

## UNE APPLICATION DE LA RECONNAISSANCE DE LA PAROLE:

COMMANDE D'UN ROBOT DE MONTAGE PAR LA PAROLE

HANS PETER AMANN, HEINZ HÜGLI

INSTITUT DE MICROTECHNIQUE DE L'UNIVERSITÉ

RUE DE LA MALADIÈRE 71, 2000 NEUCHÂTEL 7

## Résumé

Une unité de communication comprenant un système de reconnaissance et un synthétiseur de la parole a été élaborée pour commander par la voix la phase d'apprentissage des mouvements d'un robot de montage. Les essais effectués montrent que la commande du robot entièrement par la parole est possible, même dans un environnement bruyant, si l'en utilise un dialogue bien structuré.

## 1. Introduction

Depuis plusieurs années, des efforts considérables ont été entrepris dans le but d'améliorer la communication homme/machine. La communication de la machine vers l'homme a été améliorée notamment par l'introduction de la technique des menus. La communication inverse, de l'homme vers la machine a été favorisée par l'introduction d'outils supplémentaires comme la souris, le manche à balai, le crayon lumineux, la tablette de digitalisation. Une possibilité supplémentaire, qui se situe encore plus proche de la communication humaine, est l'utilisation de la voix dans l'un ou les deux sens de la communication homme/machine. Dans un sens, les locutions de l'utilisateur sont reconnues comme commandes ou plus généralement comme entrées de la machine. Dans l'autre sens la machine émet des messages parlés par l'intermédiaire d'un synthétiseur de la voix. Un grand avantage de cette méthode est évidemment son aspect naturel. En plus, les mains et les yeux restent libres pour d'autres tâches.

Ces dernières années, la synthèse et la reconnaissance de la voix se sont rapidement développées. Un grand nombre de systèmes sont apparus sur le marché (3). Alors que ces moyens sont maintenant bien connus, leur application a encore été explorée (1)(2). Le but du projet décrit dans ce papier, est d'explorer les méthodes de communication parlée dans une application concrète: la commande d'un robot de montage.

## Abstract

A speech communication unit constituted of a speech recognizer and synthesizer has been developed. It is used to control the teaching phase of an industrial robot by voice only. We show the importance of using a well structured speech dialog. The tests performed show positive results, even in a noisy environment.

Voici quelques questions qui se posent pour une telle application:

- Quel est le temps de réaction de la machine?
- Quelle est la fiabilité de la reconnaissance?
- Comment l'utilisateur peut-il réagir en cas de fausse reconnaissance par la machine?
- Comment réagit le système si l'utilisateur ne suit pas la syntaxe des commandes?
- Comment l'utilisateur peut-il être guidé?
- Est-ce qu'une telle solution est indiquée pour le problème choisi?

Dans ce article, nous présentons d'abord le robot à commander, puis la conception de l'unité de communication par la parole nouvellement développée. Ensuite seront présentés la structure des commandes possibles et le déroulement d'un dialogue homme/machine dans le cas d'une communication sans fautes de l'utilisateur. Plus loin, nous abordons le problème du traitement des erreurs dues aux fausses reconnaissances et au non-respect de la syntaxe par l'utilisateur. Une courte présentation du matériel utilisé et une discussion des résultats obtenus termineront ce papier.

## 2. Commande du robot

Le travail d'un robot de montage est caractérisé par la répétitivité de ses mouvements. Ces mouvements doivent être enregistrés au préalable lors d'une phase d'apprentissage au cours de laquelle l'opérateur dicte les mouvements à la machine.

## 2.1 Ancienne conception

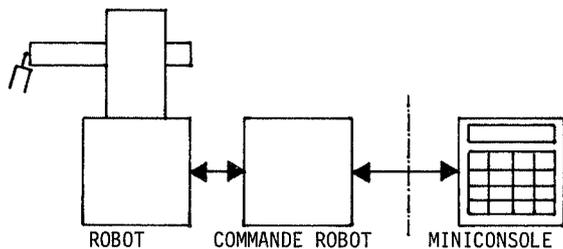


figure 1 ancienne conception

Selon la figure 1, l'apprentissage des mouvements du robot se fait par l'intermédiaire d'une miniconsole portable, avec clavier et affichage. L'opérateur commande les mouvements du robot qui s'effectuent à une vitesse ralentie. Chaque mouvement est mémorisé par le robot.

Cet apprentissage interactif demande une réaction instantanée de l'utilisateur. Celui-ci est donc contraint de surveiller les mouvements du robot et, simultanément, d'effectuer des commandes sur son clavier. Pour lui faciliter la tâche, on aimerait remplacer cette commande manuelle par une commande vocale, c'est-à-dire le clavier par un système de reconnaissance de la parole. De même on remplacera le message visible par un message vocal, c'est-à-dire, l'affichage du terminal portable par un synthétiseur de la parole.

## 2.2. Nouvelle conception

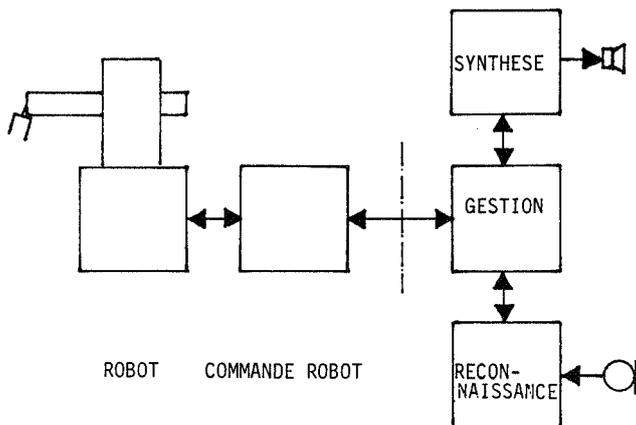


figure 2 nouvelle conception

La miniconsole de la figure 1 est donc remplacée par une unité de communication parlée (UCP) comprenant:

- un sous-système de reconnaissance de la parole
- un sous-système de synthèse de la parole
- un sous-système de gestion

Le sous-système de synthèse de la parole:

Il génère, sur demande, les messages parlés qui sont envoyés à l'opérateur par haut-parleur ou casque. Les messages sont synthétisés à partir du vocabulaire stocké dans les mémoires mortes du système dont la figure 3 dresse un exemple. Un message peut être constitué d'un seul mot du vocabulaire mais aussi d'un enchaînement quelconque de plusieurs mots.

Le sous-système de reconnaissance de la parole

Il comprend le microphone et l'unité de reconnaissance de la parole proprement dite.

Le microphone recueille la parole de l'opérateur qui le porte généralement sur lui. Il est fixé sur une monture et disposé devant la bouche, de manière à assurer, à la fois une bonne séparation entre parole et bruit de fond et les déplacements libres de l'opérateur.

L'unité de reconnaissance a pour but d'identifier un mot prononcé avec un des mots de son vocabulaire.

Il connaît deux modes de travail:

En mode d'apprentissage, l'unité stocke en mémoire les caractéristiques acoustiques des mots du vocabulaire. Pour l'opérateur, cet apprentissage consiste généralement à prononcer successivement tous les mots du vocabulaire de manière à mettre en mémoire ses propres caractéristiques.

En mode de reconnaissance, l'unité analyse donc les mots prononcés et les identifie par comparaison avec les caractéristiques acoustiques du vocabulaire. Le vocabulaire complet est donné en figure 4. Sur commande, il est possible de réduire dynamiquement le vocabulaire significatif pour la reconnaissance à un sous-ensemble du vocabulaire complet.

Le sous-système de gestion

Il est le centre de notre nouvelle unité de communication UCP. Il réalise et gère les liaisons avec les deux autres sous-systèmes et avec le robot. Vis-à-vis du robot il remplit la même fonction de commande que la miniconsole. Vis-à-vis de l'opérateur, il assure la communication parlée par un dialogue que nous décrivons maintenant.

i	m	i	g	m
1	ACTUEL	1	17	ZERO
2	AXE	2	1	UN
3	DE L'	3	2	DEUX
4	DU	4	3	TROIS
5	ENCLENCHER OU	5	4	QUATRE
	DECLENCHER	6	5	CINQ
6	ENCORE UNE FOIS	7	6	SIX
7	ENREGISTREMENT	8	7	SEPT
	EFFECTUE	9	8	HUIT
8	JE SUIS PRET	10	9	NEUF
9	LIAISON AVEC ROBOT	11	10	DIX
	INTERROMPUE	12	11	ONZE
10	NUMERO	13	12	DOUZE
11	PAS	14	13	TREIZE
12	RELAIS	15	14	QUATORZE
13	TAILLE	16	15	QUINZE
14	ZERO	17	16	SEIZE
15	UN	18	18	AXE
16	DEUX	19	18	PAS
17	TROIS	20	18	MONTE
18	QUATRE	21	18	MONTE ENCORE
19	CINQ	22	18	DESCEND
20	SIX	23	18	DESCEND ENCORE
21	SEPT	24	18	ENREGISTRER
22	HUIT	25	18	RELAI
23	NEUF	26	26	ENCLENCHER
24	DIX	27	26	DECLENCHER
25	ONZE	28	18	OUT
26	DOUZE	29	29	STOP
27	TREIZE	30	30	SORTIE
28	QUATORZE	31	31	CORRECT
29	QUINZE	32	31	FAUTE
30	SEIZE			

i index  
m mot

vocabulaire synthèse  
figure 3

i index  
g mot

g sous-groupe syntaxique

vocabulaire reconnaissance  
figure 4

### 3. Le dialogue homme/machine

#### 3.1 Syntaxe des commandes

Les commandes du robot sont constituées des phrases dont nous extrayons les mots-clés pour en constituer le vocabulaire de reconnaissance. La suite correcte de ces mots dans une commande est importante et doit suivre la syntaxe, montrée dans le graphe de la figure 5.

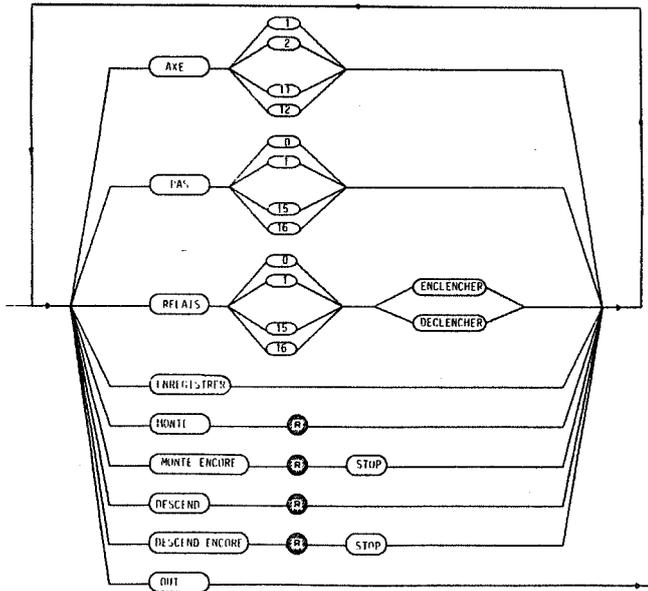


figure 5 syntaxe des commandes

La syntaxe est donnée par les possibilités du robot, par ses commandes et par l'équipement.

#### 3.2 Dialogue

Pour pouvoir dialoguer avec l'opérateur, le UCP doit connaître cette syntaxe. Cela lui permet de poser des questions adéquates, qui indiquent à l'opérateur quel type de réponse est demandé. A chaque niveau, l'UCP quitte la réponse précédente de l'utilisateur et pose la question suivante. Ceci permet à l'utilisateur de se situer dans le contexte de commande et de contrôler immédiatement si le mot énoncé précédemment à été bien reconnu. Pratiquement, quittance et question suivante sont liées dans un même message par soucis de simplifier et d'accélérer le dialogue.

Pour valider la commande entière et pour démarrer l'action du robot, l'opérateur énonce le mot de confirmation "correct".

Voilà un exemple typique de dialogue:

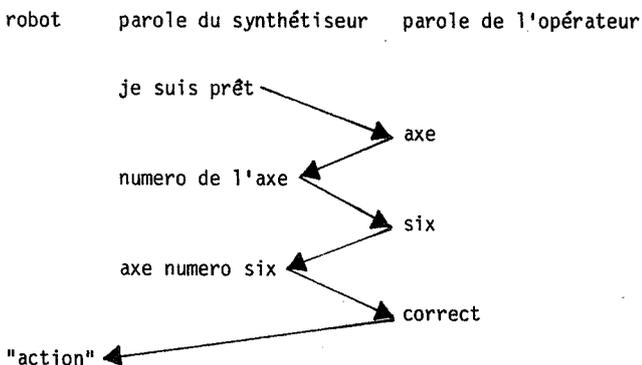


figure 6 dialogue idéal

Les problèmes qui peuvent apparaître sont les suivants:

- fausse reconnaissance: Le mot reconnu est possible d'après l'organigramme, mais pas le mot voulu par l'utilisateur.
- mauvaise acquisition: Le signal analysé ne présente pas les caractéristiques d'une bonne locution. (Un petit bruit a fait démarrer l'acquisition, le niveau est trop faible etc..).
- mot prononcé hors contexte: L'utilisateur prononce un mot qui est non conforme à la syntaxe de commande.

#### 3.3 Traitement d'erreurs

Voici les solutions adoptées:

- une fausse reconnaissance: L'utilisateur reconnaît ce cas grâce à la question suivante du système et doit donc pouvoir réagir à ce niveau-là. Il faut donc introduire une branche supplémentaire dans le graphe des commandes qui permet de revenir à la question précédente (figure 8).
- mauvaise acquisition: Elle est détectée automatiquement par le système de reconnaissance qui rend un code d'erreur au système de gestion. Le programme de dialogue réagit alors en demandant une répétition de la locution.

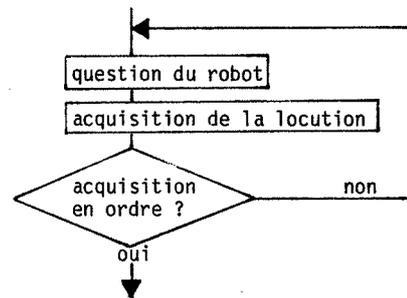


figure 7 mauvaise acquisition

- Le mot prononcé est hors contexte: Deux cas se présentent suivant si l'unité de reconnaissance n'a pas reconnu le mot ou si un (faux) mot a été reconnu. Si aucun mot a été reconnu on se trouve dans le cas de la mauvaise acquisition mentionnée ci-dessus: l'UCP réagit en demandant une répétition de la locution ce qui informe l'utilisateur de sa faute. Dans le deuxième cas, l'utilisateur est informée de sa faute en constatant que la quittance donnée n'est pas la bonne.

### 4. Réalisation et résultats

#### 4.1 Réalisation

L'UCP a été développée pour commander l'apprentissage d'un robot "Ecureuil" de MICROBO SA, St.Aubin et c'est sur ce équipement que nos essais ont été fait.

L'UCP est constituée des éléments suivants:

- a) une carte de reconnaissance

La carte SDR600/SMP au format simple-europe de SPEECH-DESIGN utilise pour la reconnaissance un ensemble de circuits de NEC, Nippon Electric Company. Elle permet de stocker et de reconnaître un maximum de 128 mots, regroupés en un maximum de 30 sous-ensembles syntaxiques. La reconnaissance est 'mono-locuteur', c'est-à-dire que elle ne fonctionne bien que pour la personne qui a enregistré les références. La communication avec l'extérieur se fait dans la version utilisée par un interface parallèle.

Le microphone directionnel, monté sur la tête est du type Shure SM-10.

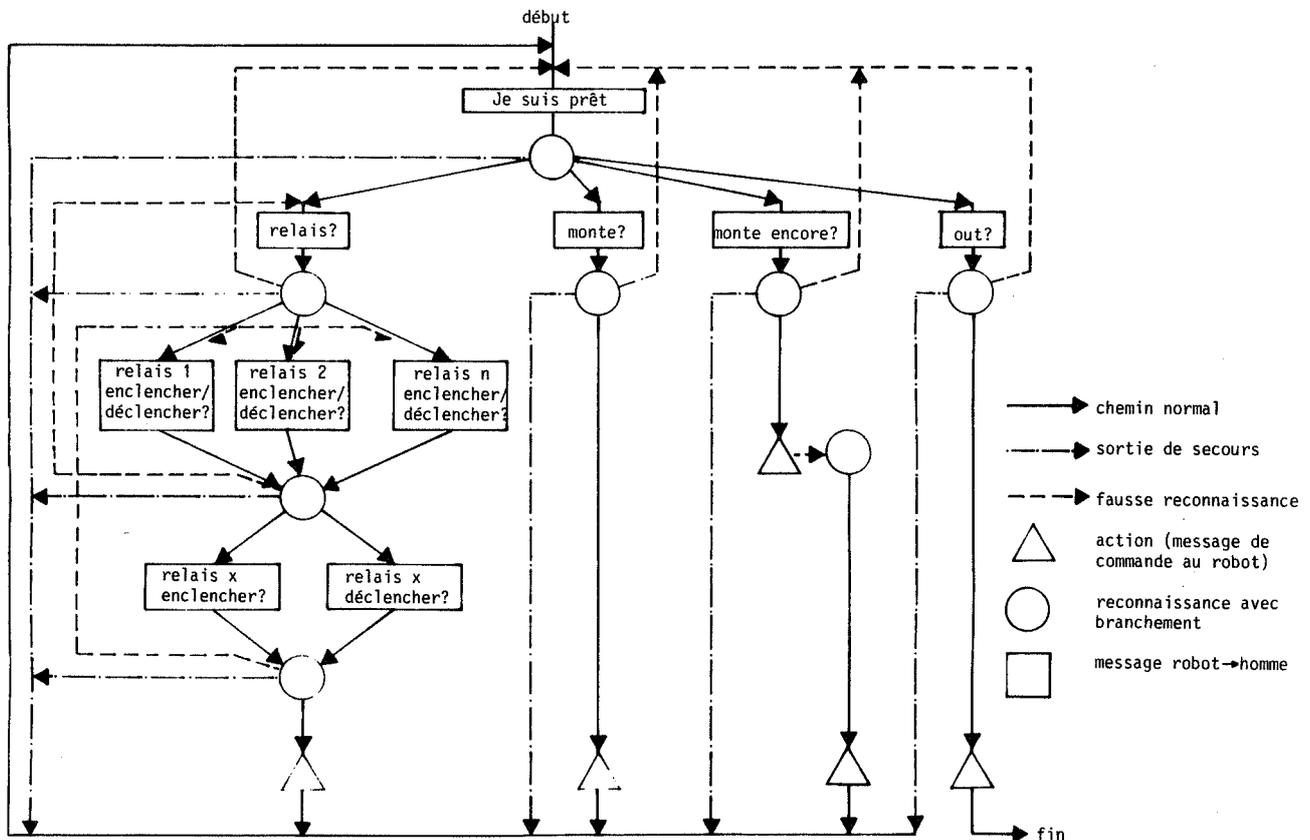


figure 8 organigramme

#### b) une carte de synthèse

La carte SDC100/SDX300, également de SPEECH-DESIGN, une carte au format simple-europe, permet de restituer les locutions stockées sur ROM. Ces locutions ont été enregistrées au préalable à l'aide de la même carte et copiées sur des mémoires mortes. La carte se connecte, dans la version utilisée, par un interface parallèle à la carte de gestion.

#### c) une carte de gestion

Il s'agit d'une carte processeur 8 bit standard équipée des interfaces parallèles pour la communication avec les cartes de reconnaissance et de synthèse et d'un interface sériel pour la liaison avec l'électronique du robot. Un deuxième interface sériel permet l'adjonction optionnelle d'une console de supervision et de développement.

#### 4.2 Résultats

Les essais effectués avec l'ensemble UCP-Ecureuil ont montrés que l'apprentissage des mouvements du robot pouvait se faire exclusivement avec la voix, même dans un environnement bruyant. La fiabilité de la reconnaissance, pour un utilisateur entraîné et appliqué, est suffisante pour pouvoir travailler d'une manière efficace. Grâce à la syntaxe, qui permet de réduire le nombre de comparaisons lors d'une reconnaissance, le temps de réaction reste raisonnable.

Pour toutes les commandes qui déclenchent un mouvement ou qui font une sélection, le processus de reconnaissance est suffisamment rapide. Par contre pour les commandes qui demandent une grande précision temporelle, par exemple pour stopper un mouvement, il est très difficile d'intervenir au bon moment. Ce problème n'est pas seulement dû au processus de la reconnaissance mais également à l'homme qui met un temps pour s'exprimer.

Un autre problème réside dans le fait que le dialogue avec le robot est monotone. Si ce point ne peut pas être amélioré, le dialogue avec la machine pourrait devenir très ennuyant pour un opérateur qui travaille à long terme sur un tel système.

#### 5. Conclusion

Le but envisagé de ce projet, la commande de la phase d'apprentissage d'un robot uniquement par la parole, a été atteint. Même dans un environnement bruyant (ventilation, mouvement du robot), un travail correct est possible.

Nos essais ont démontré que l'utilisation de la reconnaissance de la parole comme seul moyen de commande n'est pas satisfaisant. Une bonne combinaison des différentes moyens d'entrée/sortie avec la reconnaissance de la parole doit être trouvée dans chaque cas particulier.

#### 6. Remerciements

Nous remercions tous ceux qui ont rendu possible ce projet intéressant, notamment Mlle Constance Martin, MM Pascal Kunz et Denis Jeanrenaud. Une partie de ce travail a été supportée par la CERS et les maisons suivantes: Asulab SA, Autophon SA, CIR SA, CSEM SA, Hasler SA, Mettler SA.

#### 7. Références

- (1) Marek Okulewicz, 'Speech Control of Robots', Sensor Review, Vol. No 4, Oct 1984, p 184 ff
- (2) Speech Tech'85, Proceedings, April 85, New York
- (3) "1984 Voice processing market directory", Speech Technology, Vol 2, Nr 2, Jan/Feb 1984.