise la dictée de programmes Fortran. NTT conçoit un système de réservation de places de trains [45]. À Nec, Chiba [17] reconnaît la parole par programmation dynamique directe sans segmentation phonétique. Il parvient à reconnaître 99% des suites de chiffres connectés (1 à 4 chiffres) à 99,7% pour 5 locuteurs. En 1978 est commercialisée par Nec la première machine de reconnaissance de mots isolés et de suite de chiffres connectés. En 1979, Fujitsu et Hitachi annoncent leurs premiers systèmes de reconnaissance et espèrent s'implanter sur le marché.

3.2 Union Soviétique.

L'Union Soviétique possède une tradition de recherche en parole très ancienne. Dans sa monographie [77], source d'information exceptionnelle sur l'histoire de la recherche en parole en URSS, Vysotsky mentionne comme l'une des principales raisons de cette tradition « la prise de conscience très tôt par l'État du rôle vital que [la reconnaissance de la parole] pourrait tenir dans le domaine de la sécurité », et mentionne explicitement le célèbre roman d'Alexandre Soljenitsyne, *Le premier cercle*.

Dès 1943 paraît la première publication de L. Myasnikov sur un système de reconnaissance de différents sons (75 à 80% de reconnaissance des voyelles) [52]. De 1955 à 1958, G. Tsemel à l'Institut des problèmes de transmission à Moscou réalise un système capable de reconnaître les voyelles sans apprentissage avec affichage sur un tableau électronique [69]. Plus tard en 1965, la collaboration des équipes de Tsemel et de Trunin-Donskoj à Moscou permet de réaliser un système de reconnaissance analytique indépendant du locuteur atteignant 93% de reconnaissance. En 1966, le premier séminaire national annuel sur la reconnaissance de la parole (Arso) est organisé par N. Zagoruiko, à Akademgorodok, Novosibirsk. C'est la première fois que les chercheurs soviétiques en parole se rencontrent. En 1968, T. Vintsjuk à Kiev développe le premier système de reconnaissance à base de programmation dynamique (100% de reconnaissance en monolocuteur sur 12 mots [72]). Plus tard (1976 [73, 74]), ces recherches déboucheront sur un programme de reconnaissance

d'un vocabulaire de 200 mots, monolocuteur (95% de reconnaissance), basé sur des méthodes statistiques markoviennes. Ce programme nécessite 20 à 30 fois le temps réel sur un ordinateur BESM 6. En 1969, N. Zagoruiko et V. Velichko (Novosibirsk) obtiennent 95% de reconnaissance en monolocuteur (vocabulaire de 203 mots, programmation dynamique). Ils obtiennent plus tard de très bons résultats en reconnaissance de grands vocabulaires (200 à 2000 mots) et en parole continue en utilisant la programmation dynamique et des modèles de perception. 1975 voit le développement du premier système de reconnaissance opérationnel (50 mots, tous locuteurs) et la première application temps réel de cette technique au contrôle de la distribution de l'énergie à Moscou.

3.3 Etats-Unis

Davis et ses collègues sont parmi les premiers à avoir développé aux Etats-Unis des recherches sur la reconnaissance de la parole [20]. Leur appareil est le premier système capable de reconnaître les chiffres prononcés isolément par un locuteur pour lequel a été effectué un apprentissage préalable. Vicens en 1969 [71] est le premier à reconnaître un mot dans un grand vocabulaire (561 mots). Il est d'autre part le premier à utiliser les contraintes linguistiques (syntaxe) pour améliorer la reconnaissance dans le cadre de deux applications : commande vocale d'un robot en parole continue (16 mots) et commande vocale d'un ordinateur de bureau. À partir de novembre 1971, les choses s'accroissent. Sous l'influence du rapport Newell [54], l'agence Arpa lance un projet ambitieux, sur 4 ans. Il s'agit de réaliser un système capable de « comprendre » des phrases formées d'un vocabulaire de 1000 mots, sans contrainte d'élocution pour l'utilisateur, et multilocuteur. Le taux d'erreur ne doit pas dépasser 10 %. Après une première sélection de deux ans, trois projets sont retenus. HWIM (*Hear What I Mean*) de la firme Bolt Beranek and Newman [10] atteint 45 à 50 % de reconnaissance sur un langage de 1000 mots relativement complexe. À l'université Carnegie Mellon de Pittsburg, Hearsay [60, 28] parvient à 90 % de reconnaissance. Ces deux systèmes ont des caractéristiques similaires : ils font coopérer un grand nombre de « sources de connaissance » suivant des stratégies fort complexes issues de l'intelligence artificielle, afin de reconnaître la parole. Une description très détaillée du projet Arpa et de ses résultats a été présentée par D. Klatt du MIT [44].

Paradoxalement, la réalisation la plus marquante est le système Harpy, réalisé par Lowerre à l'université Carnegie Mellon de Pittsburgh, en marge du projet Arpa [47]. Sur la même application que Hearsay, il atteint 95 % de reconnaissance en environ 80 fois le temps réel, mais il est beaucoup plus simple que ce dernier.

IBM a été un pionnier de la reconnaissance de la parole. Sous la ferme direction de Jelinek, une équipe très importante a été réunie dès 1969 et a suivi une voie originale. Considérant la reconnaissance comme un problème de décodage de canal de transmission, Jelinek se fait l'avocat de techniques « aveugles » inspirées d'une variante de l'algorithme de Viterbi [39], qui se combineront plus tard au modèle dit « markovien caché » [40]. En 1978, cette approche permet d'atteindre 95 % de reconnaissance de phrase sur un vocabulaire de 250 mots de complexité moyenne [41]. Baker [1, 2] est semble-t-il le premier à avoir utilisé l'approche markovienne pour reconnaître la parole. AT&T, dans ses fameux « *Bell Labs* », s'est intéressé essentiellement à la reconnaissance des mots isolés, et ce n'est qu'à partir de 1980 que les applications de la reconnaissance de mots connectés y ont été étudiées. Le MIT, en collaboration avec le laboratoire Lincoln [29], a aussi été actif dans la reconnaissance de la parole.

3.4 Europe

En Europe, la reconnaissance de la parole a essentiellement été explorée en Allemagne de l'Ouest, en Grande Bretagne [11], en Italie [51], et bien sûr en France.

La première utilisation d'informations linguistiques en reconnaissance est due à P. Denes [21] à Londres : les fréquences des suites de deux phonèmes (diphones) sont utilisées pour améliorer la reconnaissance des phonèmes.

L'une des réalisations les plus originales est celle de Tubach, qui au Centre d'Études sur la Traduction Automatique de Grenoble, en 1970, avec l'aide du centre de recherche IBM de La Gaude, construit un programme de reconnaissance pour le langage de programmation Algol, et utilise ainsi pour la première fois

des contraintes syntaxico-sémantiques[70]. À Grenoble toujours, C. Bellissant à l'université de Grenoble I [8] et Groc et Tuffeli [34] au Laboratoire de la communication parlée de l'Enserg poursuivent des travaux sur la reconnaissance. A l'université de Nancy, Jean-Paul Haton et son équipe constituent un des pôles de recherche les plus importants sur ce sujet. Dans les années 1970, ces recherches aboutissent aux systèmes de reconnaissance de la parole continue Myrtille I et II [56, 37]. Le Limsi à Orsay est l'un des grands laboratoires de recherche sur la parole. Dès 1972, Liénard décrit dans sa thèse des méthodes de reconnaissance de la parole [46]. Dans cette lignée, Mariani [48] réalise plus tard le système de reconnaissance Esope. Citons encore à Toulouse les recherches sur le projet Arial [22], ainsi que la contribution très originale de Caelen [13] qui est un des rares chercheurs à avoir exploité un modèle d'audition pour la reconnaissance de la parole. À Marseille, Battani et Méloni utilisent la première version de l'interpréteur Prolog pour réaliser un système de reconnaissance de la parole [5]. Baudry et Dupeyrat au CEA de Saclay [6] poursuivent aussi des recherches en reconnaissance.

3.5 Les modes d'échange

Jusqu'au début des années 1970, les échanges se développent à l'occasion de conférences ou de colloques, mais au moins autant via des voyages d'étude organisés. C'est ainsi qu'en 1975, un second voyage en URSS de chercheurs français permet d'approfondir les connaissances des travaux, très riches mais peu connus, de leurs collègues soviétiques [36]. De même, en 1972, J.Y. Gresser visite les principaux laboratoires américains à l'occasion de la conférence internationale *Speech Communication and Processing* à Boston.

L'importance croissante du sujet amène rapidement les chercheurs à organiser ces échanges par l'organisation de groupes à l'intérieur des sociétés savantes.

En France, le Galf⁴ [31] et l'Afcet⁵ permettent aux chercheurs d'organiser des rencontres régulières, extrêmement fructueuses⁶. Citons René Carré⁷ :

« C'est au cours d'une rencontre organisée à Grenoble les 19 et 20 février 1970 et consacrée à quelques thèmes particuliers sur la parole que fut créé, sur proposition de Monsieur Mattei⁸, un groupe spécialisé Communication parlée au sein du Galf. La création de ce groupe se justifiait par le développement des recherches sur la parole, tant fondamentales qu'appliquées, effectuées en France, développement correspondant à un accroissement du nombre de personnes et d'équipes concernées par ces recherches [...] Cette division correspond à celle adoptée dans le Jasa par la section 9 *Speech Communication*.

« [...] Au cours des journées de Grenoble] les thèmes suivants ont été abordés :

- « – les problèmes de perception de la parole,
- « – les synthétiseurs et leurs applications,
- « – la synthèse par règles,
- « – la reconnaissance de la parole.

Pour l'étude de ce dernier thème nous avons la participation du comité spécialisé *Intelligence artificielle et reconnaissance des formes* de l'Afcet.

[...] Nous avons participé à l'Afcet à l'organisation d'une conférence sur le thème « communications homme-machine sur support vocal » par P. Vicens, le 28 mai 1970. »

4. Groupement des Acousticiens de Langue Française.

5. Association Française pour la Cybernétique Économique et Technique, aujourd'hui appelée Association française des sciences et technologies de l'information et des systèmes.

6. On se reportera à l'article de M. Grossetti [35] pour plus d'information sur l'histoire de la communauté « parole » en France.

7. Extraits du premier rapport d'activité (14 janvier 1971) du Groupe Communication Parlée, rédigé par son président René Carré (Enserg).

8. Président du Galf (NDLR).

Les Journées d'études sur la parole se poursuivent à un rythme d'abord annuel, puis bisannuel : la 20ème édition s'est tenue en 1994 à Lannion. L'intensité des échanges dans la communauté francophone – 25 équipes et 178 chercheurs recensés dans une brochure éditée par le Galf en 1977 – apparaît dans les actes de ces journées. Par exemple, le volume II des 5èmes journées d'Orsay (1974), qui rassemble les textes des quatre exposés invités, les comptes rendus des deux tables rondes et les discussions des communications, consacre 36 pages à la partie discussion. Notons qu'une session porte sur l'application des contraintes linguistiques à la reconnaissance automatique de la parole. En 1976, à Nancy, la session « reconnaissance de la parole continue » comporte 10 exposés dont 3 du Cnet (P. Quinton, J. Vaissière et R. Vivès). D.-R. Reddy, professeur à CMU, présente un exposé d'introduction sur ce thème; J.-Y. Gresser préside la session après une introduction où il oppose les préoccupations des chercheurs, « plus fondamentales », à « l'utilitarisme des travaux », et où il insiste sur le dialogue homme-machine :

« La notion de dialogue homme-machine est pratiquement absente des travaux étrangers, alors qu'elle est au centre des recherches avancées menées en France ... »([42, p. 18]).

Les contraintes linguistiques et la pragmatique sont au centre des débats.

Au niveau international, les échanges se font essentiellement via les sociétés savantes américaines, en l'occurrence l'Asa (*Acoustical Society of America*) et l'IEEE. Les réunions annuelles de l'Asa favorisent les échanges entre les linguistes et les ingénieurs. Elles sont très largement orientées vers la recherche fondamentale, et notamment la perception et les modèles de production de la parole. En 1976, la conférence internationale annuelle Icassp⁹ est créée et, désormais, constitue le lieu de rencontre international obligé. Soutenue par l'IEEE, elle donne naissance au chapitre ASSP de cette société savante. Beaucoup plus tournée vers l'ingénierie de la parole, la conférence Icassp est le reflet des progrès technologiques de la reconnaissance dans la période 1970 – 1980.

Le soutien des organismes publics est aussi déterminant pour le développement des recherches. On a vu l'importance du projet Arpa aux États-Unis. En France, les recherches sont soutenues dans les années 1970 par le Sessori (successeur du CRI), et une ATP (Action Thématique Programmée) du CNRS montre aussi l'intérêt que suscite la reconnaissance de la parole.

4 Les facteurs

Quels facteurs ont été déterminants pour la recherche sur la reconnaissance de la parole ? Ils sont selon nous à rechercher dans trois directions : les connaissances linguistiques disponibles, les modèles mathématiques, et enfin les progrès de la technologie des ordinateurs.

4.1 Connaissances linguistiques

Bien entendu, la reconnaissance de la parole s'est appuyée peu ou prou sur des connaissances acoustiques et linguistiques. Celles-ci ont du reste très largement progressé, à la fois parce qu'elles ont bénéficié de l'intérêt porté au sujet, et parce que l'instrumentation a elle aussi largement progressé. Les échanges entre linguistes et ingénieurs ont été activement recherchés pendant la période qui nous intéresse, et favorisés par le rôle de carrefour que jouaient les sociétés savantes. Parmi bien d'autres, M. Rossi, de l'Institut de phonétique d'Aix en Provence, a beaucoup œuvré pour l'utilisation des connaissances linguistiques dans les techniques de reconnaissance de la parole[61]. Le débat a souvent été vif entre les tenants d'une approche « linguistique pure », selon laquelle la reconnaissance de la parole devrait calquer la structure de production et de perception de la parole par l'être humain, et les « pragmatiques » dont la linguistique n'était pas la préoccupation majeure. On attribue à F. Jelinek, responsable de l'équipe de recherche d'IBM et inconditionnel des techniques de reconnaissances statistiques, cet avis définitif sur la question :

« *Every time we fire a phonetician or a linguist, the performance of our system goes up*¹⁰ ».

9. *International Conference on Acoustics, Speech and Signal Processing*

10. À chaque fois que nous licencions un phonéticien ou un linguiste, les performances de notre système augmentent.

Nombreux sont, à cette époque, les scientifiques qui doutent de l'intérêt, sur le plan fondamental, des recherches en reconnaissance de la parole. Citons Jean Gagnepain :

« ...le phonème ne se confond pas avec la réalité naturellement continue du son, n'en déplaie aux acousticiens qui, prenant pour argent comptant ses frontières et comme si le tracé permettait de distinguer, par exemple, l'éternuement de l'éjectivité, croient pouvoir le définir en formants;.... ... le linguiste n'a guère plus à tirer de la cybernétique que l'ornithologue des progrès de la balistique ou de l'aviation. »¹¹.

Ces débats sont loin d'être clos. Avec quelques années de recul, force est de constater, comme on le verra ci-dessous, que les « meilleurs » systèmes de reconnaissance sont fondés sur des modèles assez peu « linguistiques ». Toutefois, ces modèles reposent fondamentalement sur une structuration du signal de parole largement inspirée des connaissances fondamentales de l'acoustique et de la linguistique, qui se reflètent dans les techniques de traitement du signal, le choix des unités de reconnaissance, ou les modèles structurels.

4.2 Modèles mathématiques

L'émergence de modèles mathématiques et algorithmiques puissants est un autre facteur de progrès. L'utilisation de la *programmation dynamique* – une technique de programmation mathématique due à Bellman – pour la reconnaissance de la parole est due à Vintsjuk en 1968 [75]. Issue de la recherche opérationnelle, elle permet de comparer selon des critères de distance les séquences de paramètres acoustiques issus de l'analyse du signal de parole. La programmation dynamique permet de s'affranchir de déformations temporelles non linéaires qui apparaissent lors de la prononciation des mots, mêmes prononcés par une même personne. Très bien adaptée à la reconnaissance dépendante du locuteur en l'absence de perturbations acoustiques, cette technique est dispendieuse en puissance de calcul et surtout difficilement extensible à la reconnaissance indépendante du locuteur, et au traitement de la parole continue. Elle bénéficie cependant d'une procédure d'apprentissage extrêmement simple, limitée à l'enregistrement acoustique des mots de référence.

L'emploi de cette technique a été longtemps freiné par les exigences de puissance de calcul disponibles. Ainsi, au Cnet, le système Dynamo (voir paragraphe 2.2) a employé une forme heuristique de la programmation dynamique fondée sur des calculs de distances locales. La véritable programmation dynamique « temps réel » n'a été utilisée en France, grâce aux circuits VLSI et aux processeurs de traitement du signal, qu'à partir des années 80, avec les systèmes comme Mozart du Limsi [32] ou Séraphine du Cnet [30].

Une étape essentielle a été franchie en 1973, lorsqu'à l'Université Carnegie Mellon de Pittsburg, J.K. Baker [1] a l'idée d'employer une technique de modélisation statistique connue depuis plusieurs décennies, les modèles de Markov. Dans le domaine de la parole, cette technique est connue sous le nom de *modèle de Markov caché* (de l'anglais *hidden Markov model*). Elle s'appuie sur la programmation dynamique (algorithme de Viterbi) pour la phase de reconnaissance, mais possède l'avantage de permettre un *apprentissage* : les formes acoustiques ne sont pas mémorisées de façon passive, elles sont analysées statistiquement, ce qui permet de construire un modèle de ces données optimal pour la reconnaissance (algorithme de Baum et Welch [7]). Ce modèle possède des paramètres qui, par opposition au « non modèle » de la programmation dynamique, sont adaptés de manière optimale aux données fournies au cours de l'apprentissage.

Ainsi, dans un contexte indépendant du locuteur, on ne conservera plus, à l'issue de l'apprentissage, plusieurs dizaines de formes acoustiques brutes comme avec la programmation dynamique, mais un seul modèle statistiquement représentatif de l'ensemble des formes présentées à l'apprentissage.

Comme pour la programmation dynamique, des formes simplifiées de cet algorithme ont été utilisées pour l'apprentissage des modèles de Markov.

Très vite, le modèle markovien caché, par sa flexibilité, sa généralité et ses performances, a remplacé la programmation dynamique – ainsi que la plupart des autres techniques – pour devenir *la* technique de reconnaissance universellement employée.

Devant la difficulté de reconnaître les sons prononcés, l'idée d'utiliser la redondance inhérente aux langues naturelles pour « aider » la reconnaissance a été avancée au début des années 1970, sous l'influence des

11. JEP Lannion, 1-2 juin 1972, Jean Gagnepain

chercheurs en intelligence artificielle. Elle donna naissance aux systèmes dits de « compréhension de la parole », capables de saisir le sens des phrases proposées sans s'occuper à priori d'identifier de façon précise les sons. Deux grands types de modèles ont été utilisés à cette fin : les *grammaires formelles* et les *réseaux sémantiques*. Les grammaires formelles [18] visent à décrire les langues naturelles. Elles sont aussi à la base de la compilation dirigée par la syntaxe. Pour la reconnaissance de la parole, le principe était de définir de façon précise les phrases du langage – artificiel ou naturel – que le système de reconnaissance attendait, selon l'application choisie. Tubach à Grenoble en 1970 [70] a certainement été le premier à faire usage de ce principe. Plus tard, Bates, du Stanford Institute [4] reprend cette technique dans le cadre du projet Arpa. Dans le système Hearsay de l'Université Carnegie-Mellon de Pittsburgh, Hayes-Roth et Mostow [38] font aussi appel à la syntaxe et la sémantique. Le système Keal décrit dans le paragraphe 2.3 utilise aussi des connaissances syntaxiques [58].

Les réseaux sémantiques décrivent les relations de sens, de proximité ou d'opposition, entretenues par les mots prononcés, et leur utilité en parole est de prédire les mots attendus en fonction du contexte.

Qu'il s'agisse de syntaxe ou de sémantique, les deux modèles reposent sur des fondements linguistiques. En fait, le projet Arpa a bien montré que, si cette idée est sûrement intéressante, il est toutefois nécessaire que la reconnaissance des mots atteigne des performances élevées en pratique. De méchantes langues n'ont pas manqué de faire remarquer que les applications de certains systèmes de reconnaissance étaient tellement contraignantes que la machine devinait les réponses avant d'avoir à les analyser...¹²

4.3 La technologie

La reconnaissance a toujours été tributaire des puissances de traitement disponibles. Aujourd'hui encore, certains pensent que la reconnaissance de la parole continue ne sera atteinte que par le biais de traitements massifs et aveugles exigeant des puissances de calcul faramineuses. Il est vrai que les meilleurs systèmes actuels sont les plus « gourmands » en puissance de calcul, et les performances des modèles markoviens sont – jusqu'à un certain point – directement liées au nombre de paramètres de ces modèles.

Une étape dans le « réalisme » des maquettes a été atteinte dans les années 70, par l'emploi des mini-calculateurs puis des micro-processeurs. Les traitements acoustiques couramment utilisés – calculs de spectres, filtres linéaires – sont devenus des algorithmes et non plus des montages électroniques. De nouvelles techniques d'analyse plus performantes sont devenues réalistes, comme la prédiction linéaire, le filtrage homomorphique, et ont pu être employées dans les systèmes.

Il faut noter que jusque là, les études sur le dialogue vocal frisaient le surréalisme, avec des échanges entre l'utilisateur et le système exigeant plusieurs dizaines de secondes (le démonstrateur de Keal en 1977 permettait à l'utilisateur d'oublier la question avant que la réponse n'arrive!). Les recherches menées par l'équipe de Jelinek à IBM nécessitaient des puissances de calcul gigantesques¹³, ce qui faisait douter largement de leur utilité. L'évolution de la technologie montra que la réalité dépassait les plus folles espérances, puisque les processeurs actuels permettent de faire les mêmes calculs en temps réel.

Les circuits spécialisés ont permis d'atteindre des produits industriels (Nec vers les années 1980) qui ne se sont pas vendus, faute de marché réel et parce que les performances et l'ergonomie étaient insuffisantes. Depuis, les circuits de traitement du signal « universels » répondent bien mieux aux besoins.

5 Conclusion

Entre 1965 et 1980, la reconnaissance de la parole est passée d'une expérimentation tâtonnante à la réalisation de maquettes pré-industrielles. S'appuyant sur des connaissances linguistiques de plus en plus précises et opératoires, adaptant à ses besoins des modèles mathématiques et algorithmiques issus de l'optimisation combinatoire, des techniques de codage, et de l'intelligence artificielle, servie par les extraordinaires progrès

12. Ce phénomène est d'ailleurs courant dans la communication personne-personne.

13. L'un des auteurs a encore en mémoire la vision d'une salle de calcul gigantesque peuplée de calculateurs IBM de la série 370, entièrement consacré à la reconnaissance de la parole, lors d'une visite du centre IBM de Yorktown Heights, en 1976.

technologiques des circuits et des machines informatiques, elle a atteint une maturité technique permettant d'aborder dans de bonnes conditions la mise au point de dispositifs industriels.

A partir des années 1980, la question posée n'est plus : « un ordinateur peut-il reconnaître la parole ? », mais plutôt « à quoi un ordinateur qui reconnaît la parole peut-il servir ? ».¹⁴

C'est ainsi qu'en 1980 le Cnet décide d'organiser « une journée destinée à trouver le relais industriel pour le savoir-faire acquis [...] en traitement du signal de parole, et notamment pour trois produits : Dynamo, système de reconnaissance de mots isolés, [...] le SPI, Synthétiseur de Parole Intégré, [...] Phonex, système de boîtes postales vocales. » 130 participants provenant de plus de 60 organismes assistent à cette journée¹⁵.

Les applications potentielles de la reconnaissance de la parole sont multiples et connues depuis longtemps, depuis les utilisations « mains occupées – yeux occupés » jusqu'à la dictée vocale en passant par les contrôles d'environnement, les commandes de machines à laver ou de clignotants...

Le développement réel de services à base de commande vocale n'est que tout récent (fin des années 1980) et a résulté de plusieurs facteurs :

- la disponibilité banalisée de processeurs de traitement de signal bon marché et puissants,
- l'emploi de techniques de reconnaissances fiables et bien maîtrisées,
- la prise en compte des problèmes liés à la confrontation utilisateur – système.

Le troisième point, essentiel, résulte des opérations pionnières lancées à la fin des années 70 et au volontarisme de l'époque de « sortir » les systèmes de reconnaissance des laboratoires où ils se complaisaient depuis longtemps.

Cette volonté s'est d'abord traduite par des échecs retentissants : pas de produits réellement utilisables, et effet négatif produit sur l'industrie par ces échecs quant à l'intérêt de la reconnaissance de la parole. Cependant, c'est au prix de ces échecs de la période allant de 1978 à 1984 que l'on a pu enfin comprendre les véritables problèmes posés par la mise en exploitation d'un système de reconnaissance, problèmes jusque là masqués tant que les systèmes restaient dans les laboratoires.

À la fin des années 1970, le critère unique d'évaluation d'un système était son taux de reconnaissance mesuré dans des conditions plus ou moins réalistes. Les expérimentations ont montré que ce critère était surestimé et mineur, comparé à d'autres bien plus importants : robustesse du système face au milieu (locuteur, bruit, transmission), ergonomie (traitement des erreurs de reconnaissance, conception des mécanismes de dialogue...).

Citons pour illustrer notre propos l'exemple du composeur vocal, un des rares services de télécommunications directement lié à la reconnaissance de la parole¹⁶. En 1979, France-Télécom affiche sa volonté de développer des terminaux offrant cette fonction et met en concurrence deux techniques de reconnaissance par programmation dynamique : Moïse du Limsi et Dynamo du Cnet. Deux industriels sont retenus : Thomson et Jeumont-Schneider. En 1983–1984, le produit Thomson est vendu à la Fnac (au prix de 7500F), alors que le produit Jeumont-Schneider n'existe que sous forme de prototype. Dans les deux cas, il y a échec : pas de vente, ergonomie désastreuse, performances insuffisantes..., mais ceci met en évidence les problèmes réels. En 1988, des composeurs moins chers sont proposés aux États-Unis, et aboutissent à un nouvel échec. En 1993, le service est expérimenté puis généralisé aux États-Unis dans le réseau téléphonique, avec cette fois un succès apparent.

14. Il est intéressant, à ce point, d'indiquer que la question de l'utilité des recherches sur la reconnaissance a toujours été posée en arrière plan des recherches. Citons Pierce, en 1969, dans un article très célèbre[55] :

The purpose of this letter is to examine both motivations and progress in the area of speech recognition [...]. There is a reason for this beyond the fun and insight that such an investigation affords. The country is supporting such work. Is this wise? Are we getting our money's worth?

Malgré son urgence, cette question n'a reçu de réponse qu'une fois les principaux problèmes techniques maîtrisés, bien après 1980.

15. Note technique Cnet NT/LA-A/2, mai 1980

16. Le lecteur intéressé trouvera dans [15] des éléments sur l'histoire des relations entre les télécommunications et le traitement de la parole.

La reconnaissance de la parole est aujourd'hui annoncée par AT&T comme l'une des technologies majeures des prochaines années : plusieurs services fonctionnent aujourd'hui aux États-Unis avec des systèmes de reconnaissance simples (quelques mots plus ou moins isolés) utilisés par des millions d'utilisateurs.

Citons pour finir deux extraits de la note technique du Cnet NT/LAA/DIR/8, datée de décembre 1980 :

« Programme à *moyen terme*¹⁷ (axe Communications homme-machine) :

« – En reconnaissance, les priorités doivent être mises sur la reconnaissance de la parole continue en temps réel . . . , la reconnaissance multilocuteur . . . le niveau acoustique est essentiel (page 3).

« – Il faut tourner la page pour certaines étapes des systèmes de reconnaissance (*mots isolés, phrases constituées de mots séparés par des pauses avec des contraintes syntaxiques rigides, mots clés dans un flot de parole continue*¹⁸).

»

15 ans après, la page n'est pas encore tournée.

Références

- [1] Baker (J.K.). – *Machine-Aided Labeling of Connected Speech. Working Papers in Speech Recognition.* – Rapport technique, CMU, avril 1973.
- [2] Baker (J.K.). – The Dragon System. An Overview. *IEEE Trans. on ASSP*, vol. 23, n° 1, février 1975, pp. 24–29.
- [3] Bars (J.F.). – *Reconnaissance automatique de la parole continue; repérage d'un mot à l'intérieur d'une phrase.* – Note technique n° CEI/CSI/25, Cnet, octobre 1972.
- [4] Bates (M.). – The Use of Syntax in a Speech Understanding System. *In: IEEE Symposium on Speech Recognition*, pp. 226–233.
- [5] Battani (G.) et Meloni (H.). – *Mise en œuvre de contraintes phonologiques, syntaxiques et sémantique dans un système de compréhension automatique de la parole.* – Thèse de doctorat de spécialité en informatique, Université d'Aix-Marseille, 1975.
- [6] Baudry (M.). – *Étude du signal vocal dans sa représentation amplitude-temps. Algorithmes de segmentation et de reconnaissance de la parole.* – Thèse de docteur ès sciences mathématiques, Université Paris VI, 1978.
- [7] Baum (L.E.) et Eagon (J.A.). – An Inequality with Applications to Statistical Estimations for Probabilistic Functions of Markov Processes and to a Model for Ecology. *American Mathematical Society Bulletin*, vol. 73, 1967, pp. 360–362.
- [8] Bellissant (C.). – *Contribution à l'analyse et à la reconnaissance automatique de la parole.* – Thèse de docteur ès sciences mathématiques, Université de Grenoble I, 1978.
- [9] Boë & al (L.J.). – Informatique et parole en France : un quart de siècle après la rencontre. *In: 3ème Colloque Histoire de l'informatique.* Inria.
- [10] Bolt (Beranek) et Newman. – *Speech Understanding System.* – BBN report n° 3438, Boston, EU, 1976.
- [11] Bridle (J.S.) et Sedgwick (N.G.). – A Method for Segmenting Acoustics Patterns with Applications to Automatic Speech Recognition. *In: IEEE Icassp-77*, pp. 656–659.

17. Souligné par les auteurs.

18. Souligné par les auteurs.

- [12] Buisson (L.), Génin (J.), Hutin (M.), Mercier (G.), Querré (M.), Quinton (P.), Stéphan (P.) et Vives (R.). – Simulation d'un mini centre de renseignements automatiques expérimental. *In : F.A.S.E.*, pp. 260–270.
- [13] Caelen (J.). – *Un modèle d'oreille. Analyse de la parole continue. Reconnaissance phonétique.* – Thèse d'État, Toulouse, 1979.
- [14] Calliope. – *La parole et son traitement automatique.* – Masson, 1989.
- [15] Cartier (M.) et Gagnoulet (C.). – Télécommunications et traitement de la parole : quelques éléments historiques. *In : 20èmes JEP.* – Lannion, 1994.
- [16] Cartier (M.), Gresser (J.Y.), Lucas (J.J.) et Querré (M.). – Speaker Adaptation to an Automatic Speech Recognition System. *In : IEEE Conference on Speech Processing*, pp. 287–290. – Boston, États-Unis, 1972.
- [17] Chiba (S.). – A Speaker-Independent Word Recognition System. *In : Int. Joint Conf. on Pattern Recognition*, pp. 995–999. – Kyoto, Japon, 1978.
- [18] Chomsky (N.). – On Certain Formal Properties of Grammars. *Information and Control*, vol. 2, 1959.
- [19] Courbon (J.L.), Cartier (M.) et Lorand (P.). – Reconnaissance automatique de la parole: « Charles ». *Revue d'Acoustique*, n° 43, 1977, pp. 340–346.
- [20] Davis (K.H.), Biddulph (R.) et Balashek (S.). – Automatic Recognition of Spoken Digits. *Journal of The Acoustical Society of America*, vol. 24, 1952, pp. 637–642.
- [21] Denes (P.). – The Design and Operation of the Mechanical Speech Recognizer at University College, London. *J. Brit. IRE*, vol. 19, n° 4, 1959, pp. 219–229.
- [22] Dours et al. (D.). – L'apport de l'analyseur du projet Aria sur quelques exemples d'analyse phonétique. *In : 6ièmes journées d'Etudes sur la Parole*, pp. 214–227. – Toulouse, France, 1975.
- [23] Dreyfus-Graf (J.). – Sonograph and Sound Mechanics. *JASA*, vol. 22, n° 6, 1950, pp. 731–739.
- [24] Dreyfus-Graf (J.). – Le typo-sonographe phonétique ou phonétographe. *Bull. Techn. PTT*, n° 5, 1952.
- [25] Dreyfus-Graf (J.). – Phonétographe: Présent et Futur (Phonétographe III). *Bull. Techn. PTT*, n° 5, 1952.
- [26] Dreyfus-Graf (J.). – Compresseurs d'amplitude et cybernétique. *Bull. Techn. PTT*, n° 7, 1964.
- [27] Dreyfus-Graf (J.). – Parole codée (phonocode): reconnaissance automatique de langages naturels et artificiels. *Revue d'Acoustique*, n° 21, 1972.
- [28] Erman (L.D.). – A Functional Description of the Hearsay II Speech Understanding System. *In : IEEE Icassp*, pp. 799–802.
- [29] Forgie et al. (J.W.). – An Overview of the Lincoln Laboratory Speech Recognition System. *In : 88-ème Meeting de l'Acoustical Society of America.* – St-Louis, États-Unis, novembre 1974.
- [30] Gagnoulet (C.) et Couvray (M.). – Seraphine: a Connected Word Speech Recognition System. *In : Icassp 82.* IEEE, p. 887.
- [31] Comptes–rendus des Journées annuelles d'Etudes sur la Parole du Groupe Communication parlée du Galf, 1969–1979.
- [32] Gauvain (J.L.) et Mariani (J.). – Mozart: un système de reconnaissance globale de parole continue. *In : 11ème ICA.* – Paris, 1982.

- [33] Gresser (J.Y.), Cartier (M.) et Carré (R.). – *Rapport de mission en URSS*. – Note Technique n° CEI/7 - TMA/ETA/5, Cnet, décembre 1971.
- [34] Groc. – Compilation d'un métalangage de définition de structures relationnelles et son utilisation en reconnaissance de la parole. *In: Congrès Afcet-IRIA Reconnaissance des formes*, pp. 97–106. – Chatenay-Malabry, France, 1978.
- [35] Grossetti (M.). – Sciences de la parole : genèse d'une communauté scientifique en France. *In: XXèmes Journées d'Étude sur la Parole*. – Trégastel, France, 1994.
- [36] Haton (J.P.), Liénard (J.S.) et Mercier (G.). – *Compte-rendu d'une mission d'étude en URSS*. – Note Technique n° CEI/CSE/28, Lannion, France, Cnet Lannion, décembre 1975.
- [37] Haton (J.P.) et Pierrel (J.M.). – Data Structures and Architecture of Myrtille II Speech Understanding System. *In: 4th International Joint Conference on Pattern Recognition*, pp. 1033–1038. – Kyoto, Japon, 1978.
- [38] Hayes-Roth (F.) et Mostow (D.J.). – Syntax and Semantics in a Distributed Speech Understanding System. *In: IEEE Icassp*, pp. 421–424.
- [39] Jelinek (F.). – Fast Sequential Decoding Algorithm Using a Stack. *IBM Journal Research Development*, 1969, pp. 675–685.
- [40] Jelinek (F.). – Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech. *IEEE Trans. on Information Theory*, vol. IT 21, n° 3, mai 1975, pp. 250–256.
- [41] Jelinek et al. (F.). – Recognition of continuously read natural corpus. *In: IEEE Icassp*, pp. 422–424. – Tulsa, États-Unis, 1978.
- [42] Actes des 7èmes Journées d'Études sur la Parole. Volume 2, J. P. Haton (éditeur), Nancy, 1976.
- [43] Kato (Y.), Chiba (S.) et Nagata (K.). – Spoken digit recognizer. *Journal of the Institute of Electrical and Communication Engineers of Japan*, vol. 29, n° 11, 1964, pp. 690–691.
- [44] Klatt (D.H.). – Review of the Arpa Speech Understanding Project. *In: 92ème Meeting de l'Acoustical Society of America*. – San Diego, États-Unis, novembre 1976.
- [45] Kohda (M.), Nakatsu (R.) et Shikano (K.). – Speech Recognition in the Question-Answering System Operated by Conversational Speech. – Proc. Icassp 76, 1976.
- [46] Liénard (J.S.). – *Analyse, synthèse et reconnaissance automatique de la parole*. – Thèse d'État, Université de Paris VI, avril 1972.
- [47] Lowerre (B.T.). – *The Harpy Speech Recognition System*. – Thèse de PhD, Carnegie Mellon University, Pittsburg, États-Unis, 1976.
- [48] Mariani (J.J.) et Liénard (J.S.). – Acoustic-Phonetic Recognition of Connected Speech Using Transient Information. *In: IEEE Icassp*. pp. 667–670. – Hartford, États-Unis, 1978.
- [49] Mercier (G.). – Approximation stochastique et reconnaissance acoustique d'un vocabulaire limité. *Annales des Télécommunications*, vol. 25, n° 5–6, 1970, pp. 207–216.
- [50] Mercier (G.), Quinton (P.) et Vivès (R.). – Keal : un système pour un dialogue oral avec une machine. *In: Congrès de l'Afcet - Tome 2*, pp. 304–314. – Gif sur Yvette, France, novembre 1978.
- [51] Mori (R. De) et Piccolo (E.). – Automatic Detection and Description of Syllabic Features in Continuous Speech. *IEEE Trans. on ASSP*, vol. 24, n° 5, octobre 1976, pp. 365–379.

- [52] Myasnikov (L.L.). – Obektivnoe rasposnavanie zvukov reči (reconnaissance objective des sons de parole). *Revue Tech. Phys.*, vol. 13, n° 3, 1943, pp. 109–115.
- [53] Nakagawa (S.) et Sakai (T.). – A Real Time Spoken Word Recognition System in a Large Vocabulary with Learning Capability of Speaker Differences. *In: 4th IJCP*, pp. 985–989.
- [54] Newell (A.), Barnett (J.), Forgie (J.W.), Green (C.C.), Klatt (D.H.), Licklider (J.C.R.), Munson (J.), Reddy (D.R.) et Woods (W.A.). – *Speech understanding systems: final report of a study group*. – North Holland American Elsevier, 1973. (Le rapport original est daté de novembre 1971).
- [55] Pierce (J. R.). – Whither speech recognition? *Journal of the Acoustical Society of America*, vol. 46, n° 4, octobre 1969, pp. 1049–1051.
- [56] Pierrel (J.M.). – *Contribution à la compréhension automatique du discours continu*. – Thèse de spécialité, Université de Nancy I, 1975.
- [57] Querré (M.), Mercier (G.) et Gresser (J.Y.). – *Reconnaissance automatique de la parole: application de la programmation dynamique à l'identification des mots isolés. Résultats comparés sur 1000 mots*. – Note Technique n° CEI/CSI/44, Cnet, mars 1974.
- [58] Quinton (P.). – A Syntactic Analyzer Adapted to Speech Recognition. *In: IEEE Iccasp-76*, pp. 453–456. – Philadelphie, États-Unis, avril 1976.
- [59] Quinton (P.), Vivès (R.) et Gresser (J.Y.). – Dialogues avec un robot. *In: 6èmes Journées d'Etudes sur la parole*, pp. 411–421. – Toulouse, France, 1975.
- [60] Reddy (R.), Erman (L.D.) et Neely (R.B.). – A Model and a System for Machine Recognition of Speech. *IEEE Trans. on ASSP*, vol. AU 21, n° 3, juin 1973, pp. 229–237.
- [61] Rossi (M.). – Les contraintes phonologiques dans un système de reconnaissance de la parole. *In: Journées d'Etudes sur la Parole*, pp. 160–190. – Toulouse, France, 1975.
- [62] Roux (G.) et Poncin (J.). – *Reconnaissance de la parole; études préliminaires*. – Note Technique n° CRL-2011, ETA/CEI, Cnet-Lannion, 1967.
- [63] Sakai (I.) et Doshita (S.). – Speech recognition system of conversational sounds. *Journal of the Institute of Electrical and Communication Engineers of Japan*, vol. 41, n° 11, 1963, pp. 696 – 1702.
- [64] Sakoe (H.) et Chiba (S.). – A Dynamic Programming Approach to Continuous Speech Recognition. *In: 7ème ICA*. – Budapest, 1971.
- [65] Shigenaga (M.) et Sekiguchi (Y.). – Speech Recognition of Spoken Fortran Programs. *In: 4th. Int. Joint Conf. on Pattern Recognition*, pp. 990–994. – Kyoto, Japon, novembre 1978.
- [66] Stakushi (S.), Suzuki (H.) et Kido (K.). – Discrimination of some consonants in words with the aid of dictionary. *Journal of the Institute of Electrical and Communication Engineers of Japan*, vol. 54-C, n° 1, 1971.
- [67] Suzuki (J.) et Nakata (K.). – Phonemic classification and recognition of Japanese mono-syllable. *Journal of the Institute of Electrical and Communication Engineers of Japan*, vol. 46, n° 11, 1963, pp. 1680–1687.
- [68] Suzuki (J.) et Oizumi (J.). – The binary coding and learning for recognition of Japanese vowels. *Journal of the Institute of Electrical and Communication Engineers of Japan*, vol. 46, n° 3, 1963, pp. 291–299.
- [69] Tsemel (G.). – Reconnaissance des sons de la parole à partir de signaux échantillonnés. *In: Congrès de la section Acoustique de l'Académie des sciences de l'Union Soviétique*, 1958.

- [70] Tubach (J.P.). – *Reconnaissance automatique de la parole*. – Thèse d'État, Université de Grenoble, juillet 1970.
- [71] Vicens (P.). – *Aspects of Speech Recognition by Computer*. – Thèse de PhD, Stanford University, États-Unis, avril 1969.
- [72] Vintsjuk (T.K.). – Recognition of spoken words by methods of dynamic programming. *Kibernetika*, vol. 1, 1968, pp. 81–88.
- [73] Vintsjuk (T.K.). – Generative Grammars and Dynamic Programming in Speech Recognition with Learning. *In: IEEE Iccasp-76*. pp. 446–449. – Philadelphie, États-Unis, avril 1976.
- [74] Vintsjuk et al. (T.K.). – Phoneme by Phoneme of Speech Composed of the Words of a Given Vocabulary. *In: IEEE Iccasp-76*. pp. 450–452. – Philadelphie, États-Unis, avril 1976.
- [75] Vintsuk (T.K.). – Recognition of spoken words by methods of dynamic programming. *Kibernetika*, vol. 1, 1968, pp. 81–88.
- [76] Vivès (R.). – Élaboration d'un indice de ressemblance entre deux chaînes de longueur finie, quelconque. Application à la reconnaissance de mots isolés à l'aide d'un dictionnaire. *Annales des Télécommunications*, vol. 38, n° 3–4, 1973, pp. 123–129.
- [77] Vysotsky (G.). – *Automatic Speech Recognition Research in the USSR*. – Fall Church, États-Unis, Delphic Associates, 1986, *Monograph Series on Soviet Union*.
- [78] Wakita (H.) et Makino (S.). – *Recent work on speech recognition in Japan*, pp. 483–490. – Prentice-Hall, 1980. édité par W. A. Lea.