



HAL
open science

Technologies de la Parole au LIMSI: les années pionnières, de 1970 à 1990.

Joseph J Mariani, Jean-Sylvain Liénard

► To cite this version:

Joseph J Mariani, Jean-Sylvain Liénard. Technologies de la Parole au LIMSI: les années pionnières, de 1970 à 1990.. Journées d'Etudes sur la Parole, Jun 2018, Aix -en-Provence, France. hal-04413225

HAL Id: hal-04413225

<https://hal.science/hal-04413225>

Submitted on 23 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Technologies de la Parole au LIMSI: les années pionnières, de 1970 à 1990.

Joseph Mariani, Jean-Sylvain Liénard
LIMSI, CNRS, Université Paris-Saclay

Joseph.Mariani@limsi.fr, Jean-Sylvain.Lienard@limsi.fr



Icophone: Système de synthèse par diphonèmes à partir du texte utilisant un banc de 44 oscillateurs, résultant des travaux menés en collaboration entre le LIMSI (J.S. Liénard, D. Teil), le LAM (M. Castellengo, E. Leipp) et le LMP (J. Sapaly). L'Icophone V fut commercialisé en deux exemplaires par TITN.



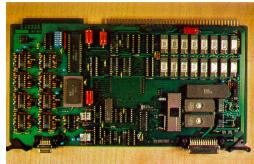
Banc de 32 filtres utilisé pour l'acquisition et le filtrage du signal de parole en vue de sa reconnaissance, réalisé en collaboration entre le LIMSI (J.S. Liénard et J. Mariani, dont le premier travail a été le réglage de leur fréquence centrale et de leur bande passante...) et le LMP.



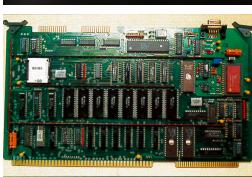
IBM 1130 utilisé dans les années 70 pour les recherches en Traitement Automatique de la Parole au LIMSI, alors dirigé par L. Malavard. On distingue l'unité centrale, le lecteur de cartes et l'imprimante. Le banc de filtres et le synthétiseur de parole Icophone lui étaient connectés. Pour le système de compréhension de la parole continue ESOPÉ (J. Mariani, 1978), l'IBM 1130 était connecté à un IBM 370/168, fonctionnant en temps partagé, situé au CIRCÉ (à présent IDRIS), centre de calcul du CNRS.



Travaux sur le dialogue Pilote-Avion de Chasse dans la salle machine du LIMSI en 1976. On reconnaît J. Mariani derrière le casque et le masque à oxygène. Les travaux étaient menés en collaboration avec le GIE Ispena et la société Crouzet en vue d'équiper le futur avion Rafale de Dassault Aviation. A défaut d'une implantation effective sur le Rafale, des essais furent conduits en mode opérationnel sur Mirage III B et débouchèrent sur la synthèse vocale de l'altimétrie pour avions et hélicoptères, et l'utilisation en simulateur de vol (système EVA). On notera qu'un tel système de dialogue n'existe toujours pas dans les cockpits actuels, s'agissant sans doute d'une des tâches les plus difficiles à accomplir.



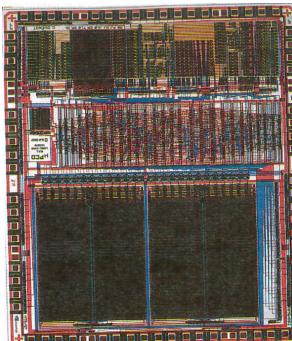
L'arrivée des microprocesseurs à la fin des années 70 permit la réalisation de systèmes complets sur une seule carte. En haut, la carte de synthèse vocale à partir du texte issue de l'Icophone (J.S. Liénard, D. Teil, B. Prouts, J.J. Gangolf). En bas, la carte de reconnaissance de mots isolés monolocuteur MOISE, fondée sur un algorithme de compression non linéaire et réalisée en 1979 (J. Mariani, J.S. Liénard, J.J. Gangolf).



Ces cartes furent commercialisées en 1981 par la société Vecsys, sous le nom d'ICO85 pour la synthèse et de RMI85, puis RMI88 pour la reconnaissance (240 mots). Cette dernière carte fut étendue à la reconnaissance multilocuteur, et fut suivie par la carte de reconnaissance de la parole continue RME186 (J.L. Gauvain, J. Mariani et J.S. Liénard) présentée à ICASSP 1982 à Paris.



Les cartes de reconnaissance et synthèse furent rassemblées dans un même boîtier en 1982 pour constituer un système complet de dialogue oral, également commercialisé par Vecsys (TLV02). L'existence de ce boîtier autonome d'E/S vocales permit le développement de nombreuses applications.



L'étape suivante fut la réalisation en 1989 d'un circuit intégré VLSI Full Custom pour effectuer la programmation dynamique (ou algorithme de Viterbi). Ce circuit dessiné par G. Quénot et réalisé en collaboration avec la société Bull, comportait 127.309 transistors. Il fut intégré dans une nouvelle carte de reconnaissance (5000 mots isolés ou 600 mots connectés), commercialisée également par Vecsys.



Le système de reconnaissance de parole Moïse fut modifié pour effectuer la reconnaissance du locuteur. Une Serrure Vocale fut présentée à la Fête de la Science en 1984, où on gagnait un Louis d'or (en chocolat) si on arrivait à se faire passer pour l'un des locuteurs autorisés. Sur ces bases, le système SESAME fut installé pour permettre l'accès au LIMSI à l'ensemble de ses membres pendant les années 80-90. Le mot à prononcer était « généalogie ».



Dans les années 80, d'autres travaux très précurseurs ont porté sur les modèles de langage probabilistes (A. Andreevski, G. Adda) dès 1979, l'évaluation des systèmes de reconnaissance multilingues dans le cadre d'une campagne internationale organisée par le groupe OTAN RSG10 (J. Mariani, 1981) ou des unités de décision (M. Adda-Decker, 1988) et l'adaptation au locuteur par Quantification Vectorielle (H. Bonneau-Maynard, 1987).



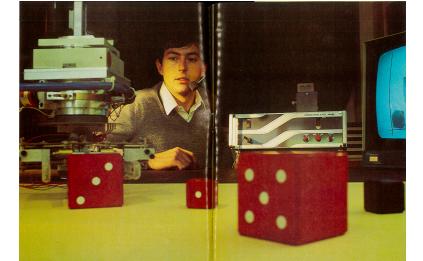
Les premiers travaux en analyse de la parole au LIMSI ont amplement utilisé le sonographe de Kay Elemetrics donnant une représentation fréquentielle, en bande large ou en bande étroite, de la parole (spectrogramme). D'un maniement complexe, le résultat était donné sur du papier spécial placé sur un rouleau. Il fut remplacé par des logiciels comme UNICE réalisé par J.L. Gauvain et diffusé par Vecsys en 1989.



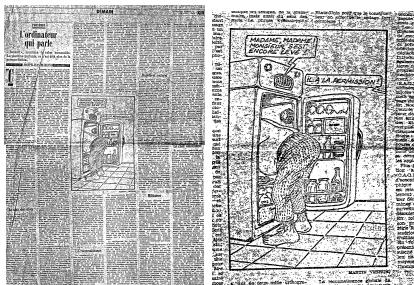
D'autres travaux en matière de reconnaissance vocale portèrent sur l'utilisation de systèmes experts pour la reconnaissance de spectrogrammes en 1983 (J. Mariani, M. Eskinazi, D. Memmi, P.E. Stern).



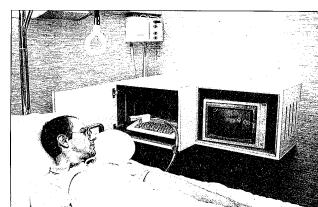
Les approches neuronales firent également l'objet de travaux précurseurs (réseaux à propagation guidée et *Never Ending Learning Process* de D. Béroule en 1985, modèle hybride TDNN par L. Devillers en 1987).



Au sein du département « Communication Homme-Machine », de nombreuses expériences furent menées sur l'utilisation d'interfaces vocales en robotique (D. Béroule), en synthèse d'images, et en communication multimodale dont la résolution du « Put That There » (Y. Bellik).



Dans Le Monde du 24 Août 1980, J. Mariani écrit un article de prospective sur les applications des technologies vocales, « L'Ordinateur qui Parle », avec une illustration de M. Veyron. On y parle déjà de robots humanoïdes.



Les systèmes de reconnaissance et synthèse vocale conçus au LIMSI et commercialisés par Vecsys furent utilisés dans différents types d'application, comme la commande pour les handicapés, en collaboration avec Proteor, ou la numérotation téléphonique vocale en collaboration avec Jeumont-Schneider en 1984. Un autre domaine d'application fut le contrôle aérien soit en mode opérationnel en collaboration avec le CENA, soit pour l'entraînement des élèves contrôleurs avec l'ENAC (F. Néel, K. Matrouf, M. Garnier-Rizet). Une autre collaboration avec la SNCF conduisit à la réalisation d'un système de réservation en ligne et d'un kiosque pour la réservation des billets de train (projet MASK).



Les années 90 virent ensuite l'utilisation massive des approches probabilistes (Modèles de Markov Cachés discrets puis continus) et neuronales pour la reconnaissance de parole continue de très grands vocabulaires indépendante du locuteur, la dictée vocale et la transcription de parole, et l'abandon progressif des travaux en électronique pour aller vers le tout logiciel.

