



THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Un système de messagerie vocale utilisant un éditeur de sons

Mousel, Josiane

Award date:
1985

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

FACULTES UNIVERSITAIRES
NOTRE DAME DE LA PAIX
DE NAMUR

CENTRE DE RECHERCHES EN INFORMATIQUE
DE NANCY

**UN SYSTEME DE MESSAGERIE VOCALE
UTILISANT UN EDITEUR DE SONS**

Mémoire présenté en vue de l'obtention
du grade de licencié et maître en informatique

par

Josiane MOUSEL née MULLER

en Septembre 1985
devant le Jury composé de:

Henri Leroy (Président du Jury)
Claude Cherton (Directeur de mémoire)
Jean-Marie Pierrel (Promoteur de stage)
Jacques Berleur (Membre invité)

TABLE DES MATIERES

Chapitre 1: INTRODUCTION	1
Chapitre 2: TRAITEMENT AUTOMATIQUE DE LA PAROLE	3
Sect. 1: Reconnaissance automatique de la parole	3
Sect. 1.1: Quelques notions de base	3
Sect. 1.2: L'intérêt des entrées vocales	4
Sect. 1.3: Typologie des systèmes	5
Sect. 2: Synthèse de la parole	10
Sect. 2.1: Quelques généralités	10
Sect. 2.2: Typologie des systèmes	10
Sect. 3: Les applications	12
Sect. 3.1: Les applications de la reconnaissance de la parole	12
Sect. 3.2: Les applications de la synthèse de la parole	13
Chapitre 3: COMMUNICATION HOMME-MACHINE	14
Sect. 1: Généralités	14
Sect. 2: Typologie des dialogues	18
Chapitre 4: EDITEURS DE SONS	21
Sect. 1: Généralités	21
Sect. 1.1: Rôle de l'éditeur de sons	21
Sect. 1.2: L'architecture matérielle	22
Sect. 2: L'éditeur de sons EDIT	23
Sect. 2.1: Les types d'objets d'EDIT	23
Sect. 2.2: Les fonctions principales d'EDIT	27
Sect. 3: Applications de l'édition de sons	30
Sect. 3.1: Applications professionnelles	30
Sect. 3.2: Applications grand public potentielles	30
Chapitre 5: SYSTEMES INFORMATIQUES DE MESSAGERIE VO- CALE	33
Sect. 1: Généralités	33
Sect. 1.1: L'intérêt des systèmes informatiques de communication	34
Sect. 1.2: L'intérêt des systèmes de messagerie vocale	35

Sect. 1.3: Les applications de la messagerie informatique vocale	37
Sect. 2: Le système de messagerie vocale MVOC	37
Sect. 2.1: Organisation générale de MVOC	38
Sect. 2.1.1: Les types d'utilisateurs et modes d'accès	38
Sect. 2.1.2: Les types de messages	39
Sect. 2.1.3: Les services offerts	39
Sect. 2.1.4: L'émission des messages	40
Sect. 2.1.5: La consultation de messages	41
Sect. 2.1.6: L'interface avec l'utilisateur	43
Sect. 2.1.7: Les informations nécessaires	44
Sect. 2.1.8: Les types d'objets utilisés	45
Sect. 2.2: Le dialogue oral utilisateur-système	50
Sect. 2.3: Evaluation du système MVOC	54
Chapitre 6: CONCLUSION	56
Annexe A: ENCHAINEMENT DU DIALOGUE ORAL	58
Annexe B: COMPARAISON AVEC LE SYSTEME D'AUDIO MES- SAGERIE D'IBM	72

TABLE DES FIGURES

Fig. 1. Environnement matériel de l'éditeur de sons	22
Fig. 2. Liste des types d'objets d'EDIT	25
Fig. 3. Schéma des relations entre les différents types d'objets	26
Fig. 4. Les fonctions d'EDIT	28
Fig. 5. Liste des types d'objets de MVOC	46
Fig. 6. Schéma des relations entre les différents types d'objets	47
Fig. 7. Les fonctions primitives définies dans MVOC	48
Fig. 8. Dialogue en début et en fin de session, cas de l'abonné	51
Fig. 9. Dialogue en début et en fin de session, cas du non-abonné	52
Fig. 10. Dialogue lors de l'écoute d'un message	53
Fig. 11. Schéma du dialogue	58
Fig. 12. Tableau comparatif	72

REMERCIEMENTS

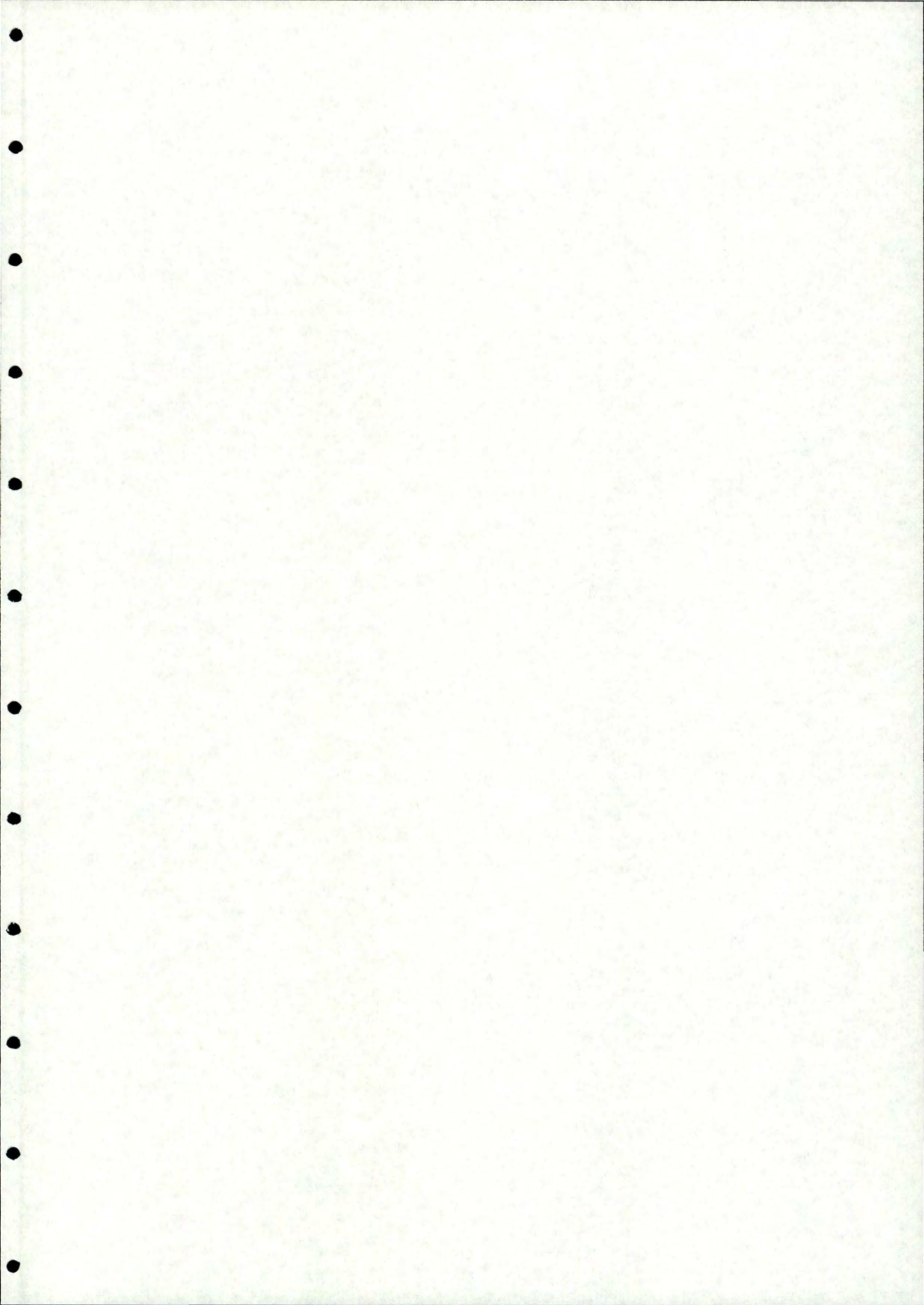
Je tiens à exprimer ma reconnaissance envers toutes les personnes dont la compréhension et la patience ont permis la réalisation de ce mémoire. En particulier je voudrais remercier

Monsieur C. Cherton, Professeur aux Facultés Universitaires de Namur, d'avoir accepté la direction de ce mémoire

Monsieur J.P. Haton, Professeur à l'Université de NancyI, d'avoir eu la gentillesse de m'accueillir dans son équipe de recherche

Monsieur J.M. Pierrel, Professeur à l'Université de NancyI, pour les conseils qu'il m'a prodigués tout au long de mon stage à Nancy

Messieurs H. Leroy et J. Berleur pour avoir accepté de bien vouloir juger ce travail.



CHAPITRE 1

INTRODUCTION

Depuis le lancement du projet ARPANET en 1968, les réseaux de transmission de données n'ont cessé de se développer, évolution qui a favorisé l'essor des services informatiques de messagerie. Actuellement ces derniers se limitent encore principalement à la gestion de messages écrits, mais, grâce à l'évolution des techniques de traitement du signal vocal (en particulier dans le domaine de la reconnaissance, de la synthèse et de la compression de la parole) et à la numérisation progressive des réseaux téléphoniques, ils vont également pouvoir traiter des messages vocaux. Des recherches sont notamment en cours sur la transmission de parole par paquets (analogue à celle des données). Dans [COLE-82] on souligne que cette transmission de la parole par paquets ne remplacera pas, dans l'immédiat, le système téléphonique analogique actuel, mais que d'un point de vue technique et économique elle conviendrait bien à de la parole qui soit est déjà numérisée, soit fait partie d'une communication non-interactive, soit est envoyée avec des données d'un autre type (alphanumérique ou graphique) dans la cadre de services multi-média.

Implémentés sur des réseaux de communication informatiques, les systèmes de messagerie offrent à leurs utilisateurs comme atouts majeurs, l'indépendance face aux horaires et à la distance ainsi qu'une diffusion multiple simultanée et confidentielle des messages [BERGER-84]. Un système de messagerie vocale y ajoute d'un côté l'introduction de la parole, qui est le moyen de communication privilégié de l'homme, et de l'autre la possibilité d'accéder au système par un simple combiné téléphonique, généralisant ainsi l'accès à ces réseaux de communication [HATON-84-a].

Notre stage au CRIN avait comme objet de réfléchir sur l'organisation générale d'un système de messagerie vocale. Il s'agissait de définir et de mettre en oeuvre un système qui, en s'appuyant sur des outils disponibles au laboratoire (en particulier sur le système EDIT, éditeur de sons offrant un ensemble de fonctions de manipulation d'objets sonores et des modules d'acquisition, de stockage, de reconnaissance et de restitution de la parole), permettrait l'enregistrement, la sauvegarde et la consultation de messages. L'accès au système devait pouvoir se faire en langage parlé via un téléphone standard en plus du mode clavier-écran classique. L'étape suivante de notre travail était la conception et la réalisation d'un dialogue oral utilisateur-machine simulé sur terminal en vue d'essais et de démonstrations.

Dans ce qui suit, il sera d'abord question de techniques de traitement automatique de la parole. On y abordera le domaine des systèmes de reconnaissance et de

synthèse de la parole, le système de messagerie vocale y faisant appel respectivement pour comprendre les énoncés de gestion du dialogue avec l'utilisateur et pour restituer des messages vocaux enregistrés. Les avantages et inconvénients d'entrées vocales à un système seront discutés avant de passer en revue les différents types de systèmes de reconnaissance et de synthèse de la parole ainsi que quelques applications envisagées.

La communication homme-machine sera abordée ensuite. Ayant exposé quelques idées générales sur la manière de concevoir un dialogue qui soit adapté à la fois à l'utilisateur et à l'application considérée, on présentera une typologie de dialogues allant de systèmes à question-réponse aux systèmes de dialogue en langage quasi-naturel, c'est à dire langage naturel restreint à un domaine d'expertise.

Nous continuerons avec une présentation générale des éditeurs de sons, de leur utilité et architecture matérielle. Le logiciel EDIT, sur lequel s'appuie notre système de messagerie vocale, sera décrit plus en détail. En particulier nous verrons quels sont les types d'objets qu'il manipule. Les applications mentionnées ensuite nous montreront que l'édition de sons peut servir les utilisateurs spécialistes étudiant le signal lui-même ainsi que les non-spécialistes travaillant sur des objets sonores.

Après avoir situé notre travail dans le cadre des problèmes du traitement de la parole, de la communication homme-machine et des éditeurs de sons, nous aborderons le domaine des systèmes de messagerie, plus particulièrement celui des messageries vocales. Nous en spécifierons le rôle et l'utilité avant de passer au système de messagerie vocale MVOC que nous avons conçu et dont nous présenterons la configuration, les types de messages manipulés, les fonctions offertes aux différents utilisateurs et le type de dialogue utilisé pour la communication avec le système. Nous terminerons avec la description de la réalisation du dialogue oral utilisateur-machine et avec une évaluation comparative de MVOC.

Une conclusion finale fera le point en matière de messageries vocales et de leurs perspectives d'avenir.

CHAPITRE 2

TRAITEMENT AUTOMATIQUE DE LA PAROLE

Les progrès des recherches menées sur le traitement automatique de la parole ces trente dernières années, ont contribué largement à l'apparition récente de systèmes de messagerie vocale. Non seulement des algorithmes plus performants ont été développés au niveau du traitement du signal vocal lui-même, mais l'intégration de connaissances linguistiques a également aidé à améliorer la qualité des logiciels de reconnaissance et de synthèse de la parole.

Nous allons d'abord exposer les notions de base de la reconnaissance automatique de la parole, les possibilités qu'offrent les entrées vocales à un système informatique ainsi que les limites de ces interfaces. Ensuite nous proposerons une classification des systèmes de reconnaissance en fonction de leur complexité, classification qui reflète aussi leur évolution historique. Une typologie analogue sera proposée pour les systèmes de production automatique (synthèse) de la parole. Enfin, différentes applications seront passées en revue.

1. Reconnaissance automatique de la parole

1.1. Quelques notions de base

D'après W.A. Lea [LEA-80-a], les chercheurs en reconnaissance de la parole sont d'accord sur le fait qu'un système de reconnaissance automatique de la parole doit être un ordinateur ou une machine qui accepte en entrée des commandes ou des questions introduites par téléphone ou par micro et qui génère en sortie une réponse ou une action appropriée. A cet effet, le système transforme le signal vocal continu en une représentation discrète qu'il est capable d'interpréter.

Ainsi, un système de reconnaissance automatique de la parole essaie essentiellement de dégager le contenu d'un énoncé: il s'agit de comprendre le sens de l'énoncé dans le but de produire une réponse adéquate. Or, dans [LEA-80-a], on souligne qu'un énoncé oral transmet, outre son contenu, également des informations sur la personne qui parle (sexe, accent), sur sa manière de prononcer les mots (distinctement ou non) et sur les conditions environnantes (bruit). Ces facteurs influent beaucoup sur les caractéristiques du signal acoustique en entrée, si bien que certains systèmes de reconnaissance de la parole intègrent des connaissances sur le locuteur, sa prononciation particulière et sur l'environnement afin d'améliorer la compréhension de la phrase.

En toute généralité, on peut dire qu'un système de reconnaissance automatique de la parole est d'autant plus performant qu'il:

- produit des réponses correctes, même dans un environnement bruyant;
- admet une grande population de locuteurs et n'exige pas d'apprentissage préalable;
- permet le discours continu d'une manière naturelle, impliquant ainsi un vocabulaire considérable, la parole continue et une structure des messages assez flexible;
- est indépendant de l'application donnée;
- garantit un temps de réponse acceptable et n'est pas trop exigeant quant à la taille de mémoire et la puissance de CPU requises;

Notons encore que certains auteurs distinguent entre compréhension et reconnaissance de la parole. En compréhension automatique de la parole, l'objectif est de comprendre suffisamment le contenu global d'un énoncé vocal pour pouvoir sélectionner la bonne réponse tandis qu'en reconnaissance automatique de la parole, on s'efforce de reconnaître à tout prix la totalité des éléments (voyelles, consonnes ou autres sous-entités) de la phrase prononcée. Or cette reconnaissance au sens strict est considérée comme très difficile voire impossible, de sorte que la plupart des réalisations ont recours à des techniques de compréhension. Ainsi, cette distinction s'estompe [PIERREL-81].

1.2. L'intérêt des entrées vocales

Dans [LEA-80-b], on avance les trois critères d'évaluation suivants:

a) **Utilisation efficace des facultés humaines de communication**

Les avantages sont dûs à la grande facilité qu'a l'homme de communiquer par le langage parlé (spécialement par sa langue maternelle):

- la parole est le moyen de communication le plus naturel, familier et spontané de l'homme;
- son utilisation ne nécessite aucun entraînement spécial;
- c'est aussi le moyen de communication humain le plus rapide, comparé à l'écriture et la dactylographie;
- elle peut être utilisée pour la communication simultanée entre hommes et machines, hommes et machines étant tous les deux capables de la comprendre;
- elle permet la communication homme-machine multi-mode

Il y a cependant des restrictions à ces possibilités, dues surtout aux capacités limitées des machines:

- l'homme peut prononcer des phrases qui, bien que correctes, sont incompréhensibles par la machine;

- certains systèmes imposant des contraintes plus ou moins fortes, l'homme doit s'entraîner à restreindre ses productions linguistiques à l'ensemble des énoncés compréhensibles par la machine;
- sa rapidité d'élocution risque d'être fortement entravée s'il doit prononcer les mots isolément ou s'il doit utiliser des constructions de phrases inhabituelles;
- dans le cas de communications simultanées entre deux hommes et une machine, un homme peut oublier à qui il s'adresse et la communication homme-machine peut être troublée.

b) **Compatibilité avec des circonstances inhabituelles**

Ces arguments influent moins sur la décision d'achat d'un système de reconnaissance de la parole, mais néanmoins ils peuvent parfois entrer en considération:

- la parole peut être employée dans l'obscurité, à travers des obstacles, par des aveugles ou certains handicapés;
- elle n'est pas affectée par l'apesanteur et est peu sensible aux fortes accélérations et autres contraintes mécaniques;
- les systèmes de reconnaissance de la parole peuvent servir à vérifier l'identité du locuteur, mais par contre, sont sensibles aux dialectes et aux différences de prononciation;
- ils permettent de surveiller un environnement acoustique, mais sont sensibles aux bruits et aux distorsions;
- ils ne nécessitent aucun espace supplémentaire pour un écran, tableau de bord ou d'autres appareillages complexes, seul un micro doit être porté en permanence.

c) **Mobilité et liberté pour d'autres activités**

Il s'agit des raisons principales décidant les gens à développer et à utiliser des systèmes de reconnaissance automatique de la parole:

- on peut utiliser ces systèmes à distance et à partir de différentes positions;
- ils permettent l'utilisation simultanée des mains, des yeux et de la parole à des tâches différentes;
- le téléphone peut servir de terminal d'ordinateur, à condition que les problèmes posés par une grande population de locuteurs, par le bruit et les distorsions soient résolus.

1.3. Typologie des systèmes

La classification proposée dans [LEA-80-a] reflète bien l'évolution des systèmes de reconnaissance de la parole, évolution caractérisée par une complexité croissante de la tâche de reconnaissance des systèmes dans le temps. On distingue:

- **La reconnaissance de mots isolés**

Dans ces systèmes, les moins complexes et les plus commercialisés actuellement, le locuteur ne peut pas parler de manière naturelle: il doit marquer une pause après chaque mot prononcé, de façon à bien délimiter les frontières entre les mots composant la phrase (la notion de mot ne doit pas nécessairement correspondre à la notion de mot au sens courant, mais éventuellement à une expression comprenant plusieurs mots). L'approche de résolution la mieux adaptée à ce type de reconnaissance est la méthode globale, à condition qu'on n'accepte qu'un vocabulaire restreint (Nous reviendrons ultérieurement sur cette approche).

- **La reconnaissance continue de séquences de nombres ou de mots strictement formatées**

Dans ce cas, le locuteur peut parler d'une manière plus naturelle puisqu'il ne doit plus marquer d'arrêt entre les mots, mais il se trouve encore fortement limité par une syntaxe et un vocabulaire réduits (on parle de langages artificiels entièrement définis par leur syntaxe). L'imposition de ces contraintes permet de limiter les problèmes posés par la parole continue. Moyennant quelques adaptations, la méthode globale peut encore être utilisée pour ce type de reconnaissance: la recherche de la comparaison dynamique parmi toutes les séquences de mots possibles est limitée en fonction des contraintes syntaxiques [HATON-84-b] et [DIMARTINO-84].

- **La reconnaissance continue par mots-clés**

Ces systèmes recherchent dans l'énoncé vocal certains mots-clés leur permettant de reconnaître suffisamment d'éléments dans l'énoncé pour en deviner le contenu. Cette approche peut être utilisée dans la surveillance de conversations afin d'en présélectionner les plus intéressantes.

- **La compréhension restreinte de phrases**

Ces systèmes reconnaissent des énoncés prononcés d'une manière naturelle dans le cadre d'une application donnée: ils travaillent sur une syntaxe proche des langues naturelles, avec des limitations imposées par le contexte de l'application sur le vocabulaire (on parle de langages pseudo-naturels, c'est-à-dire des sous-ensembles de langues naturelles). La méthode de résolution globale n'est plus applicable dans ces systèmes, la compréhension nécessitant la coopération de différentes sources d'informations (syntaxe, sémantique, lexicque etc) et impliquant ainsi une démarche analytique (Nous reviendrons ultérieurement sur cette méthode).

- **La reconnaissance continue et autonome de la parole**

Ces systèmes devraient être capables de reconnaître n'importe quel énoncé oral sans aucun recours à des connaissances syntaxiques, sémantiques ou

pragmatiques et devraient pouvoir être appliqués, sans trop d'ajustements ou de modifications, à des applications diverses. Dans l'état actuel des choses, leur réalisation est considérée comme très difficile ou même impossible par beaucoup de chercheurs.

Des implémentations de ces différents types de systèmes de reconnaissance de la parole, on peut dégager deux approches fondamentales [HATON-84-b], [PIERREL-81] et [BENNANI-84]:

- **La méthode globale**

La forme inconnue à reconnaître est considérée comme une entité globale, indivisible. Dans une phase préalable d'apprentissage, les formes de référence correspondant aux différents mots du vocabulaire sont mémorisées. Le problème de reconnaissance consiste alors à comparer la forme à reconnaître aux différentes formes de référence stockées et à sélectionner celle qui ressemble le plus à la forme inconnue au vu d'un certain nombre de critères de mesure de distances. Cette approche globale de reconnaissance de la parole a l'avantage d'être plus facile à mettre en oeuvre que la méthode analytique. En effet, le traitement acoustique du signal est simplifié par les pauses entre les mots, permettant une segmentation triviale en mots. De plus, la programmation dynamique a permis de résoudre le problème des distorsions dues aux variations de vitesse de prononciation du locuteur (en procédant à une normalisation dans le temps), de sorte que cette méthode offre de bons résultats pour des vocabulaires limités (pas plus d'une centaine de mots) même s'ils contiennent quelques mots phonétiquement proches tels que 'capot', 'cabot', 'cadeau', 'caveau' ou 'cachot' [WELLEKENS-83], [SAMBUR-80], [LEVINSON-81] et [DIMARTINO-84]. Ces restrictions sur le vocabulaire sont nécessaires afin d'éviter des problèmes d'espace mémoire (il faut en effet garder en mémoire les différentes représentations possibles de tous les mots) et des dégradations de performance des algorithmes de comparaison. Etant donné le problème d'espace mémoire, cette méthode est plutôt adaptée à des systèmes mono-locuteur, puisque, pour obtenir de bons résultats en mode multi-locuteur, chaque utilisateur devra disposer de son propre lexique de référence. Enfin, son principal inconvénient est son inadaptation au discours continu puisqu'elle nécessite l'élocution mot à mot. Néanmoins des recherches sont en cours pour étendre le principe de reconnaissance globale au traitement de mots enchaînés, c'est-à-dire des mots prononcés continûment mais sans lien logique (syntaxique), par exemple des suites de chiffres.

- **La méthode analytique**

Cette approche a pour objectif de dégager dans un premier temps des éléments minimaux ou sons pour ensuite reconstruire la phrase de départ comme séquence de ces sons élémentaires afin de comprendre et d'interpréter le message. A cet

effet, le processus de traitement fait intervenir plusieurs sources d'informations, en majorité d'origine linguistique:

acoustiques: pour la saisie du signal vocal, son prétraitement et l'extraction des paramètres pertinents. On distingue en général trois grands types de méthodes d'analyse du signal:

- méthodes spectrales: transformées de Fourier (F.F.T.), bancs de filtres, analyses cepstrales [DIMARTINO-84];
- méthodes par filtrage inverse: codage par prédiction linéaire (LPC) [DIMARTINO-84] et [BOITE-83];
- méthodes temporelles: nombre de passages par zéro du signal ou de sa dérivée [FAWE-83].

Citons quelques paramètres pertinents de la production vocale:

- la hauteur et la mélodie de la voix
(la hauteur de la voix est la fréquence de vibration des cordes vocales et son évolution correspond à la mélodie);
- l'intensité, c'est-à-dire l'amplitude des vibrations sonores;
- le rythme qu'on peut étudier en analysant l'évolution en fonction du temps, de l'énergie totale ou de l'énergie dans une bande de fréquence bien choisie;
- les caractéristiques fréquentielles des sons;

phonétiques: afin de passer du signal paramétrisé à sa description en termes d'unités phonétiques, la phonétique étudiant les sons du langage d'un point de vue purement physique;

phonologiques: complémentaires aux connaissances phonétiques, celles-ci rendent compte des variations individuelles (accent etc) et des phénomènes d'altération contextuelle des sons (phénomènes de co-articulation, de liaison etc), la phonologie analysant l'aspect phonique des langues naturelles du point de vue des fonctions linguistiques;

prosodiques: pour tenir compte du rythme, de l'intensité et de la mélodie de la voix. Ces informations sont spécifiques à la communication orale et peuvent influencer sur la compréhension d'un message;

lexicales: liées aux mots, elles servent à regrouper les sons respectant les règles phonologiques;

syntaxiques: concernant la structuration d'un énoncé en fonction des règles grammaticales du langage, elles permettent de formuler ou de valider des hypothèses sur la présence de mots dans l'énoncé à reconnaître. Ainsi l'analyse syntaxique peut lever certaines ambiguïtés (choix entre deux mots candidats) et minimiser le

temps de recherche lexicale par l'émission d'hypothèses;

sémantiques: liées à la signification des mots, ces connaissances sont utilisées pour dégager un sens de la phrase reconnue en l'interprétant à partir du sens de chaque mot. L'analyse sémantique peut éventuellement signaler des erreurs dans la construction de la phrase, notamment en cas de présence simultanée de mots s'excluant mutuellement;

pragmatiques: plus spécifiques que les autres, elles contiennent les informations relatives à l'univers de l'application, en liaison avec la notion de dialogue.

Pour des définitions plus détaillées de la phonologie, prosodie, phonétique, syntaxe, sémantique, pragmatique et du lexique, nous renvoyons le lecteur à [PIERREL-81] et [MOUSEL-83].

Les deux étapes de la méthode analytique, détection des unités élémentaires (l'analyse acoustico-phonétique produit un treillis de phonèmes) et reconstruction de la phrase (à partir du treillis de phonèmes) pour comprendre et interpréter le message, peuvent s'exécuter:

séquentiellement: dans ce cas, lors de l'analyse acoustico-phonétique, il faudra conserver un grand nombre d'hypothèses puisqu'on ne peut pas revenir dessus;

simultanément: chacun des deux niveaux contribue alors à la reconnaissance en proposant des hypothèses ou en validant ou infirmant des hypothèses émises par l'autre niveau. Cette approche semble préférable puisque la transcription du signal vocal en treillis phonémique comporte toujours des erreurs dues aux difficultés d'interpréter des phrases mal prononcées et aux limites des algorithmes du niveau acoustique.

La méthode analytique, du fait qu'elle a recours à autant de sources d'informations différentes, est a priori la seule adaptée au discours continu et au traitement multi-locuteur. Des systèmes basés sur ce principe ont été réalisés, mais leur temps de calcul dépasse le temps réel et les scores de reconnaissance sont encore inférieurs à ceux de la reconnaissance globale [BENNANI-84]. De plus, il subsiste encore des problèmes de mise en oeuvre pratique, spécialement en ce qui concerne la représentation et l'utilisation des différentes sources d'informations (par exemple les informations prosodiques). Dans [PIERREL-81] et [MOUSEL-83] on peut trouver un aperçu sur les différentes représentations et stratégies envisagées jusqu'ici. Enfin dans [HATON-84-b] on signale qu'il reste encore à résoudre le problème important de l'auto-apprentissage et de l'adaptabilité (à un nouveau locuteur, un nouvel environnement, un nouveau langage, un nouveau domaine d'application etc) des systèmes de compréhension de la parole, peu de travaux ayant vraiment débouché dans ce domaine.

2. Synthèse de la parole

2.1. Quelques généralités

Le but de la synthèse de la parole est de simuler à l'aide de dispositifs électroniques la production de mots ou de sons par l'appareil phonatoire humain. Elle présente sensiblement les mêmes intérêts que l'utilisation de la parole comme moyen d'accès à un système informatique, c'est-à-dire qu'elle rend la communication homme-machine plus efficace, qu'elle est compatible avec des circonstances inhabituelles et qu'elle augmente la mobilité et la liberté, pour d'autres activités, de l'utilisateur. La synthèse se fait à partir d'éléments primitifs (phrases, mots ou phonèmes) décrivant tout ou une partie du message vocal à émettre.

2.2. Typologie des systèmes

On peut classer les différentes approches selon une mise en oeuvre de plus en plus complexe:

- La synthèse par phrases

Elle consiste en une restitution des messages qui ont été saisis, numérisés, comprimés et stockés en mémoire. La qualité du son restitué dépend de l'analyseur utilisé pour coder le signal vocal. Une qualité presque parfaite, conservant l'aspect naturel de la voix du point de vue du rythme, de l'intonation et de l'articulation, nécessite un débit binaire de 64000 bits par seconde de parole. Or, ce débit exige, outre une grande capacité mémoire pour le stockage de la parole, un canal de transmission particulier. C'est pourquoi on s'efforce de réduire ce débit élevé à des taux variant entre 2400 bps et 9600 bps, débit acceptable pour une ligne téléphonique ou une ligne de transmission de données, moyennant des techniques d'analyse complexes conservant plus ou moins bien la qualité de la voix.

Pour plus de détails, nous renvoyons le lecteur à [BOITE-83], [ADANT-83], [FAWE-83] et [VANLIEFFERINGE-83]. Dans [SAMBUR-82] on peut trouver une liste donnant les débits de transmission minimums nécessaires aux diverses techniques d'analyse et de compression de la parole, pour garantir une qualité:

- téléphonique: c'est-à-dire équivalente à celle offerte par le téléphone analogique. Les débits varient entre 16 Kbps à 56 Kbps;
- communicative: c'est-à-dire une voix intelligible et naturelle, mais cependant avec une dégradation sensible. Les débits vont de 7,2 Kbps à 36 Kbps;
- synthétique: c'est-à-dire une voix encore intelligible mais dont le locuteur ne peut plus être identifié et dont le son est quelque peu dénaturé. Le débit minimal est alors de 2,4 Kbps.

Sambur y souligne aussi que ces techniques de compression du signal vocal sont très sensibles aux bruits environnants et aux erreurs de transmission. Dans [ASEO-84], on laisse entendre que la réduction du taux d'échantillonnage jusqu'à 24 Kbps est assez aisée par l'élimination des pauses entre les mots, mais, qu'en-deça des 24 Kbps, les modèles mathématiques ne sont plus applicables. On a alors recours à des processeurs de signaux dédiés ou à des algorithmes spéciaux qui comptent sur les facultés de l'oreille humaine pour comprendre, malgré certaines lacunes, la parole restituée (l'humain est notamment capable de comprendre une personne en dépit de la mauvaise qualité d'une ligne téléphonique ou malgré son accent). Le fait que l'oreille humaine est incapable de différencier entre sons de même fréquence, mais déphasés, peut également être d'une grande utilité. Plus récemment, certains algorithmes, basés sur des heuristiques qui s'inspirent des résultats d'analyse de milliers d'heures d'enregistrements vocaux (ces études avaient pour but de mesurer l'influence de facteurs tels que le sexe, l'âge ou l'accent), ont été développés pour pallier aux insuffisances de la parole restituée.

La synthèse par phrases est utilisable dans le cas où le nombre total de messages à émettre est restreint ou si la qualité du son à produire est un facteur important. Sa limite principale est l'encombrement en mémoire, les messages étant saisis globalement. On ne peut donc pas réutiliser des éléments de phrases (des 'mots' au sens large) pour construire de nouveaux messages.

- **La synthèse par mots ou éléments de phrases**

Dans ce cas, la machine dispose d'une représentation phonétique de chaque mot ou suite de mots connus: c'est le lexique de référence. La restitution d'un message consiste alors à concaténer, à la suite l'une de l'autre, les réalisations relatives aux différents éléments de base composant le message. Une prononciation mot à mot du message peut être évitée par un traitement spécifique résolvant le problème des liaisons entre les mots.

Cette méthode de synthèse, plus complexe que la précédente, nécessite moins de place mémoire pour stocker les messages oraux puisque les éléments de base communs à plusieurs messages ne sont mémorisés qu'une seule fois. Cependant, la qualité du son restitué est comparable à celle d'un locuteur prononçant le message mot à mot: la phrase ne possède pas de mélodie globale.

- **La synthèse par phonèmes**

Cette méthode de synthèse s'impose lorsqu'il s'agit de produire n'importe quel message vocal. La représentation d'un phonème variant en fonction de son contexte et de sa position dans le mot ou le message, la machine doit disposer, en plus d'une référence pour tous les phonèmes d'une langue (environ 37 phonèmes pour le français et 40 pour l'anglais), de règles lui permettant de tenir compte de ces variations intra- et extra-phonémiques.

Enfin, dans le cadre de la synthèse phonétique, on pourrait encore mentionner le cas particulier de la *synthèse à partir d'un texte*. Ce type de synthèse nécessite, à part les informations lexicales et phonologiques, des informations syntaxiques, sémantiques et pragmatiques pour générer, à partir d'un texte écrit donné, sa transcription en une chaîne de phonèmes qu'il s'agit ensuite de synthétiser [BERLEUR-83]. Ces informations supplémentaires devraient permettre de résoudre les problèmes posés par des énoncés du genre: les poules du couvent couvent ou: les fils à soie et les fils à papa.

3. Les applications

3.1. Les applications de la reconnaissance de la parole

En ce qui concerne l'utilisation potentielle de la reconnaissance automatique de la parole dans la vie pratique, trois domaines principaux se profilent [LEA-80-a], les domaines:

militaire, la plus grande partie des fonds investis dans ces recherches sur la parole provenant de sources militaires. Entre autres, on envisage des systèmes:

- de cartographie;
- d'entraînement de contrôleurs aériens;
- de communication dans les postes de pilotage;
- de surveillance de conversations;
- d'aide aux officiers supérieurs pour l'accès aux ordinateurs.

humanitaire, par le développement de systèmes d'enseignement assisté par ordinateur, de systèmes d'aide aux handicapés ou autres personnes défavorisés. Citons par exemple:

- le développement de chaises roulantes à commande vocale;
- un système hospitalier proposé par Dialog Systems Inc., qui permet aux malades immobilisés de contrôler par commande vocale les mouvements de lit, de climatisation, les appels aux infirmières, la télévision etc;
- SIRENE, un système développé au CRIN pour la rééducation vocale des enfants non-entendants [HATON-84-c].

commercial ou économique,

l'utilisation des systèmes de reconnaissance automatique de la parole, surtout de mots isolés, ayant eu comme effet, dans certains cas, une réduction des besoins en personnel de 50% à 90% et des gains de temps de l'ordre de 30% à 95% selon les tâches à effectuer. Les applications commerciales sont assez diverses:

- systèmes de contrôle d'environnement, du même type que le système DIALOG;
- systèmes de contrôle et de programmation par commande numérique vocale de machines puissantes;
- systèmes de tri pour les PTT ou les compagnies aériennes;
- standards téléphoniques;
- systèmes de contrôle de qualité et d'inspection, permettant au contrôleur de vérifier en même temps la qualité d'une pièce et d'en enregistrer les défauts;
- systèmes de home-banking par téléphone;
- systèmes de sécurité et de contrôle d'accès.

3.2. Les applications de la synthèse de la parole

En tant que complément ou même substitut de l'information visuelle, la production automatique de la parole peut convenir à des applications diverses [VANLIEFFERINGE-83]:

- les annonces publiques: Elle peut servir dans le domaine des transports publics: annonces à la fois sur l'aire d'attente (quai, station, aéroport etc) et dans le moyen de transport (train, bus, avion, bateau etc), dans le domaine des distributeurs automatiques (Bancontact, Mister Cash etc): guidage de la manipulation de ces automates et dans les bâtiments publics: renseignements divers;
- la transmission par câble téléphonique d'informations vocales relatives à différents domaines: bancaire (état de compte, taux de change, marché des actions etc), renseignements généraux (météo, lotto, horloge parlante etc), PTT (coût des communications, répondeur automatique pour personnes absentes, changement de numéro d'appel etc) ou contrôle à distance (surveillance de consignes, alarme etc);
- le contrôle de processus (processus industriel, robotique, simulateur, appareil de mesure etc);
- l'informatique (aide à l'utilisation de la CAO, télématique et bureautique);
- les jeux vidéos;
- l'aide aux handicapés (illettrés, aveugles etc);

CHAPITRE 3

COMMUNICATION HOMME-MACHINE

C'est vers le début des années soixante, avec l'apparition des premiers systèmes informatiques interactifs capables de répondre en temps réel aux sollicitations des utilisateurs, que le concept de dialogue homme-machine a vu le jour. Cependant, les coûts du matériel informatique étant de loin supérieurs à ceux du personnel, on avait tendance à optimiser l'utilisation du temps CPU et des capacités mémoire de la machine au détriment de l'interface avec l'utilisateur: le dialogue était fortement dirigé par la machine, c'est-à-dire se faisait en des termes compréhensibles seulement par des personnes spécialisées. C'est pourquoi, jusqu'au début des années soixante-dix, beaucoup d'utilisateurs confiaient le traitement de leurs applications à de grands centres de calcul. L'introduction de l'informatique dans le domaine public, industriel et académique ayant augmenté le nombre d'utilisateurs non-spécialistes, et le développement de mini- et micro-ordinateurs ayant inversé le rapport des coûts entre matériel et personnel, le besoin en interfaces allant à la rencontre des désirs et besoins des utilisateurs se fait de plus en plus sentir. Actuellement il s'agit en effet de concevoir des systèmes permettant d'optimiser le temps de l'utilisateur et non plus le temps de l'ordinateur [MAGUIRE-82]. Certains auteurs estiment même que le succès des nouveaux produits informatiques (surtout en bureautique) ne dépend plus uniquement de leur fonctionnalité mais aussi de leur facilité d'utilisation et de leur acceptabilité par les futurs utilisateurs, ces deux facteurs déterminant les gains en efficacité apportés par le système [ROHLFS-81]. Cette tendance est renforcée en outre par le fait que l'acceptabilité des systèmes est devenue un sujet de recherches interdisciplinaires (faisant appel à des méthodes technologiques, organisationnelles et socio-psychologiques) ayant comme objet d'empêcher des développements technologiques qui ne respecteraient pas les besoins des utilisateurs [HELMREICH-81].

Nous verrons maintenant quelques notions générales sur ce qu'on appelle couramment 'dialogue': son importance en communication orale ou écrite et les composants indispensables à une bonne interaction seront mis en évidence avant la présentation de points critiques en conception de dialogues. Enfin, nous passerons en revue certains types de dialogues.

1. Généralités

Il semble que le dialogue soit essentiel pour la communication orale entre humains. Sans cette forme de protocole elle se réduit à une simple transmission

d'informations peu satisfaisante pour l'homme. Le dialogue a en effet plusieurs rôles [PIERREL-81]:

- il établit un *contact* entre les différents interlocuteurs de sorte que quand quelqu'un parle, il est sûr d'être écouté;
- il permet la *confirmation* ou l'*infirmité* montrant au locuteur si oui ou non il a été compris;
- il permet d'*obtenir des précisions* sur ce qui vient de se dire.

Etant donné qu'on cherche à développer des systèmes informatiques offrant une communication homme-machine ressemblant autant que possible au modèle humain, il s'agit d'élaborer des procédures de dialogue réalisant ces trois fonctions de base. Cependant, vu l'état actuel des recherches en intelligence artificielle, il faut être conscient de ce que l'interaction homme-machine ne sera pas aussi riche en informations que celle entre êtres humains: outre la parole, la communication orale humaine fait intervenir tout un langage qualifié de non verbal (regard, expression du visage, mouvements de la tête, des mains et du corps, bruits tels que toussotements, soupirs ou râclements de gorge etc) alors que celle entre l'homme et la machine se restreint le plus souvent à un seul canal (la parole ou le texte) [PIERREL-81].

Offrir un interface en langage plus ou moins naturel n'est généralement pas suffisant pour permettre une interaction agréable avec une machine: Ball, Hayes et Reddy [BALL-81] et [HAYES-83], reprochent aux systèmes actuels offrant de tels interfaces, qu'ils ne savent habituellement, dans le cadre de leur domaine d'expertise, répondre de manière exacte et appropriée qu'à un ensemble restreint de sollicitations. La majorité d'entre eux seraient notamment incapables de répondre raisonnablement à des questions ne respectant pas une grammaire rigide, de demander et de comprendre des explications si la sollicitation était ambiguë et d'expliquer leurs propres réponses à la demande de l'utilisateur. D'après eux, l'interface en langage écrit ou parlé apparaît d'autant plus coopératif et accessible à des utilisateurs non expérimentés que l'interaction est gracieuse. Pour cela, il faut pouvoir traiter convenablement n'importe lequel de leurs messages. Partant des deux hypothèses suivantes:

- l'homme a tout un ensemble de conventions et de besoins lorsqu'il dialogue avec autrui,
- la communication avec la machine est difficile parce qu'elle ne satisfait pas ces exigences,

Ball, Hayes et Reddy ont essayé d'identifier ces besoins et conventions de la communication humaine et de trouver, pour les interfaces, des moyens de les satisfaire, qui ne correspondent pas nécessairement à ceux des humains. Il ont ainsi établi une liste de critères relativement indépendants, permettant de définir certaines qualités indispensables à tout système interactif convivial mais probablement non suffisantes pour garantir une interaction agréable dans le cas général:

- **La robustesse de la communication**

Le système doit être capable d'indiquer à l'utilisateur si oui ou non il l'a compris. Afin d'éviter un allourdissement du dialogue par des confirmations explicites, le système peut procéder à des interprétations, mais il doit en informer son interlocuteur de façon à lui permettre d'éventuelles rectifications. De plus le système doit prendre garde à ce que l'utilisateur le comprenne et corriger les malentendus qu'il détecte. Les techniques envisagées pour mener à bien le dialogue sont inspirées du modèle humain: confirmation implicite, indication explicite d'incompréhension ou de reconnaissance partielle, répétition directe, indirecte ou paraphrasée, répétition dans le but de gagner du temps (par exemple lors de la recherche d'un renseignement).

- **La flexibilité de l'analyse des énoncés en entrée**

Le système devrait pouvoir traiter les messages de l'utilisateur même s'ils présentent certaines malformations par rapport aux normes grammaticales, malformations apparaissant très fréquemment en langue parlée mais aussi de temps en temps en langage écrit:

- les idiomes (énoncés qu'on ne peut pas interpréter à partir de leurs composants, mais qui, ayant un sens particulier doivent être interprétés globalement);
- les énoncés fragmentés (énoncés déformés par des bruits);
- les erreurs grammaticales (temps, déclinaisons ou prépositions incorrects, mots répétés ou omis);
- les ellipses (énoncés incompréhensibles ou ambigus hors du contexte du dialogue);
- etc.

Quant aux techniques d'analyse envisagées, la *stratégie descendante et de la gauche vers la droite* est considérée comme mal adaptée tandis que les méthodes *d'analyse conceptuelles* (celles-ci utilisent des contraintes sémantiques pour faire entrer la forme inconnue dans une structure conceptuelle prédéfinie), *d'analyse basées sur la reconnaissance de formes* et *d'analyse spécifiques aux constructions* (celles-ci font appel à plusieurs stratégies d'analyse dont chacune renferme des connaissances spécifiques sur le genre d'erreurs pouvant se présenter dans les différents types de constructions) paraissent plus prometteuses.

- **La capacité de fournir des explications sur soi-même**

Afin d'apparaître coopératif à l'utilisateur, le système doit pouvoir répondre à d'éventuelles questions portant sur:

- * son domaine d'application (ce qu'il sait et ne sait pas faire),
- * la forme d'une requête,
- * les actions entreprises et les motifs sous-jacents,
- * les résultats d'actions antérieures.

- **La capacité de gérer un historique du dialogue**

Le dialogue étant en général soumis à une certaine *finalité* et centré sur un *sujet* donné, le système en tant que participant au dialogue doit être capable de les déterminer et d'en garder une trace pour savoir faire face à des changements de contexte dus à des conversations divagantes. Cet historique sert également à résoudre des ellipses et des anaphores.

- **La capacité d'identifier un objet à partir d'une description**

Comprendre les requêtes d'un utilisateur exige du système la capacité d'identifier des objets qu'il connaît, à partir de la description fournie par l'utilisateur. Par exemple, dans le cadre d'un standard téléphonique automatique, lorsque l'on demande le numéro d'une certaine personne, il faut pouvoir identifier cet abonné sur base de la description fournie (nom, prénom, adresse, profession etc). Les informations données dans un premier temps peuvent ne pas être suffisantes pour reconnaître sans ambiguïté l'objet référencé. Dans ce cas, le système doit être capable de mener un dialogue clarifiant la requête, soit en proposant un choix entre les objets répondant à la description, soit en demandant des précisions si la description était tellement vague qu'il y a trop d'alternatives, sans pour autant perdre de vue le contexte général du dialogue.

- **La génération de réponses en langage 'naturel'**

L'homme peut compenser dans une certaine mesure les déficiences du langage en sortie du système, néanmoins, ce dernier lui apparaîtra d'autant plus coopérant et gracieux qu'il produira des réponses appropriées dans un langage proche du sien: le système pourra disposer d'un modèle de l'utilisateur, en particulier de son expérience, afin d'y adapter les messages et les explications.

Selon Ball, Hayes et Reddy, certains systèmes actuels possèdent quelques-unes de ces qualités, qualités qui théoriquement seraient d'ailleurs à la portée de la technologie actuelle, mais, à leur connaissance, il n'y aurait pas de système qui les réunirait toutes à la fois. Ils soulignent en outre que celles-ci, insuffisantes pour garantir une interaction agréable dans le cas général, forment cependant une bonne base de travail pour construire bon nombre de systèmes interactifs, particulièrement dans le domaine des *services simples* tels que: standard téléphonique, services de réservation (restaurants, compagnies aériennes etc), centres de renseignement (météo, valeurs boursières etc), messagerie électronique etc. Ces services sont qualifiés de simples parce qu'en principe le système n'exige de l'usager que l'identification d'un nombre limité d'entités qui sont des paramètres du service à rendre. Par exemple pour une réservation de table dans un restaurant les paramètres principaux à identifier seraient: le nombre de personnes, la date et l'heure, pour un système de messagerie électronique: l'émetteur, le destinataire et le corps (non analysé) du message.

D'après Spérandio et Scapin, [SCAPIN-85], un logiciel est un outil dont, par conséquent, l'efficacité doit être évaluée sur base de l'aide qu'il apporte aux

utilisateurs pour atteindre leurs objectifs plutôt que sur base de la vitesse ou du degré de sophistication. De ce fait, avant la conception proprement dite d'un logiciel, il faut identifier et analyser les caractéristiques des futurs usagers du système (leurs façons de procéder, besoins, contraintes, capacités d'adaptation etc) ainsi que les spécificités de la tâche (sa complexité, fréquence d'exécution etc). En effet, la forme du dialogue à concevoir dépend fortement:

- du degré de l'initiative accordée à l'opérateur,
- du niveau de la flexibilité offerte,
- du mode d'interaction choisi: multi-mode ou non;
 - textuelle, graphique ou parlée,
- du langage de communication défini: vocabulaire plus ou moins adapté aux besoins des usagers; langage de commande rigide ou interface en langage pseudo-naturel,
- du niveau de détail adopté pour les messages d'aide ou d'erreur et pour les explications.

(Dans [DEAN-82], on peut trouver un exposé détaillé sur la manière de concevoir des messages adaptés à la situation et au récepteur.)

Cependant, prévoir les besoins des usagers potentiels n'est pas suffisant: les utilisateurs évoluant avec l'utilisation du système, en particulier leur niveau d'expérience augmentant, il faut également prendre en compte l'évolution de ces besoins. Il s'avère donc nécessaire d'étudier dès le départ plusieurs niveaux d'interface convenant à des niveaux différents d'expérience [SCAPIN-85] et [MAGUIRE-82]. Une façon de réaliser ceci serait de définir un ensemble de permissions et d'options par défaut pour chaque utilisateur [ROHLFS-81]: tandis que les permissions ne peuvent être changées que par une personne désignée par l'organisation (cette personne assure le contrôle organisationnel du système en gérant notamment la quantité d'espace allouée aux divers usagers), les options par défaut sont modifiables à volonté par chaque utilisateur, lui permettant de choisir le mode de communication qui lui convient le mieux.

2. Typologie des dialogues

Lorsqu'on veut classifier les dialogues, une première distinction peut se faire entre *dialogues libres* et *dialogues dirigés*, les dialogues dirigés étant les seuls à pouvoir être automatisés. Cependant, dans un souci de confort d'utilisation d'un système informatique, ces dialogues sont conçus et réalisés de manière à minimiser les aspects contraignants [PIERREL-81].

Les dialogues libres peuvent être de deux types [PIERREL-81]:

- **Les dialogues enfantins**

Leurs caractéristiques essentielles sont une structure syntaxique relativement pauvre (phrases se limitant le plus souvent à un sujet, un verbe et un complément) qui néanmoins peut couvrir un vaste champ sémantique, et un vocabulaire assez limité mais cependant varié.

- **Les dialogues courants**

Outre une structure syntaxique plus riche, ceux-ci sont caractérisés par un vocabulaire étendu, une grande liberté par rapport à la syntaxe et des champs sémantiques très divers permettant de changer aisément de sujet au cours d'un même dialogue.

Quant aux dialogues dirigés (automatisables), on peut distinguer différents types selon leur degré de complexité et leur niveau de flexibilité [PIERREL-81], [SCAPIN-85] et [PIERREL-85]:

- **Le mode question-réponse**

Simple mais très rigide, ce type de dialogue convient aux utilisateurs occasionnels pour la saisie de données bien connues et se présentant toujours dans le même ordre. En effet, il provoque peu d'erreurs, la structure du dialogue étant entièrement imposée par la machine. Cependant, il ne permet pas d'avoir de vue globale sur l'ensemble des alternatives et il devient vite contrariant et parfois même agaçant pour les personnes expérimentées.

- **Le mode remplissage de formulaire (form-filling)**

Ce mode devient préférable lorsque l'entrée comporte des valeurs de paramètres plutôt que des commandes.

- **Le mode de sélection (par menus)**

Ce type d'interface, facilement réalisable, offre à l'utilisateur une information complète sur les options locales à un niveau. Outre une organisation cohérente des menus (un choix par niveau, présentation uniforme des alternatives, vocabulaire cohérent etc), il exige toutefois un dialogue dirigé par la machine et la disponibilité d'un écran pour présenter les menus. Afin d'éviter que les menus ne deviennent trop détaillés pour une personne non habituée au système, on prévoit souvent des menus 'cachés' qui répondent directement aux requêtes des utilisateurs et ne présentent les menus sous-jacents qu'après une demande d'aide explicite.

- **Le mode touches de fonction**

L'utilisation de touches de fonction s'avère utile et efficace en cas de tâches qui sont exécutées assez fréquemment et requièrent peu de commandes différentes: Le nombre de fonctions étant restreint, leur apprentissage est relativement aisé et le risque d'erreur très réduit.

- **Le mode langage de commande**

Les langages de commande (y compris les langages de programmation) constituent le moyen de communication principal entre l'homme et la machine: complexes mais cohérents, ils offrent aux utilisateurs entraînés plusieurs avantages tels que l'adaptabilité aux matériels et aux diverses applications permettant ainsi une bonne utilisation des ressources et une exécution efficace des tâches, l'usage international grâce à leur standardisation etc. Leur principal inconvénient réside dans la nécessité d'un entraînement assez fastidieux et d'une adaptation du raisonnement humain à l'algorithmique. En outre, ces langages impliquent un dialogue homme-machine qui est le plus souvent restreint au clavier et à l'écran.

- **Le mode dialogue expert**

D'une réalisation beaucoup plus difficile, ce type de dialogue offre d'avantage de flexibilité à l'utilisateur: Le contrôle de la communication est plus ou moins partagé entre machine et utilisateur, permettant ainsi une interaction plus naturelle et satisfaisante. Ces dialogues font appel à des techniques complexes qui sont en grande partie du domaine de l'intelligence artificielle. Appelés aussi 'dialogues experts' par analogie aux systèmes experts, ils se limitent à un domaine particulier, tout en disposant d'un vocabulaire suffisamment consistant (c'est-à-dire riche de quelques centaines de mots) pour permettre un dialogue naturel sur un sujet précis. La syntaxe des énoncés, assez riche, est le plus souvent respectée et la structure du dialogue est en général assez précise.

CHAPITRE 4

EDITEURS DE SONS

L'adjonction d'entrées-sorties vocales et sonores aux systèmes informatiques offre des possibilités nouvelles aux interfaces futurs: une communication multi-média pourrait améliorer d'une manière significative le dialogue homme-machine. Notamment dans le domaine de l'enseignement assisté par ordinateur (EAO), l'utilisation d'éléments sonores (parole ou musique) pourrait gagner en importance et conduire à de nouvelles applications intéressantes.

Nous allons d'abord montrer ce que sous-entend le terme éditeur de sons: son rôle et son architecture matérielle seront présentés. Ensuite nous verrons un exemple d'un tel outil, le logiciel EDIT réalisé au CRIN pour enrichir un système d'EAO. Ce logiciel servant également de base à notre système de messagerie vocale, nous en exposerons les éléments principaux, surtout les types d'objets qu'il manipule et les opérations définies sur ces objets. Enfin, nous mentionnerons quelques applications possibles de l'édition de sons, à la fois pour des spécialistes étudiant les caractéristiques de signaux sonores et pour des utilisateurs moins spécialisés ayant recours à des objets sonores dans le cadre d'applications diverses en EAO, bureautique ou dans d'autres domaines.

Pour cet exposé, nous nous référons essentiellement à [HATON-84-a] et [BENNANI-84].

1. Généralités

1.1. Rôle de l'éditeur de sons

En toute généralité, un éditeur spécialisé pour un type donné d'objets est un outil informatique constitué d'un ensemble de fonctions qui sont censées permettre une manipulation aisée de ces objets. En particulier, il doit faciliter, à l'utilisateur averti ou non, la création, la restitution, la modification, la concaténation etc des objets.

Un éditeur de sons offre des fonctions de traitement d'objets sonores tels que:

- des éléments de parole continue;
- des mots isolés;
- des sons enregistrés;
- des sons synthétisés.

La nature physique de ces sons peut varier selon l'utilisation prévue et la qualité de son requise. On peut distinguer notamment:

- des portions de signal numérisé (parole ou sons) pour une restitution ultérieure de bonne qualité;
- des portions de signal prétraité pour optimiser le stockage ou en vue d'une analyse;
- des codes permettant la création de sons synthétiques etc.

1.2. L'architecture matérielle

Les opérations d'entrées-sorties sonores nécessitent la présence de certains composants particuliers. Le schéma de la figure 1 présente la configuration nécessaire à l'édition de sons.

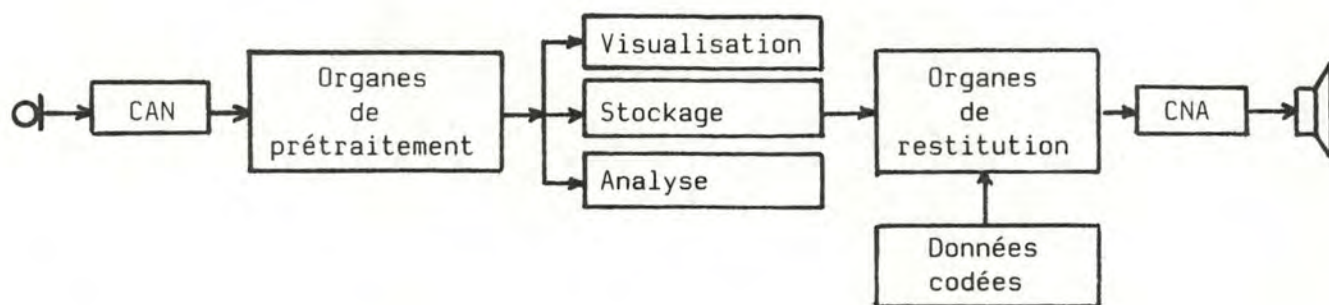


Fig. 1. Environnement matériel de l'éditeur de sons

Un *microphone* fournit un signal analogique à un certain nombre d'*organes de prétraitement* qui en donnent une forme numérique pouvant être visualisée, stockée ou analysée:

Un *convertisseur analogique-numérique* (CAN) chargé d'échantillonner et de numériser le signal. En vertu du principe de Shannon, la fréquence d'échantillonnage doit être au moins le double de la largeur de bande utile du signal [BOITE-83]. La bande de fréquence se situant entre 300 Hz et 3400 Hz en téléphonie, on choisit souvent une fréquence d'échantillonnage de 8 KHz. Chaque échantillon étant codé sur 8 bits, il en résulte un débit d'informations de 64 Kbps (la mémorisation d'une seconde de parole ou de sons musicaux exige

alors 64 Kb). Ce débit convient pour stocker des sons de bonne qualité (cf. 'La synthèse par phrases').

Un *analyseur spectral* donnant une représentation amplitude-fréquence-temps du signal qui peut être visualisée. En outre, l'analyse fréquentielle permet de comprimer le débit d'informations en le réduisant jusqu'à des taux minimums de 2400 bps, débit suffisant si le message ainsi codé est destiné à passer par une phase d'identification: c'est le cas des mots énoncés un par un dans des systèmes de commande ou de réponse vocale.

Un *détecteur de mélodie* fournissant l'évolution de la hauteur de la voix, hauteur et mélodie de la voix étant des paramètres intéressants à étudier dans un signal sonore. Dans la production de la parole, la hauteur de la voix est donnée par la fréquence de vibration des cordes vocales et son évolution correspond à la mélodie de la voix.

A la restitution, un *convertisseur numérique-analogique* (CNA) combiné avec un *haut-parleur* fournit, suivant les besoins, un signal de parole de bonne qualité sonore ou des sons synthétisés à partir de données sur la forme d'onde, les fréquences composantes etc.

2. L'éditeur de sons EDIT

EDIT est un éditeur d'objets de type parole, musique et graphique développé pour enrichir un système d'EAO: il permet aux enseignants de créer des objets de différents types et de les utiliser dans des didacticiels. Nous allons nous limiter ici à la présentation des divers types d'objet et des principales fonctions définies dans EDIT.

2.1. Les types d'objets d'EDIT

L'éditeur EDIT met à la disposition des utilisateurs un ensemble de fonctions lui permettant une manipulation aisée des types d'objet suivants:

- **parole** (noté par la suite: P);
- **texte parlé** (TP), type composé d'objets du type **message élémentaire** (ME);
- **texte musical** (TM), type composé d'objets du type **symbole musical élémentaire** (SM).

L'objet *parole* est considéré comme une seule entité correspondant à un énoncé vocal préalablement enregistré (c'est-à-dire numérisé, codé et comprimé). La séquence de parole ne peut être ni modifiée, ni tronquée: sa restitution doit se faire de façon continue, garantissant ainsi une qualité de son maximale, compte-tenu des contraintes de place-mémoire pour la stocker (la qualité est proportionnelle à la quantité d'informations sauvegardées sur le signal). Ce type d'objet convient à la synthèse par phrases (cf. 'La synthèse par phrases')

L'objet *texte parlé*, par opposition au type parole, n'est pas considéré comme formant un tout indivisible: il est formé par la concaténation de messages élémentaires (le sous-type ME, message élémentaire, correspond à un mot ou un groupe de mots) prononcés par un locuteur unique. L'auteur d'un message de type texte parlé introduit le texte du message à l'aide d'un éditeur de texte vidéo, puis grâce à une commande d'apprentissage, il permet au système d'acquérir une réalisation vocale de tous les messages élémentaires encore inconnus rencontrés dans le texte. La restitution du message sous forme vocale consiste alors à reproduire la suite des messages élémentaires les uns après les autres. Pour le confort de l'écoute, on exige que, pour un même message, tous les messages élémentaires soient prononcés par une seule et même personne. De cette façon, la restitution vocale d'un texte parlé revient à une synthèse par mots ou par éléments de phrases (cf. 'La synthèse par mots ou éléments de phrases').

Comparé au type parole, l'objet texte parlé permet une économie en place-mémoire, un même message élémentaire dont la prononciation orale n'est mémorisée qu'une seule fois par locuteur, pouvant servir à construire plusieurs textes parlés. En outre, il permet l'utilisation de variables qui, étant instanciées au moment de la restitution, permettent de paramétrer les messages. Par exemple en EAO, ces variables peuvent être associées au nom de l'apprenant, rendant ainsi les messages plus personnels.

Notons encore que la définition du type texte-parlé ne constitue en fait qu'une étape dans l'édition de sons: par la suite, la restitution mot à mot d'un message devra être substituée par une synthèse à partir du texte qui ne nécessitera plus aucun apprentissage préalable. Cependant il faut encore attendre que cette technique de synthèse fasse ses preuves et donne lieu à des systèmes d'une intelligibilité acceptable.

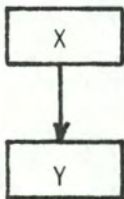
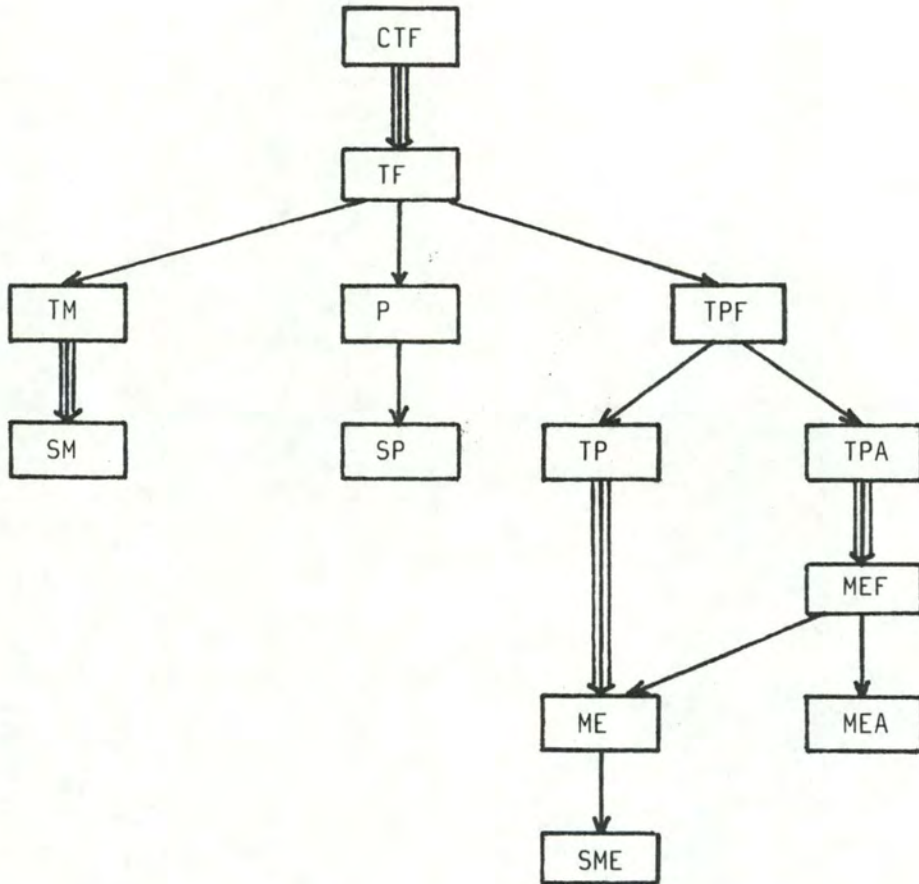
L'objet *texte musical*, d'une manière analogue au texte parlé, est composé d'une suite d'éléments primitifs: les notes musicales (de type symbole musical élémentaire ou SM). Les textes musicaux sont composés en fournissant, à l'aide d'un éditeur de texte vidéo, les symboles correspondant à chaque note du texte musical ou par la transcription de la partition correspondante sur une tablette graphique. La restitution sonore ou graphique d'une partition se fera en reproduisant, les unes après les autres, chacune des notes du texte musical, soit en les synthétisant à partir de leurs caractéristiques (fréquence, durée etc), soit en les visualisant sous forme symbolique sur un écran graphique.

Dans sa thèse M. Bennani propose une spécification formelle des types d'objets en termes de types abstraits. Cette spécification a l'avantage de conduire à une modularité des données et des programmes: les données sont structurées en types d'objets définis par les opérations qui leur sont applicables. Nous nous limitons ici à la présentation de la liste des types d'objets définis dans EDIT (tableau de la figure 2) et des relations entre ceux-ci (schéma de la figure 3).

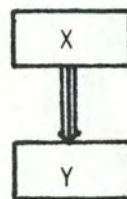
Fig. 2. Liste des types d'objets d'EDIT

TYPE	SPECIFICATION INFORMELLE
CTF	Type catalogue formé d'une liste d'objets de type TF.
TF	Type générique regroupant les objets de type parole (P), texte parlé fictif (TPF) et texte musical (TM).
TM	Type texte musical défini comme une liste de symboles musicaux élémentaires (SM).
SM	Type symbole musical élémentaire.
P	Type parole.
SP	Type signal de parole associé à un objet parole.
TPF	Type texte parlé fictif regroupant les objets de type texte parlé (TP) et de type texte parlé en attente d'apprentissage (TPA).
TP	Type texte parlé composé d'une liste de messages élémentaires connus (ME).
TPA	Type texte parlé en attente d'apprentissage composé de messages élémentaires fictifs (MEF).
MEF	Type message élémentaire fictif soit connu (ME), soit en attente d'apprentissage (MEA).
ME	Type message élémentaire connu.
MEA	Type message élémentaire en attente d'apprentissage.
SME	Type signal de parole associé à un message élémentaire connu.

Fig. 3. Schéma des relations entre les différents types d'objets



un objet Y intervient
dans la composition de X



une liste d'objets Y interviennent
dans la composition de X

2.2. Les fonctions principales d'EDIT

L'ensemble des fonctions définies dans EDIT peut être partitionné en deux classes:

- les fonctions de manipulation d'objets de type parole (P), texte parlé (TP), message élémentaire (ME), texte musical (TM) et symbole musical élémentaire (SM). Le tableau de la figure 4 résume ces fonctions.
- les fonctions utilitaires servant à guider l'utilisateur lors de l'utilisation du logiciel EDIT:

Aide:	fournit des informations complémentaires sur les différents catalogues (P, TP, ME, TM, SM), les différentes opérations possibles sur le catalogue courant, la syntaxe d'une commande etc;
Mode:	permet de changer de mode de travail, de passer du mode 'guidé' ou 'bavard' en mode 'libre' ou 'muet' ou inversement (en mode bavard, EDIT propose pour chaque catalogue un menu des opérations possibles alors qu'en mode muet, l'utilisateur peut directement formuler ses commandes);
Parole:	produit des échos sonores correspondant aux principaux messages affichés sur l'écran;
Réécoute:	reproduit le dernier message vocal restitué;
Menu:	permet de revenir au niveau du dialogue précédent (choix d'un catalogue ou d'une fonction à exécuter);
Suite:	si la machine doit visualiser à l'utilisateur un texte dépassant la taille de la zone d'écran réservée à la visualisation, elle visualise une partie du message et attend l'activation de cette fonction 'suite' pour afficher le reste;
Fin-aide:	permet le retour au point d'appel avant l'activation de la fonction 'aide'.

Fig. 4. Les fonctions d'EDIT

Nom de la fonction	Paramètres	Types d'objets modifiés	Description informelle de la fonction
Apprendre		P, ME	déclenche la saisie au microphone pour apprendre les objets de type P ou ME encore inconnus;
Créer	iden	P, TP, ME, TM, SM	crée l'objet de nom 'iden';
Détruire	iden	P, TP, TM	détruit l'objet 'iden': 'iden' est supprimé définitivement du catalogue;
Ecouter	iden, nb	P, TP, ME, TM, SM	restitue sous forme sonore 'nb' fois l'objet 'iden' s'il est de type ME, TP ou P; on le synthétise 'nb' fois s'il est de type TM ou SM;
Générer		P, TP, ME, TM, SM	réinitialise le catalogue courant: tous les objets contenus dans le catalogue sont détruits;
Imprimer	iden, perif	P, TP, ME, TM, SM	édite sur le périphérique désigné par 'perif' les composantes de l'objet 'iden' dans le catalogue courant;

(Fig. suite)

Nom de la fonction	Paramètres	Types d'objets modifiés	Description informelle de la fonction
Lister	iden1, iden2, périf	P, TP, ME, TM, SM	édite sur le périphérique désigné par 'périf' les noms des objets contenus dans le catalogue courant dont le nom est compris entre 'iden1' et 'iden2';
Prendre	iden	P, TP, ME, TM, SM	si l'objet de nom 'iden' se trouve dans le catalogue courant, il devient l'objet courant;
Renommer	iden1, iden2	P, TP, TM	donne à l'objet 'iden1' le nom 'iden2';
Sauver		P, TP, ME, TM, SM	sauvegarde un objet créé dans le catalogue;
Visualiser	iden, orig	P, ME, TM, SM	visualise sur l'écran graphique l'objet 'iden' à partir du point d'ordonnée 'orig';

3. Applications de l'édition de sons

L'évolution technologique (circuits intégrés spécialisés de traitement du signal ou de gestion graphique, mémoires, écrans 'bit-map' etc) a largement contribué au développement d'applications industrielles de la parole, et plus généralement de sons, dans des domaines variés, de tels systèmes étant désormais implémentables sur micro-ordinateurs. Cette évolution a en même temps conduit à la conception d'éditeurs spécialisés en manipulation d'objets sonores tels que EDIT dont nous venons de décrire les composantes principales. L'aspect conversationnel et interactif de ces éditeurs rend leur utilisation agréable et efficace.

Nous allons voir d'abord l'intérêt de ces éditeurs sonores dans le cadre d'applications professionnelles, puis nous allons nous intéresser aux possibilités d'utilisation pour des non-spécialistes (applications grand public).

3.1. Applications professionnelles

Dans ce domaine, l'éditeur de sons est un outil de travail destiné à trois grands types d'utilisations:

- à la recherche fondamentale et appliquée, en particulier en ce qui concerne la parole: études fines de paramètres et leurs évolutions temporelles (ici la visualisation graphique peut être d'un grand intérêt, le signal vocal ou sonore étant un phénomène variant dans le temps), examen systématique de grandes quantités de données (bases de données ou corpus réalisés spécialement à des fins de recherches), etc;
- à la production (synthèse) de la parole, de sons ou de musique en vue d'une utilisation ultérieure (annonces parlées, jouets, alarmes, automobiles etc);
- à l'analyse de la production vocale, notamment comme aide au diagnostic médical: étude de l'importance de la déviation par rapport à la parole normale, recherche de corrélations entre pathologie et manifestations acoustiques de la parole, détermination de l'orientation à donner à la rééducation vocale, le cas échéant etc.

3.2. Applications grand public potentielles

Dans un avenir plus ou moins proche, les éditeurs de sons pourront offrir des possibilités nouvelles pour:

- les dialogues homme-machine multi-média

L'amélioration de la communication homme-machine est devenue un souci important pour les concepteurs de nouveaux systèmes (cf. 'Communication homme-machine'). Cette tendance se manifeste notamment par l'apparition de systèmes intégrant, outre le dialogue clavier-écran classique, des dialogues utilisant des écrans graphiques avec possibilité de désignation ou des écrans bit-map avec souris. L'adjonction d'entrées-sorties vocales à ces systèmes pourrait leur

permettre de conduire un véritable dialogue multi-média: la parole serait utilisée en complément aux autres modes de communication pour réaliser certaines fonctions spécifiques telles que:

- * en sortie, l'envoi vers l'utilisateur de commentaires liés à une image ou un graphisme présentés en cours de dialogue, de certains messages d'alarme, de rappel à l'ordre (par exemple en cas de réponse trop tardive à un message), d'aide ou d'information (intéressant en cas d'applications faisant appel à un dialogue sous forme d'écran pleine page, car en évitant une réponse visuelle, la structure de l'écran n'est pas perturbée).
- * en entrée, la prise en compte de commandes orales simples (par exemple liées à la gestion de l'écran), la gestion par le système d'un catalogue de doléances formulées par les utilisateurs (plus particulièrement dans le cas de logiciel mis à disposition sur un réseau: ceci rejoint un autre type d'applications possibles, les systèmes informatiques de messagerie).

- **les systèmes informatiques de messagerie**

Etant capables d'acquérir et de stocker de la parole par des techniques de numérisation et de codage, de la restituer par des méthodes de synthèse et de comprendre des commandes vocales simples par des techniques de reconnaissance automatique de la parole, les éditeurs de sons permettent d'aller bien au-delà des répondeurs-enregistreurs actuels: Le stockage séquentiel sur bande magnétique peut être remplacé par la mémorisation de parole numérisée sur supports informatiques. Cette dernière est beaucoup plus souple et, moyennant une gestion de catalogue, permet un accès direct aux différents messages vocaux enregistrés. Nous reviendrons sur cette famille d'applications dans le chapitre suivant en exposant le système MVOC sur lequel nous avons travaillé. Notons encore qu'une intégration plus complète d'un système de messagerie, de télécopieurs et d'éditeur de sons pourra conduire vers une véritable messagerie multi-média qui serait capable d'apporter de nouvelles possibilités intéressantes à la vie d'entreprise.

- **l'enseignement assisté par ordinateur**

L'interactivité entre le didacticiel et l'apprenant étant l'un des principes de base de l'EAO, les éditeurs de sons pourront contribuer à une amélioration de celle-ci par l'introduction d'un dialogue multi-média, en facilitant en particulier l'utilisation de la parole dans la conception et la mise en oeuvre de nouveaux didacticiels. En effet, tout comme le graphisme peut aider à véhiculer certaines notions, la parole peut avoir une influence pédagogique du fait que certains messages seront mieux perçus et plus vite compris par l'apprenant s'ils sont produits sous forme vocale. Ainsi, on peut imaginer de fournir, d'une part, les messages propres au contenu enseigné et, d'autre part, un certain nombre de messages ayant des fonctions différentes:

- * encouragement de l'élève en cas de réponse correcte;
- * rappel à l'ordre, par exemple s'il tarde à répondre;
- * rappel des notions précédemment traitées (ceci pourra conduire vers des didacticiels capables d'aider l'élève dans la recherche de la réponse en lui suggérant un début de solution).

Cependant, une telle utilisation de la parole comme moyen de communication reste, dans le cadre de l'EAO, encore de l'ordre du 'gadget', vu l'état actuel des recherches dans ce domaine. Bien que très intéressants d'un point de vue pédagogique, ces avantages ne peuvent motiver à eux seuls l'adjonction d'entrées-sorties vocales à un système d'EAO.

Toutefois, il existe certaines applications spécifiques nécessitant l'utilisation de la parole pour rendre l'EAO effectif et efficace (dans ce cas, il ne s'agit plus de gadgets):

- * l'enseignement de langues étrangères: l'introduction du traitement automatique de la parole (acquisition, analyse, reconnaissance et restitution de la parole) pourra conduire au développement de didacticiels d'une utilisation souple et permettant à l'apprenant d'évaluer plus objectivement ses productions sonores;
- * l'enseignement de la musique: dictée ou lecture musicale ou plus généralement l'enseignement du solfège;
- * la rééducation vocale, en particulier de jeunes enfants mal- ou non-entendants, grâce à des exercices sous forme de jeux dans lesquels la contre-réaction auditive déficiente est remplacée par un retour visuel et des messages d'évaluation;
- * l'apprentissage ou le réapprentissage de la lecture et de l'écriture: enseignement primaire, jeunes gens ayant connu l'échec scolaire, travailleurs (immigrés le plus souvent) devant participer à de nouvelles formations pour pouvoir affronter les mutations industrielles;
- * l'enseignement professionnel, plus particulièrement dans le cadre de mise en oeuvre de certains simulateurs (par exemple des simulateurs d'entraînement pour pilotes ou contrôleurs aériens): dans ce cas, l'utilisation de la parole s'impose pour pouvoir conserver le caractère naturel de certaines situations de dialogue.

CHAPITRE 5

SYSTEMES INFORMATIQUES DE MESSAGERIE VOCALE

Les progrès en techniques de traitement automatique de la parole (acquisition, stockage, reconnaissance et synthèse) ont largement contribué à l'apparition de nouveaux systèmes de messagerie: les systèmes de messagerie vocale. Accessibles via un téléphone standard, ceux-ci permettent d'améliorer les communications internes et externes des sociétés commerciales.

Dans un premier temps, nous allons voir les possibilités des systèmes de communication par ordinateur, en particulier des systèmes de messagerie vocale, leurs apports à la vie des organisations, leurs limites ainsi que quelques applications concrètes envisagées.

Dans un second temps, nous allons présenter le système MVOC que nous avons développé au CRIN en nous appuyant sur certains outils disponibles (tels que l'éditeur de sons EDIT). Nous décrirons les types d'utilisateurs, les modes d'accès, les types de messages pouvant être envoyés, les fonctions offertes en émission et en consultation ainsi que l'interface utilisateur-machine choisi. Nous exposerons le travail réalisé, travail centré sur la définition de l'organisation générale du système et la conception du dialogue oral homme-machine. Aucune carte d'entrées-sorties vocales n'ayant été disponible sur la SM90 (mini-ordinateur fonctionnant sous le système d'exploitation UNIX et sur lequel le dialogue a été implémenté) l'interaction orale a été simulée par un échange de messages clavier-écran. Nous terminerons avec une évaluation des possibilités actuelles du système MVOC et de ses perspectives d'avenir.

1. Généralités

Les systèmes de messagerie vocale offrent généralement l'une ou l'autre des fonctions suivantes, fonctions d'ailleurs assez semblables à celles qu'on trouve dans les systèmes classiques de messagerie, abstraction faite de l'aspect oral des messages traités dans les nouveaux systèmes:

- enregistrement, stockage, modification et consultation de messages;
- diffusion multiple et simultanée d'un message à un groupe d'abonnés ('broadcasting');
- envoi immédiat ou différé de messages;
- possibilité de répondre à- ou de rediriger un message consulté etc.

Toutefois, ces nouveaux systèmes ont certaines caractéristiques particulières. Ils utilisent la parole qui est le moyen de communication le plus naturel, familier et spontané de l'homme (cf. 'L'intérêt des entrées vocales'). D'où la possibilité d'utiliser un simple combiné téléphonique comme terminal d'accès au système de messagerie électronique. Ces points d'accès pourront alors être multipliés grâce au réseau téléphonique, dont la numérisation progressive devrait permettre une transmission de bonne qualité (comparée à la transmission analogique, la transmission numérique est relativement insensible aux bruits ou aux signaux perturbateurs) et une intégration plus facile des informations vocales numériques dans les nouveaux réseaux de transmission de données, [ADANT-83] et [ADANS-84].

Par la suite, nous allons essayer de mettre en évidence l'intérêt des systèmes informatiques de communication pour les milieux de la recherche et la vie dans l'entreprise. Nous terminerons en proposant quelques applications possibles des messageries vocales.

1.1. L'intérêt des systèmes informatiques de communication

Les systèmes informatiques de messagerie (traitant des messages brefs, peu structurés et informels), les systèmes de téléconférence par ordinateur (permettant la mise en commun des messages concernant un sujet précis) et les systèmes de courrier électronique (transmettant des messages plus étoffés et structurés comme par exemple des documents) sont des outils développés pour faciliter un mode particulier de communication: la communication par réseaux d'ordinateurs. Dans [BERGER-84], on prétend que ce nouveau mode de communication combine les avantages du téléphone et du courrier, tout en évitant les inconvénients:

- comme le téléphone, les réseaux permettent une transmission rapide, une économie des formules de politesse etc, mais, contrairement au téléphone, ils permettent également au destinataire de consulter son courrier uniquement quand il le souhaite et non pas lorsqu'il est contacté par l'émetteur.
- comme le courrier, ils offrent la possibilité d'échanger d'importantes quantités d'informations, de faire suivre le courrier en cas d'absence temporaire, de garder une trace écrite des communications etc, mais en plus, ils n'exigent aucune intervention de tiers (secrétaire, service postal etc).

D'après R. Mahl et P. Berger, ces moyens de communication par réseaux sont d'une grande importance pour les milieux de la recherche. Ils ont tellement transformé l'activité des chercheurs y ayant accès que, selon eux, on peut actuellement distinguer trois types de départements informatiques dans les universités américaines: 'la crème', c'est-à-dire ceux qui sont connectés à ARPANET (le réseau reliant les principaux centres de recherche travaillant pour le Département de la Défense des Etats-Unis), ceux n'ayant accès qu'à des réseaux secondaires et finalement ceux qui ne sont reliés à aucun réseau et par conséquent considérés comme hors-course.

1.2. L'intérêt des systèmes de messagerie vocale

Quant aux services de messagerie vocale en particulier, ils sont censés promouvoir une communication efficiente et économique [HANSON-84]. On considère que, comparés aux services de messagerie classiques, ces nouveaux systèmes ont l'avantage d'utiliser:

- La parole comme moyen de communication: ils évitent ainsi à l'émetteur la tâche de rédiger les messages, l'emploi de la parole lui étant plus familier et spontané. En outre la parole peut l'aider à mieux s'exprimer et mieux transmettre les informations contenues dans le message (l'intonation de la voix peut influencer sur la signification d'une phrase et faciliter sa compréhension). Une entreprise américaine en a fait l'expérience: après trois mois de fonctionnement du système, on a constaté que 77% des messages vocaux, contre 25% des messages écrits, étaient assez explicites pour permettre une réponse immédiate [DRINKWATER-85].
- Le téléphone comme moyen d'accès: il en résulte une plus grande indépendance du système face à la distance. Le téléphone étant un outil très répandu et facile à utiliser, le système peut être accédé à partir de presque n'importe quel endroit.

De plus, contrairement aux services téléphoniques actuels, la présence de l'interlocuteur est facultative pour la communication verbale. De ce fait:

- La disponibilité du système, vingt-quatre heures sur vingt-quatre, permet à l'utilisateur de gagner du temps, en ne l'obligeant plus à prendre en considération la présence de son interlocuteur à une certaine heure à un certain endroit; les problèmes de décalage horaire pour des communications à longue distance se trouvent ainsi éliminés.
- Pouvant consulter les messages reçus quand il le souhaite, le récepteur n'est plus interrompu par l'émetteur et peut travailler plus à l'aise [HANSON-84].
- Comme, selon [BAKER-84], on estime que jusqu'à 90% de toutes les communications importantes dans une entreprise commerciale se font verbalement, que jusqu'à présent, celles-ci se sont limitées essentiellement aux échanges face à face et aux conversations téléphoniques, que ces deux modes de communication verbale exigent la présence de toutes les parties concernées, ce qui n'est pas toujours facile, surtout pour les communications téléphoniques, et comme, d'après des études de l'International Resource Development, seulement un sur quatre des appels téléphoniques se faisant dans un intérêt commercial permet à l'émetteur de transmettre son message complètement dès le premier essai, le coût final d'une communication téléphonique étant ainsi estimé à treize dollars [LYMAN-84], la prise en compte d'une partie de ces communications verbales par des systèmes informatiques de messagerie vocale pourra donc permettre une optimisation du temps de travail en plus d'une économie d'argent et de personnel. En outre, les communications verbales traitées par un service de messagerie

n'étant plus interactives, on pourra les transmettre, particulièrement les communications à longue distance, durant les périodes de faible charge: la méthode de transmission par paquets, préférable du point de vue technique et économique à la transmission analogique du système téléphonique actuel (car elle permet une utilisation plus optimale du réseau), est, d'après [COLE-82], tout à fait indiquée dans ce cas, la parole étant déjà numérisée.

Toutefois, l'utilisation de systèmes informatiques de messagerie vocale connaît certaines limites:

- La manipulation d'objets sonores est loin d'être aussi bien maîtrisée que celle des textes écrits. Le caractère séquentiel et continu de la parole complique voire rend inadaptes certains traitements couramment effectués sur des textes écrits: modification (insertion ou suppression d'une partie d'une phrase), saut en avant ou en arrière dans le texte restitué, présentation de menus de fonctions etc. De plus, comme nous l'avons déjà mis en évidence dans les chapitres précédents, il faut mettre en oeuvre des techniques de numérisation et de codage complexes pour réduire jusqu'à des taux acceptables le débit binaire nécessaire à l'acquisition, la mémorisation et la transmission de la parole tout en garantissant que, lors de la restitution, elle sera encore intelligible. Quant aux techniques de reconnaissance multi-locuteur, elles ne donnent actuellement de bons résultats que si les commandes vocales et les destinataires des messages sont assez restreints en nombre et phonétiquement peu ressemblants. C'est pourquoi plusieurs des systèmes proposés actuellement (par exemple le DSC-2000 VoiceServer [HANSON-84] ou le Wang DVX [DRINKWATER-84]) utilisent les douze touches du téléphone pour introduire les différentes commandes sans ambiguïté.
- Enfin, notons encore qu'un système informatique de messagerie, pour pouvoir fonctionner d'une manière un tant soit peu efficace, exige qu'une communauté entière de gens travaillant ensemble l'utilise fréquemment [BAKER-84]. Chacun d'eux doit faire preuve d'une certaine discipline, surtout à la réception, et fournir l'effort régulier de retirer son courrier électronique et d'y répondre. Mais le succès commercial de ce nouveau produit informatique dépendra en fin de compte de sa capacité de satisfaire les besoins réels des utilisateurs. Par exemple, dans [NEWMAN-81], on estime que, dans le cadre de travaux de direction ou d'administration ('management'), les systèmes de messagerie sur ordinateur ne seraient efficaces que s'ils offraient, outre certaines facilités d'accès, des possibilités supplémentaires au simple traitement de texte. Selon J. Newman, les administrateurs ne consacrent que peu de temps à des conférences, à la consultation et l'envoi de messages. Donc, pour être utilisés plus souvent, les systèmes de messagerie devront leur offrir davantage, par exemple la faculté de filtrer l'information utile, de façon à accroître l'efficacité du traitement des informations captées dans l'environnement, et des possibilités d'aide à la décision.

1.3. Les applications de la messagerie informatique vocale

Du fait de leur capacité de transmettre des communications verbales de manière non-interactive à n'importe quelle heure du jour, à partir d'un endroit quelconque et sans exiger la présence du correspondant à l'autre bout du fil, les systèmes de messagerie vocale conviennent aux applications suivantes [LYMAN-84]:

- Aux communications internes des sociétés, soit à l'intérieur d'un même bâtiment, soit entre plusieurs sièges pas trop éloignés les uns des autres (par exemple dans une même ville). L'apparition de réseaux nationaux, supports de systèmes de messagerie vocale (par exemple Voicenet aux Etats-Unis annoncé par VMX en 1982), et leur interconnexion (par exemple la liaison du réseau anglais au réseau américain par Commercial Cable) devront permettre le développement d'un réseau international de télécommunications par ordinateur.
- Aux communications des représentants commerciaux en déplacement avec leur siège central.
- Au traitement de requêtes et de commandes. Certains systèmes de messagerie vocale offrent à des personnes extérieures (c'est-à-dire inconnues du système), la possibilité d'appeler et de déposer des messages pour les abonnés du système. Une société commerciale pourra avoir recours à de tels systèmes pour prendre en compte les requêtes et commandes de service provenant de sa clientèle. Sans que la personne appelante ait à s'en soucier, ses requêtes vont aboutir à la personne compétente grâce à la possibilité de redirection des messages. De plus, la communication n'étant plus interactive, cette dernière pourra s'en occuper plus à l'aise.
- Aux 'sociétés de service'. Bien que dans la plupart des cas l'investissement en équipement de messagerie vocale puisse être amorti en moins d'un an, il peut être considérable (les coûts d'un grand système peuvent atteindre \$500.000). Ceci explique l'apparition de 'sociétés de service' qui acquièrent de grands systèmes informatiques et puis revendent des services de messagerie à des entreprises plus petites n'ayant pas les moyens ou le désir de s'offrir leur propre système.

2. Le système de messagerie vocale MVOC

Notre stage au CRIN avait comme objet la conception d'un système de messagerie vocale pouvant être exploité par le personnel du CRIN pour améliorer les procédures de communication à l'intérieur des diverses équipes de recherche ainsi que la circulation de l'information entre ces dernières. L'intérêt principal du système devait être une grande facilité d'utilisation et d'accès, c'est-à-dire:

- le dialogue d'interfaçage devait pouvoir être oral, à partir d'un téléphone standard, ou écrit, à partir d'un terminal vidéo;

- de plus, le système devait permettre à toute personne extérieure d'accéder au système et de déposer des messages à destination de personnes abonnées;
- la gestion des messages devait rendre possible l'enregistrement des messages émis, la transmission au(x) destinataire(s) et la restitution à celui(ceux)-ci quand il(s) le souhaitai(en)t.

La manipulation de la parole devait se faire à l'aide d'outils disponibles au laboratoire, à savoir l'éditeur de sons EDIT que nous avons déjà présenté.

Dans les pages suivantes, nous allons exposer en détail le système de messagerie vocale MVOC en commençant par son organisation générale, définie dans une première phase de notre travail, après quoi nous verrons le dialogue d'interfaçage avec l'utilisateur dont la conception constituait une deuxième phase. Enfin une évaluation comparative de MVOC sera faite en guise de conclusion.

2.1. Organisation générale de MVOC

2.1.1. Les types d'utilisateurs et modes d'accès

Dans notre système de messagerie vocale, on distinguera:

- Deux modes d'accès:

Le mode **oral**: via un combiné téléphonique ou un terminal vidéo incorporant un micro et un haut-parleur.

Le mode **écrit**: via un terminal clavier-écran classique.

- Deux types d'utilisateurs:

Les **abonnés** sont les seuls utilisateurs à disposer d'une 'boîte-aux-lettres' pour recevoir des messages écrits ou oraux. Outre la possibilité d'émettre également ces deux types de messages à destination d'une ou de plusieurs personnes à la fois, ils peuvent, par combinaison des deux modes d'accès, émettre des messages mixtes pouvant être restitués à la fois sous forme vocale et sous forme écrite. Dans ce cas, ils doivent disposer d'une console comprenant un clavier, un écran, un micro et un haut-parleur. Par contre, si l'accès au système se fait de l'extérieur via le réseau téléphonique, il doit se limiter à une interaction orale, s'il se fait par un terminal classique, à une interaction écrite.

Les **non-abonnés** sont des utilisateurs occasionnels inconnus du système qui, ne disposant pas d'une boîte-aux-lettres, sont quand même autorisés à émettre des messages écrits ou oraux à destination d'un seul abonné à la fois (on reviendra plus loin sur les raisons de cette restriction). Une telle interaction avec le système se fait soit en mode oral, soit en mode écrit.

2.1.2. Les types de messages

Les messages manipulés par le système peuvent être de trois types:

- message **oral**:

Un message de ce type est saisi globalement sans aucune compréhension du contenu et, lors d'une demande de consultation par le destinataire, restitué de façon continue. Il correspond au type 'Parole', un des types d'objets manipulés par EDIT.

- message **mixte**:

Ce type de message est formé d'une suite de messages élémentaires (un mot ou un groupe de mots) que l'expéditeur compose à l'aide d'un éditeur de texte. Le système devra ensuite apprendre tous les messages élémentaires encore inconnus en mémorisant une réalisation vocale pour chacun. Le confort de l'écoute exige que pour un même message, les messages élémentaires le composant soient prononcés par une seule et même personne. Ainsi, le système disposant à la fois d'une version écrite et d'une version orale du message, ce dernier pourra être restitué sous l'une ou l'autre de ces deux formes. Contrairement aux messages de type oral, la reproduction vocale du message de type mixte ne sera pas continue, mais correspondra plutôt à une restitution mot à mot du message. Toutefois, les messages mixtes ont l'avantage d'être imprimables et moins exigeants en place mémoire, les messages élémentaires prononcés par un même abonné pouvant servir dans plusieurs messages. Ce type correspond au type 'Texte Parlé' d'EDIT.

- message **écrit**:

Ce type de message est conservé dans un souci de compatibilité avec les systèmes de messagerie classiques. Traité globalement comme le type Parole, il sera appelé 'Texte Ecrit' par opposition au Texte Parlé. Ce type d'objet n'est pas défini dans EDIT.

Les abonnés pourront envoyer ou recevoir les trois types de messages, tandis que les non-abonnés auront seulement la possibilité d'expédier des messages de type Parole ou Texte Ecrit, le système ne gérant pas de catalogues de messages élémentaires pour des utilisateurs occasionnels donc inconnus du système. Ils n'auront donc à leur disposition que les procédures d'émission de messages oraux ou écrits, procédures peu complexes. Signalons encore que les messages manipulés par MVOC sont d'une taille relativement limitée: un maximum de deux minutes de parole ou une page d'écran ont été envisagés pour les messages ordinaires tandis que pour les messages urgents, le sixième de cette taille a été retenu.

2.1.3. Les services offerts

Les services offerts par la messagerie MVOC peuvent être divisés en trois groupes:

- Les procédures gérant l'**émission** de messages.

On peut distinguer deux traitements différents selon la qualité de l'utilisateur. Le non-abonné ne peut qu'émettre des messages oraux ou écrits à destination d'un seul abonné à la fois tandis que l'abonné peut s'adresser à plusieurs abonnés simultanément, y compris lui-même (**message rappel**). L'abonné a en plus la possibilité d'envoyer des messages de taille réduite sur une base prioritaire (**message urgent**). Le système étant incapable d'appeler lui-même un abonné, cette priorité signifie uniquement que lorsque le destinataire accède au système, il sera averti immédiatement de la présence de messages urgents éventuels avant qu'il ne puisse entamer un traitement quelconque.

- Les procédures gérant la **consultation** de messages.

Ces traitements sont réservés exclusivement aux abonnés du système, les abonnés étant les seuls à disposer d'une boîte-aux-lettres. Notons que chacun d'eux ne peut consulter que les messages qui lui sont destinés.

- Les procédures **complémentaires** aux traitements d'émission et de consultation.

Accessibles uniquement via console, elles permettent aux abonnés:

- De définir et de gérer un ensemble limité de listes de diffusion. Ceci peut leur être d'une grande utilité s'ils communiquent souvent avec les mêmes groupes de personnes.
- De se renseigner sur les autres abonnés du système, les informations disponibles étant semblables à celles fournies par un botin téléphonique.
- De changer la version orale ou écrite de leur mot-de-passe.

Pour la définition de listes de diffusion et le changement de mots-de-passe, la console doit comprendre, outre un clavier et un écran, un micro pour l'acquisition des formes vocales correspondant aux noms des listes et aux mots-de-passe.

Voyons maintenant le détail des traitements en émission et en consultation de messages, les fonctions sous-jacentes ainsi que les options proposées aux utilisateurs.

2.1.4. L'émission des messages

L'émission d'un message se fait en deux phases distinctes; on procède d'abord à l'**enregistrement du (des) destinataire(s) du message** puis à l'**acquisition proprement dite du message**. Pour passer à la deuxième phase, il faut que la première ait réussi pour au moins un nom d'abonné.

L'envoi simultané d'un message à plusieurs destinataires exige l'introduction des différents noms, l'un après l'autre. Afin de lui éviter l'énumération de ces noms à

chaque fois qu'il adresse un message à un groupe de gens avec lequel il communique fréquemment, le système offre à l'utilisateur abonné uniquement, la possibilité de prédéfinir des **listes de diffusion**. Ainsi, au lieu de réciter toute la suite de noms, il lui suffit de fournir le nom de la liste prédéfinie. Une liste de diffusion pouvant être référencée aussi bien en émission orale qu'en émission écrite, la gestion de ces listes (création, destruction et modification) n'est possible qu'à travers un terminal combinant les deux modes d'accès.

L'acquisition des messages, excepté celle des messages urgents, se fait en deux étapes, tout message ordinaire (non-urgent) comprenant une **entête** composée de quelques **mots-clefs** spécifiant l'objet du message et un **corps** constitué du message proprement dit. On demande à l'émetteur de décrire le sujet de son message en quelques mots afin de donner au destinataire, lorsqu'il consulte son courrier électronique, une idée du contenu du message sans l'obliger à le lire ou à l'écouter en entier. Ces mots-clefs ne sont pas exigés pour les messages urgents, ceux-ci étant déjà de taille réduite. Ils seront immédiatement restitués dans leur intégralité.

L'acquisition d'un message terminée, l'émetteur peut réécouter, un nombre limité de fois, ce qu'il va envoyer. S'il juge que le message n'est pas au point, il a la possibilité d'empêcher une transmission immédiate et peut, soit reprendre l'émission au début s'il le désire (les mots-clefs et le message seront redemandés), soit abandonner (dans ce cas le message est perdu).

Notons encore qu'au cours d'un traitement de consultation (voir 'La consultation de messages'), l'utilisateur peut, après chaque message consulté, émettre une réponse à l'intention de son correspondant. Dans ce cas, le destinataire du message étant connu, l'émission se limite à l'acquisition du message.

2.1.5. La consultation de messages

Afin de permettre aux abonnés de prendre connaissance des messages reçus, le système MVOC prévoit trois types de consultation:

- **La consultation de messages urgents** qui est prioritaire et obligatoire, c'est-à-dire que l'utilisateur ne peut ni l'éviter ni l'arrêter si des messages urgents lui sont parvenus depuis son dernier accès à la messagerie.
- **La consultation normale** qui permet de prendre connaissance de tous les messages de la boîte-aux-lettres dans l'ordre de leur arrivée, mais en reprenant d'abord les nouveaux puis les anciens (un message est considéré comme ancien s'il a déjà été consulté une fois au moins et en entier, et a été sauvegardé sur demande explicite de l'utilisateur). Cette consultation est optionnelle et peut être arrêtée après chaque message consulté.
- **La consultation particulière** qui permet à un abonné de consulter dans le même ordre que lors d'une consultation normale, les messages lui provenant d'une personne donnée. Afin d'être connue du système, cette personne doit donc être

elle-aussi abonnée à la messagerie. Ce type de consultation est également optionnel et peut être arrêté après chaque message consulté.

Lorsqu'il consulte un message, l'utilisateur obtient d'abord une description succincte du message avant de recevoir le contenu complet. Cette description comprend:

- Le nom de l'émetteur, sauf lors d'une consultation particulière, auquel cas il serait évident.
- La date et l'heure d'émission qui correspondent ici à l'instant de réception, la transmission étant immédiate.
- L'objet du message, c'est-à-dire quelques mots-clefs. Ils ne sont fournis que dans le cas de la consultation normale pour permettre au récepteur, après avoir pris connaissance de l'arrivée d'un message, soit d'écouter ou de lire le corps entier du message, soit de passer directement au message suivant. Cette possibilité n'est offerte à l'utilisateur ni en consultation urgente, ni en consultation particulière, dans le premier cas, la consultation étant obligatoire et prioritaire, dans le deuxième cas, l'utilisateur ayant demandé explicitement la consultation des messages en provenance d'un abonné particulier et spécifié par lui.
- Les autres destinataires s'il y en a. Le système fournira alors leurs noms à la demande de l'utilisateur.

Après avoir pris connaissance du contenu d'un message, l'abonné a plusieurs possibilités:

- Demander une réécoute (un nombre limité de fois). Celle-ci peut, soit se limiter aux informations sur le message ou à son contenu, soit englober les deux.
- Passer en mode émission afin de répondre à l'auteur du message qu'il vient de consulter.
- Exiger une version imprimée du message consulté (Ceci n'est possible que si le message est imprimable, c'est-à-dire s'il est de type Texte Ecrit ou de type Texte Parlé).
- Sauvegarder ou détruire le message consulté. La sauvegarde d'un message urgent a comme effet de transformer ce message en un message ordinaire avec comme mot-clef: 'urgent'. Tous les messages nouveaux consultés en entier deviennent implicitement des messages anciens tandis que tous les messages n'ayant pas été entièrement lus par l'abonné durant la session vont rester inchangés. Les messages détruits disparaissent définitivement de la boîte-aux-lettres.
- S'il est en consultation normale ou particulière, arrêter ou poursuivre la consultation des messages; la consultation des messages urgents ne pouvant être interrompue.

Enfin, le traitement de consultation prévoit encore quelques **fonctions de destruction**. Celles-ci sont proposées à l'abonné avant qu'il n'entame vraiment la consultation des messages non-urgents mais seulement après lui avoir communiqué l'état de sa boîte-aux-lettres, c'est-à-dire le nombre de messages nouveaux et anciens. Cela signifie qu'en mode d'accès oral on l'informerait du nombre de messages oraux (de type Texte Parlé ou Parole) tandis qu'en mode écrit on lui fournirait celui des messages écrits (de type Texte Parlé ou Texte Ecrit). Disposant de cette information succincte, l'utilisateur a la possibilité de faire le ménage à l'aide des fonctions suivantes:

- **Destruction totale:** Cette fonction élimine tous les messages relatifs au type de la consultation.
- **Destruction partielle:** Cette fonction limite l'effacement aux messages dont la date d'émission est antérieure à une date fixée par l'utilisateur (la date limite incluse). L'acquisition de cette date étant faisable mais coûteuse en mode oral, elle exige en effet un accroissement considérable de la taille du lexique relatif à chaque abonné, cette fonction de destruction sera réservée au mode écrit et n'aura donc un effet que sur les messages concernés par ce mode d'accès.
- **Destruction particulière:** Cette fonction détruit le(s) message(s), oral(aux) ou écrit(s) selon le mode d'accès, en provenance d'un correspondant donné. Dans le cas de destructions multiples, une confirmation est exigée pour chaque message à effacer.

2.1.6. L'interface avec l'utilisateur

Le dialogue oral avec l'utilisateur envisagé dans MVOC est fortement guidé par le système, offrant ainsi peu de flexibilité et d'initiative à l'utilisateur. Il est du type **question-réponse**, l'utilisation de la parole ne permettant pas de présenter des menus avec, à chaque niveau, un nombre important d'alternatives. L'enchaînement du dialogue a donc été conçu de manière à réduire le nombre de choix proposés à deux ou trois par niveau. Une contrainte supplémentaire est imposée par l'utilisation de la méthode de reconnaissance par mots isolés pour l'analyse des réponses orales de l'utilisateur, celle-ci étant assez bien maîtrisée. Seuls les mots-clefs et les messages proprement dits peuvent être prononcés d'une façon naturelle lors du traitement d'émission. Ils sont en effet simplement saisis et mémorisés, moyennant des techniques de numérisation de la parole, pour être restitués tels quels au(x) destinataire(s).

Quant au dialogue clavier-écran qui n'est cependant pas encore réalisé, nous pensons qu'il devra rester assez semblable au dialogue oral; le même enchaînement des fonctions d'émission et de consultation des messages sera conservé afin de ne pas trop irriter les utilisateurs. La seule différence se situera au niveau des traitements de destruction partielle, de gestion des listes de diffusion etc, traitements non-prévus en mode oral. Nous reviendrons plus loin sur la partie du dialogue oral réalisée.

2.1.7. Les informations nécessaires

Afin de mener à bien la gestion des boîtes-aux-lettres des différents abonnés, le système MVOC dispose d'un ensemble d'informations. Ces informations concernant essentiellement les abonnés sont composées pour chacun

- de son nom (version orale et version écrite);
- de quelques renseignements généraux (adresse, profession etc), dans le but de constituer une documentation du genre botin téléphonique;
- des listes de diffusion prédéfinies et des listes temporaires, une liste temporaire étant constituée par le système dès qu'un utilisateur envoie un message à plusieurs abonnés sans définir lui-même une liste de diffusion (permettant à chaque récepteur du message de prendre connaissance des autres destinataires s'il y en a plusieurs, le système doit conserver la liste temporaire jusqu'à ce que tous les destinataires aient consulté et détruit le message en question);
- du lexique contenant l'ensemble du vocabulaire dont dispose l'abonné lors des sessions de consultation ou d'émission (à l'exception bien entendu des mots clefs et du message), ceci afin que les réponses de l'utilisateur puissent être interprétées par la méthode de reconnaissance par mots isolés (produisant en effet de bons résultats en mono-locuteur avec un vocabulaire restreint, elle justifie donc le recours à un lexique par abonné, du moins pour les abonnés les plus zélés);
- du catalogue des messages élémentaires pouvant servir à la composition de messages mixtes de type Texte Parlé;
- de l'état de la boîte-aux-lettres (les totaux des différents types de messages, anciens et nouveaux, reçus);
- du contenu de la boîte-aux-lettres, c'est-à-dire d'une liste de messages avec, par message, une indication de type, des informations succinctes sur le message (émetteur, date et heure d'émission, liste des autres destinataires s'il y en a plusieurs et mots-clefs s'il s'agit d'un message ordinaire) et le corps du message;

Les informations sauvegardées sur l'utilisateur non-abonné se limitent au nom dans sa version orale ou écrite, selon le mode d'accès, spécifié lors de l'émission. Cette sauvegarde n'est que temporaire, le nom ne pouvant être référencé qu'aussi longtemps que le message émis existe dans le système. Afin de comprendre les réponses des non-abonnés, le système devra disposer d'un lexique général. Face à ces utilisateurs occasionnels le système fonctionne en multi-locuteur et un choix judicieux des mots composant le vocabulaire peut influencer fortement sur la qualité de la reconnaissance.

Finalement, le système nécessite un ensemble de messages oraux ou écrits, suivant le mode d'accès, qui lui permette de s'adresser à l'utilisateur. Le souci d'économie de place mémoire nous a amené à adopter des objets de type Texte Parlé

pour ces messages, choix qui nous permet également d'utiliser certains des messages oraux en communication écrite.

Pour permettre à MVOC de fonctionner convenablement, cet ensemble d'informations doit être tenu à jour. Il faut donc prévoir quelques fonctions système permettant au(x) responsable(s) de la messagerie:

- l'ajout et le retrait d'abonnés avec tout ce que cela implique;
- la mise à jour des messages système en cas d'évolution de ce dernier;
- de faire des contrôles, par exemple l'inspection de la taille des boîtes-aux-lettres, visant à garantir un bon fonctionnement général.

2.1.8. Les types d'objets utilisés

Les pages suivantes présenteront les différents types d'objets utilisés dans le système MVOC. Certains correspondent à des types définis dans EDIT, d'autres ont dû être ajoutés.

Comme nous venons de le voir, les messages manipulés par la messagerie peuvent être de trois types:

- de type **Texte Parlé** (noté TP et constitué d'une suite de messages élémentaires notés ME): il s'agit des messages mixtes et des messages système;
- de type **Texte Ecrit** (noté TE): il s'agit des messages écrits;
- de type **Parole** (noté P): il s'agit des messages oraux.

De plus, nous avons vu qu'en mode oral, à l'exception des mots-clefs et du message, l'utilisateur doit prononcer ses énoncés mot par mot afin de permettre une reconnaissance par mots isolés. On peut dès lors distinguer deux types de parole:

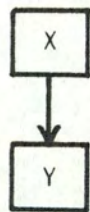
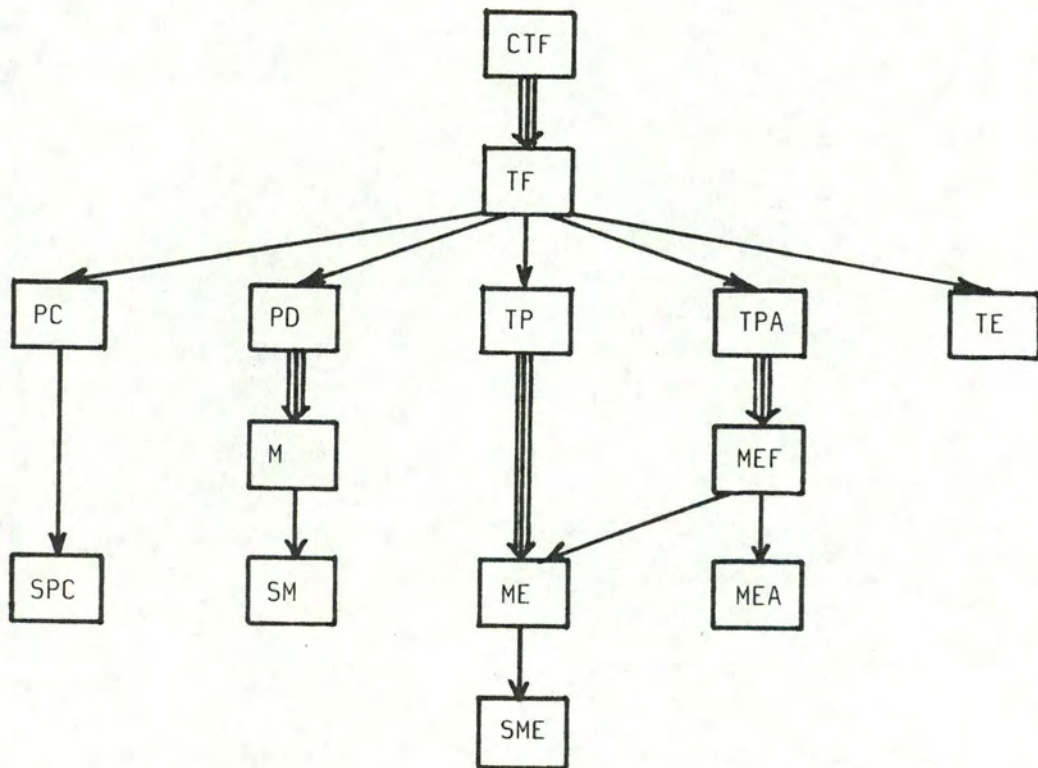
- La **Parole Continue** (notée PC) pour les messages oraux, mots-clefs compris. Elle correspond au type Parole de EDIT.
- La **Parole Discontinue** (notée PD) pour les réponses de l'utilisateur. Elle est constituée d'une suite d'objets de type Mot (noté M), la reconnaissance automatique travaillant au niveau du mot encore inconnu.

Par analogie aux figures 2,3 et 4, relatives aux types d'objets et fonctions d'EDIT, les figures 5,6 et 7 décrivent respectivement les types d'objets manipulés par MVOC, les relations entre ces différents types ainsi que les fonctions qui leurs sont applicables. Notre travail a présupposé l'existence de ces fonctions dont une grande partie a d'ailleurs été réalisée dans le cadre de l'éditeur de sons EDIT.

Fig. 5. Liste des types d'objets de MVOC

TYPE	SPECIFICATION INFORMELLE
CTF	Type catalogue d'objets de type TF.
TF	Type générique regroupant des objets de type parole continue (PC), parole discontinue (PD), texte parlé (TP), texte parlé en attente d'apprentissage (TPA) et texte écrit.
PC	Type parole continue.
SPC	Type signal de parole associé à un objet de type parole continue.
PD	Type parole discontinue constitué d'une suite d'objets de type mot.
M	Type mot
SM	Type signal de parole associé à un objet de type mot.
TP	Type texte parlé dont tous les messages élémentaires sont connus.
TPA	Type texte parlé en attente d'apprentissage composé d'une suite de messages élémentaires dont quelques-uns sont encore inconnus.
TE	Type texte écrit.
MEF	Type message élémentaire fictif regroupant les objets de type message élémentaire connu (ME) et de type message élémentaire en attente d'apprentissage (MEA).
ME	Type message élémentaire connu.
MEA	Type message élémentaire connu en attente d'apprentissage.
SME	Type signal de parole associé à un message élémentaire.

Fig. 6. Schéma des relations entre les différents types d'objets



un objet Y intervient
dans la composition de X



une suite d'objets Y interviennent
dans la composition de X

Fig. 7. Les fonctions primitives définies dans MVOC

Nom de la fonction	Paramètres	Types d'objets modifiés	Description informelle de la fonction
Acquérir		PC, PD, M, ME	déclenche la saisie au micro pour acquérir les signaux de parole relatifs à un objet de type parole continue, parole discontinue, mot ou message élémentaire;
Créer	iden	PC, PD, M, TP, ME, TE	crée l'objet de nom 'iden';
Détruire	iden	PC, PD, TP, TE	détruit l'objet 'iden': 'iden' est supprimé définitivement dans le catalogue;
Ecouter	iden	PC, TP, ME	restitue sous forme sonore l'objet 'iden';
Générer		PC, PD, M, TP, ME, TE	réinitialise le catalogue courant: tous les objets contenus dans le catalogue sont détruits;
Imprimer	iden, perif	TP, ME, TE	édite sur le périphérique désigné par 'perif' les composantes de l'objet 'iden' dans le catalogue courant;

(Fig. suite)

Nom de la fonction	Paramètres	Types d'objets modifiés	Description informelle de la fonction
Lister	iden1, iden2, périf	M, ME	édite sur le périphérique désigné par 'périf' les noms des objets contenus dans le catalogue courant et dont le nom est compris entre 'iden1' et 'iden2';
Prendre	iden	PC, PD, M, TP, ME, TE	si l'objet de nom 'iden' se trouve dans le catalogue courant, il devient l'objet courant;
Reconnaître	iden1, iden2,	M	indique si oui ou non, l'objet inconnu 'iden1' a été reconnu comme étant identique à l'objet 'iden2';
Sauver		PC, PD, M, TP, ME, TE	sauvegarde un objet créé dans le catalogue;

2.2. Le dialogue oral utilisateur-système

Après avoir défini l'organisation générale du système de messagerie MVOC, nous avons entamé la conception et la réalisation du dialogue oral entre l'utilisateur et le système. Comme nous l'avons déjà indiqué, le dialogue envisagé pour cette application est de type question-réponse. La reconnaissance se fait par mots isolés et on laisse relativement peu d'initiatives à l'utilisateur. Ce genre de dialogue convient parfaitement à des systèmes de cette catégorie. Il provoque peu d'erreurs et, le traitement des réponses étant assez simple, conduit à des temps de réponse acceptables.

Notre réalisation distingue nettement entre utilisateurs fréquents (abonnés) et utilisateurs occasionnels (non-abonnés). A ces derniers on n'offre d'ailleurs que des possibilités limitées d'émission (ils ne peuvent par exemple envoyer des messages qu'à un seul abonné à la fois), le dialogue a été conçu pour eux de manière à limiter les choix possibles à un mot significatif. De plus, afin de réduire les problèmes de reconnaissance, nous avons sélectionné des mots peu ressemblants du point de vue phonétique. Enfin, ces utilisateurs n'étant pas très expérimentés, les messages du système sont plus explicites (ils indiquent à chaque fois les réponses qu'il convient de donner). Quant aux abonnés, nous avons également cherché à réduire les choix à un mot significatif, ceci dans un souci d'économie de place, en réduisant la taille des lexiques, et afin de permettre un dialogue moins verbeux simplifiant ainsi la reconnaissance tout en améliorant les temps de réponse. Les messages systèmes prévus sont moins explicites, ces utilisateurs connaissant généralement assez bien l'enchaînement des différentes phases du dialogue. Si un abonné fournit cependant une réponse mauvaise, des messages plus détaillés sont disponibles pour mieux le guider.

En guise d'illustration, nous allons présenter un exemple de dialogue possible en début et en fin de session (pour un abonné et un non-abonné) ainsi qu'un exemple d'écoute lors d'une consultation. Comme nous l'avons déjà précisé, la réalisation s'est faite sur une SM90 fonctionnant sous UNIX, les entrées-sorties vocales ayant été simulées par un dialogue clavier-écran. Les schémas montrant l'enchaînement du dialogue peuvent être trouvés en annexe.

Fig. 8. Dialogue en début et en fin de session, cas de l'abonné

S(ystème): Bonjour, vous êtes en liaison avec le système de messagerie vocale du CRIN. Veuillez donner votre nom.
A(bonné): Josiane Muller.
S : Veuillez donner votre mot-de-passe.
A : Toto.
S : Ce mot-de-passe est incorrect, veuillez recommencer.
A : Rintintin.
S : Vous avez reçu cinq messages urgents oraux.
Un message urgent de Jean Dupont émis le dix-sept août à quinze heures.
<message urgent 1>
. . .
S : Ceci étaient tous les messages urgents oraux.
Vous avez encore reçu trois messages urgents écrits.
Que désirez-vous, émettre, consulter ou arrêter?
A : Emettre.
. . .
S : Désirez-vous encore consulter des messages?
A : Oui.
. . .
S : Avez-vous terminé?
A : Oui.
S : Merci de votre appel, bonne journée.

Fig. 9. Dialogue en début et en fin de session, cas du non-abonné

S : Bonjour, vous êtes en liaison avec le système de messagerie vocale du CRIN. Veuillez donner votre nom.

N(on-abonné): Gérard Lambert.

S : Qui est le destinataire du message que vous voulez émettre? Veuillez donner son nom.

.

.

.

S : Voulez-vous encore émettre un autre message?
Si oui, répondez par OUI sinon, par TERMINE.

N : Terminé.

S : Merci de votre appel, bonne journée.

Fig. 10. Dialogue lors de l'écoute d'un message

.
.
.
S : Un message de Christian Bernard émis le vingt juillet à neuf heures quinze concernant: <mots-clefs>. Voulez-vous l'écouter?
A : Non.
S : Voulez-vous poursuivre la consultation?
A : Oui.
S : Un message de Jean Legrand émis le premier août à cinq heures concernant: <mots-clefs>. Voulez-vous l'écouter?
A : Oui.
S : <message>. Désirez-vous une réécoute?
A : Oui.
S : Que voulez-vous réécouter, seulement le message, seulement les informations sur le message ou voulez-vous réécouter tout? Selon le cas, veuillez répondre par MESSAGE, par INFORMATION ou par TOUT.
A : Message.
S : <message>. Désirez-vous une réécoute?
A : Non.
S : Voulez-vous envoyer un message de réponse?
A : Non.
S : Voulez-vous sauvegarder le message écouté?
A : Non.
S : Voulez-vous poursuivre la consultation?
.
.
.

2.3. Evaluation du système MVOC

Comparé à d'autres systèmes de messagerie vocale, MVOC, tel que nous l'avons conçu, prévoit la plupart des fonctions généralement offertes: émission, consultation, sauvegarde, destruction, diffusion multiple et simultanée de messages, possibilité de réponse à un message consulté, par contre, dans sa version actuelle, il ne permet pas encore la redirection d'un message consulté. Pour autant que le message à rediriger reste inchangé, cette fonction devrait cependant pouvoir être ajoutée assez facilement. Autre restriction: le système ne disposant d'aucun moyen pour signaler l'arrivée d'un message à un abonné (par exemple à l'aide d'un signal lumineux disposé sur le poste), nous avons limité l'émission des messages à une diffusion immédiate (la diffusion différée étant surtout intéressante lorsqu'elle est combinée avec des messages rappel).

Les particularités de MVOC par rapport aux autres systèmes peuvent être résumées de la manière suivante:

- Il accepte un dialogue en langage parlé alors que la plupart des systèmes existants ont recours à des codes numériques, introduits à l'aide des touches du téléphone. L'utilisation de commandes numériques et le multiplexage du canal vocal, permettant ainsi la transmission simultanée des messages et des fonctions de contrôle, ont quand même l'avantage de rendre possible des manipulations sur la parole irréalisables dans MVOC (nous pensons par exemple aux sauts en avant ou en arrière dans le texte restitué, au réglage du débit des messages en cours de restitution, à l'arrêt de la reproduction sonore etc), le prix en étant un apprentissage plus long. Il semblerait en outre que la méthode par codes numériques laisse un peu plus d'initiative à l'utilisateur que le dialogue envisagé dans MVOC qui, lui, est fortement guidé. Précisons tout de même que ce dialogue pourra sûrement évoluer avec les progrès en compréhension du dialogue. Pour l'instant la solution serait plutôt de prévoir plusieurs structures de dialogue en fonction du niveau d'expertise acquise par l'utilisateur. La tâche de reconnaissance s'en verrait évidemment alourdie.
- Il permet à des personnes inconnues du système de déposer des messages à l'intention d'utilisateurs abonnés.
- Il permet au destinataire une présélection des messages à consulter par l'existence de mots-clefs. Ceci réduit l'inconvénient provenant du fait qu'on ne peut pas interrompre une restitution de message.
- Il prévoit plusieurs possibilités de destruction ou de consultation de messages.

Dans les annexes on trouvera une comparaison plus détaillée entre le système d'audio-messagerie d'IBM et le système MVOC.

En ce qui concerne l'évolution future du système MVOC, nous avons pensé à plusieurs extensions éventuelles:

- Le système manipulant trois types de messages (écrit, oral et mixte), on pourrait envisager pour des raisons de cohérence, outre les deux modes d'accès écrit et oral déjà définis, un mode mixte à partir d'un terminal intégrant clavier, écran, micro et haut-parleur, mode permettant des actions sur tous les types de messages. En début de session, l'utilisateur s'étant identifié, il pourrait ainsi choisir entre un dialogue oral ou écrit. Un abonné pourrait ensuite émettre et consulter les trois types de messages, quant au non-abonné, il pourrait émettre aussi bien des messages écrits que des messages oraux en une même session. De cette façon, l'émission de messages de type Texte Parlé ne devrait plus être traitée comme un cas particulier dans le cadre d'un accès en mode écrit. De plus, lors d'un accès en mode mixte, un abonné aurait la possibilité de consulter les messages de tous les types au cours d'une seule session.
- Le système étant destiné à être implémenté dans une organisation déjà structurée en un certain nombre de groupes de travail, il pourrait être intéressant de disposer de listes de diffusion, prédéfinies et correspondant à ces groupes, au niveau du système lui-même, accessibles à tous les utilisateurs au même titre que les abonnés. Ceci devrait conduire à une réduction sensible des listes de diffusion prédéfinies par ces derniers. Pour les non-abonnés, l'avantage serait la possibilité de s'adresser à un groupe de destinataires s'ils sont familiarisés avec la nomenclature en vigueur dans l'organisation.

Enfin, le dialogue pourrait être étendu de manière à permettre à un utilisateur non-abonné l'émission simultanée d'un message vers plusieurs destinataires en une seule phase de dialogue. Pour cela il doit suffisamment guider l'utilisateur afin d'éviter toute ambiguïté dans l'interaction.

CHAPITRE 6

CONCLUSION

Dans ce mémoire nous avons essayé de situer notre travail, c'est-à-dire le développement d'un système informatique de messagerie vocale, dans le cadre plus général des techniques de traitement automatique de la parole. Ce système utilise certains types d'objets et fonctions définis dans l'éditeur de sons EDIT pour mener un dialogue avec l'utilisateur. Nous avons donc successivement:

- exposé, de manière sommaire, les méthodes de reconnaissance et de synthèse de la parole, l'intérêt des systèmes de reconnaissance et de synthèse pour les systèmes informatiques ainsi que quelques applications concrètes envisagées;
- vu quelques aspects généraux de la communication homme-machine avec en particulier une typologie des dialogues;
- parlé de l'édition de sons et de ses utilisations possibles, nous avons présenté en détail l'éditeur de sons EDIT, développé au CRIN pour l'adjonction d'entrées-sorties vocales à un système d'EAO.

Enfin nous avons entamé le domaine des systèmes informatiques de messagerie avec un relevé des avantages de ces systèmes pour les communications internes et externes des organisations et nous avons exposé le système de messagerie vocale MVOC dont l'organisation générale ainsi que le type de dialogue envisagé pour la communication orale avec l'utilisateur ont été passés en revue. Une évaluation comparative de ce système et quelques perspectives d'avenir ont clôturé cette partie.

Il nous reste à signaler que l'apparition de systèmes électroniques de messagerie vocale s'insère dans une tendance plus vaste qui vise une intégration plus poussée de l'informatique, de la téléphonie et des télécommunications afin de donner lieu à des moyens de communications polyvalents, d'une accessibilité facile et plus généralisée. Cette intégration devient possible grâce aux techniques de numérisation de la voix. Ainsi, on peut remarquer, d'une part, le développement d'applications permettant l'emploi du téléphone comme terminal d'accès (certains auteurs vont même jusqu'à affirmer que le téléphone pourrait devenir le terminal universel dans la prochaine décennie [SCHALK-83]) et, d'autre part, une évolution vers des réseaux téléphoniques numérisés. D'après [ADANS-84], le développement d'une téléphonie digitale, moyennant quelques extensions à la fois faciles et à coût modéré du réseau téléphonique pratiquement omniprésent, pourrait conduire à un réseau unique capable de véhiculer simultanément des données et de la parole à une vitesse moyenne

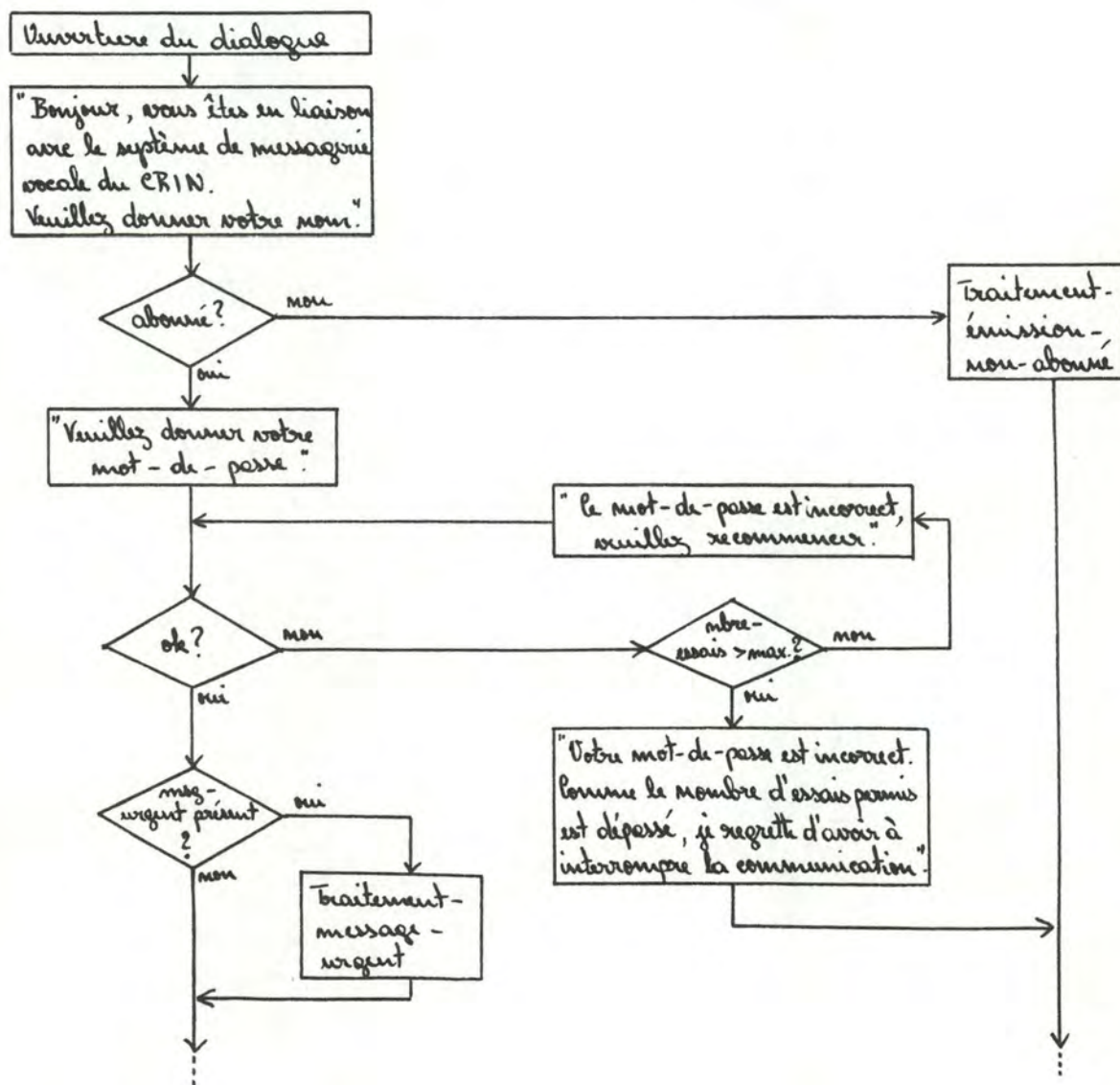
(64Kbs) et de couvrir la majorité des besoins en matière de communication des organisations. De cette manière l'installation de deux réseaux de communication en parallèle (réseau téléphonique et réseau de transmission de données) pourrait être évitée aisément. Les seules applications pour lesquelles ce type de réseau serait inapproprié sont les applications nécessitant des vitesses de transmission très élevées (de l'ordre de 10Mbps et plus). En outre, comme la digitalisation de la téléphonie nécessitera l'introduction d'autocommutateurs numériques privés qui ne seront en fait rien d'autre que des ordinateurs dédiés à la transmission en temps réel de données digitales, ces ordinateurs pourront éventuellement prendre en charge d'autres fonctions, par exemple des services de messagerie du type MVOC.

Cependant, le développement de ces moyens de communication, censé favoriser une circulation rapide et volumineuse des informations, présente également certains dangers. La saisie, la transmission et l'analyse d'informations, rendues aisées grâce à la quasi-omniprésence des nouveaux moyens de communication, peuvent engendrer une expansion des techniques de surveillance, une concentration des activités de fichage, la constitution d'élites (l'accès à certaines informations leur étant réservé; ce phénomène s'est produit dans le cas du réseau ARPANET) etc. De ce fait, le développement de nouvelles technologies capables de modifier profondément notre mode de vie et de porter atteinte à certains droits de l'homme, devrait être soumis à un contrôle très strict. Ce contrôle devrait s'exercer aussi bien au niveau de la décision de développer un type donné de technologie qu'au niveau de l'exploitation de cette dernière afin de réduire à un minimum les abus possibles. Nous pensons que la création d'organismes tels que la Commission Nationale de l'Informatique et des Libertés (CNIL) en France représente un premier pas dans cette direction.

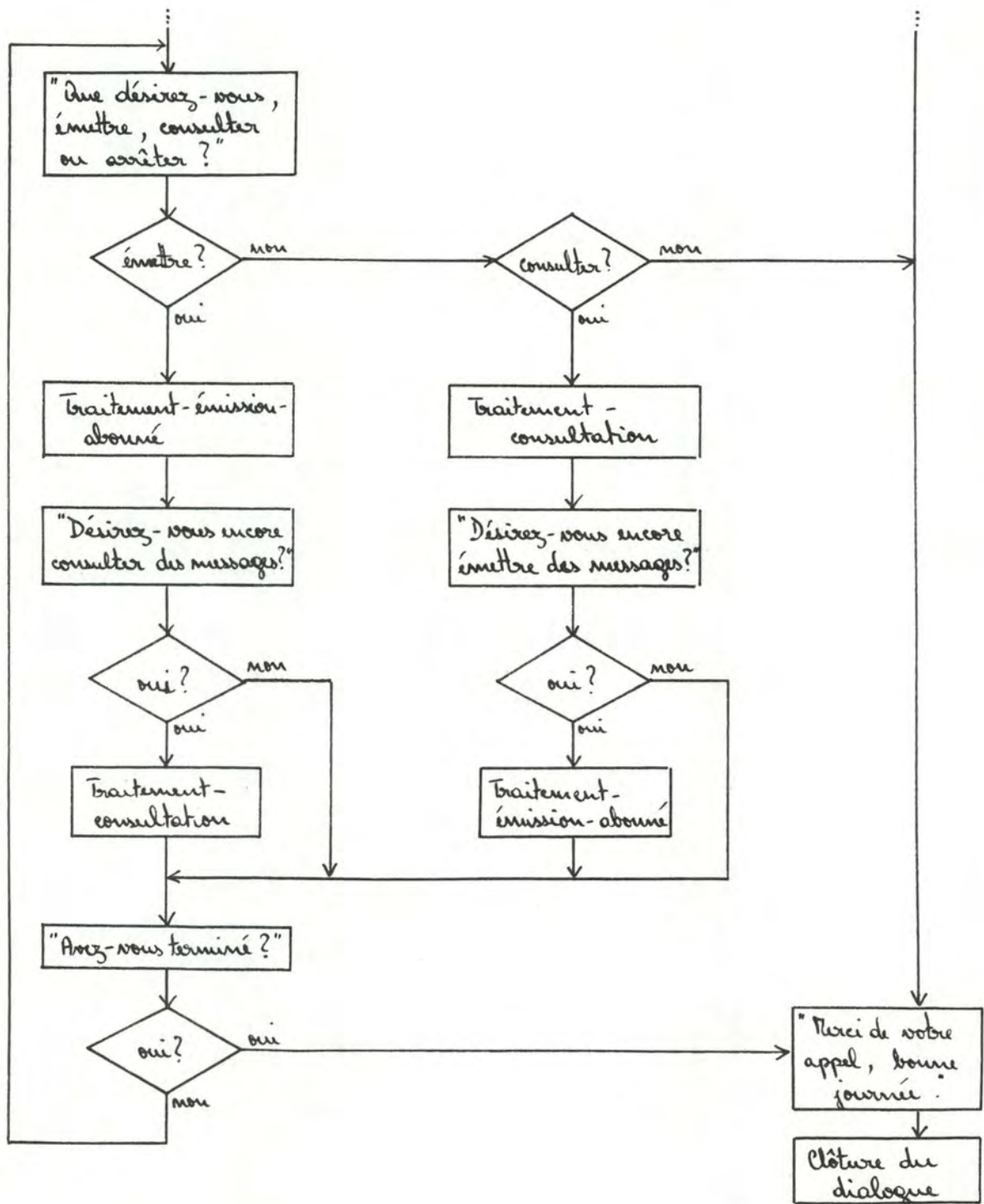
ANNEXE A

ENCHAINEMENT DU DIALOGUE ORAL

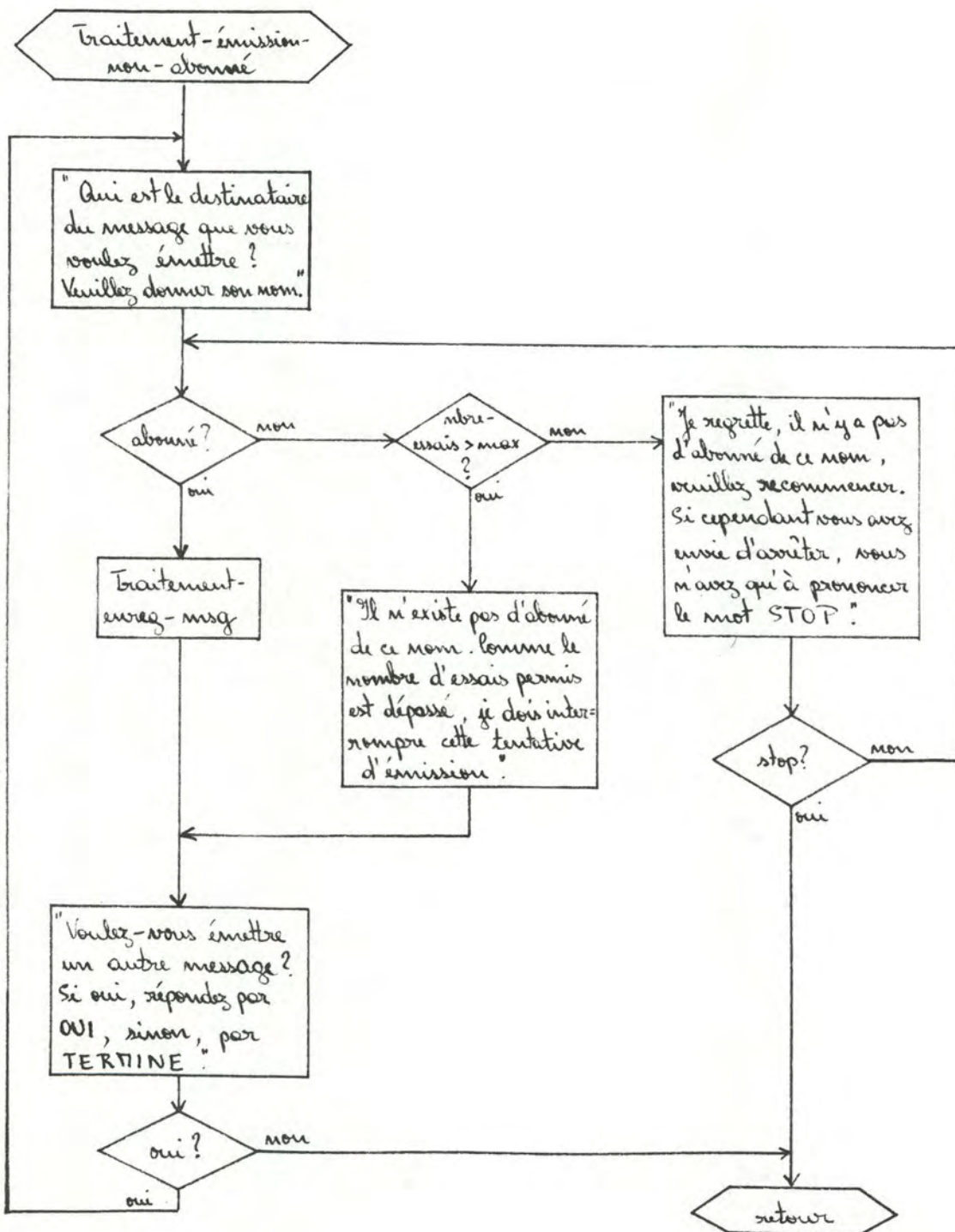
Fig. 11. Schéma du dialogue



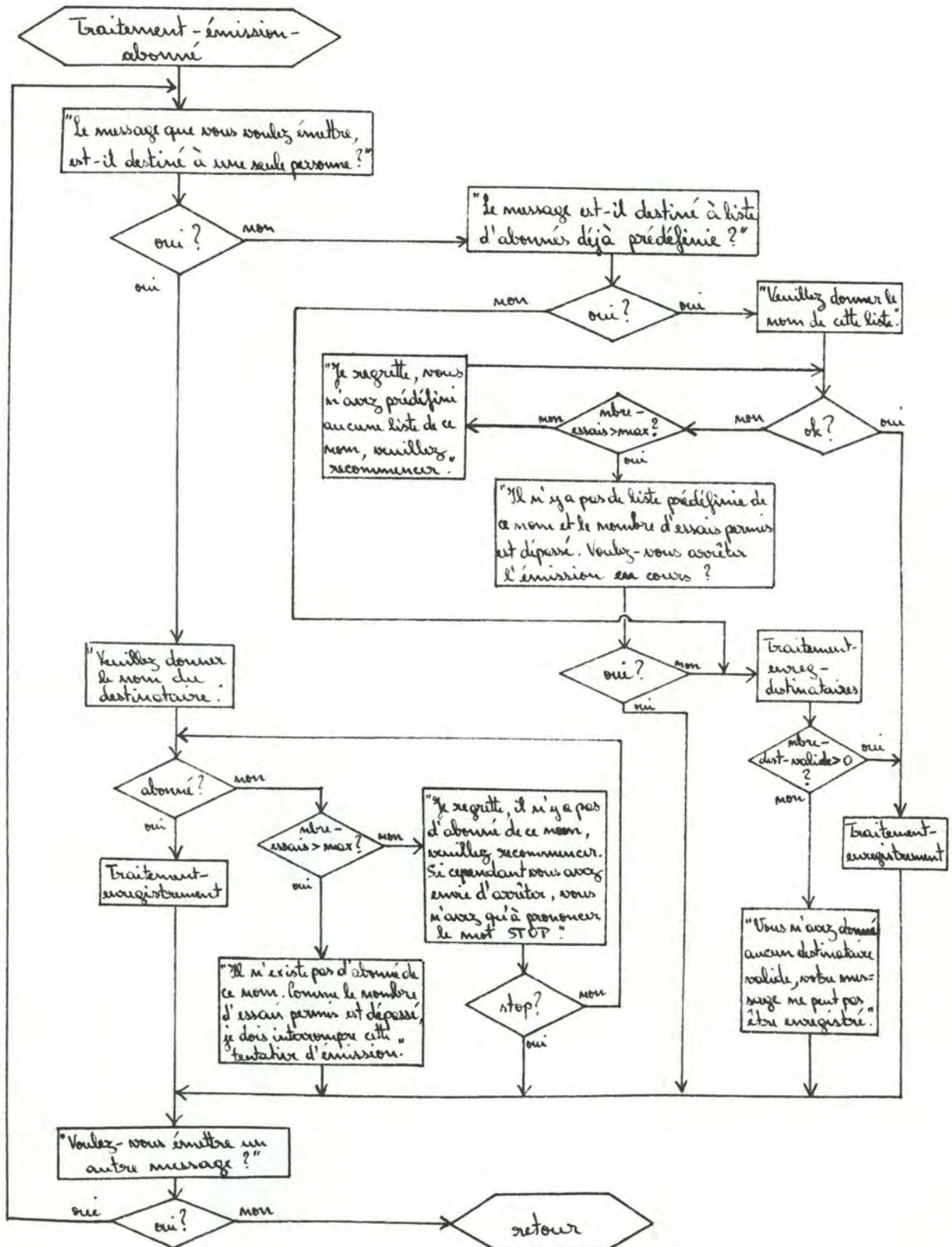
(Schéma suite)



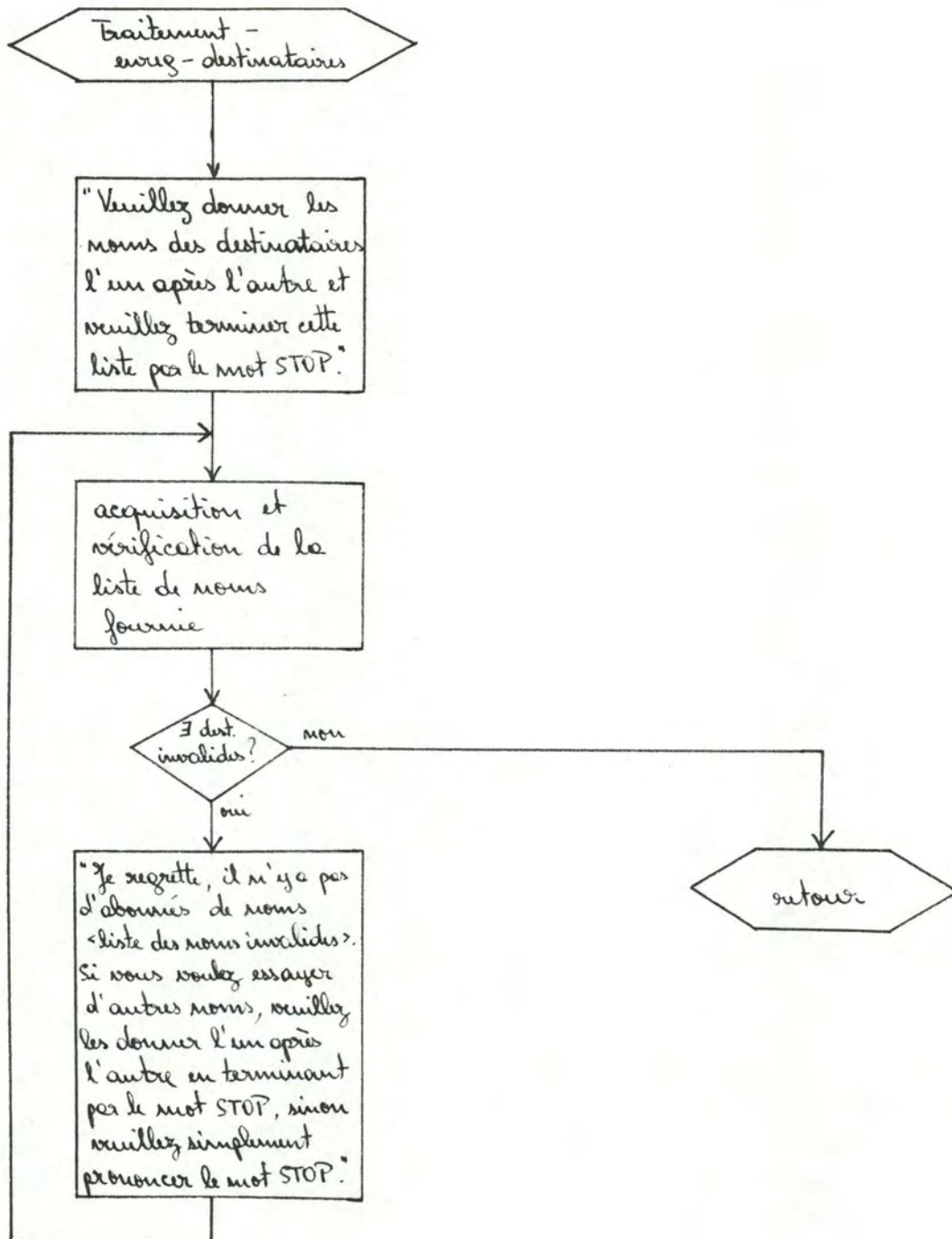
(Schéma suite)



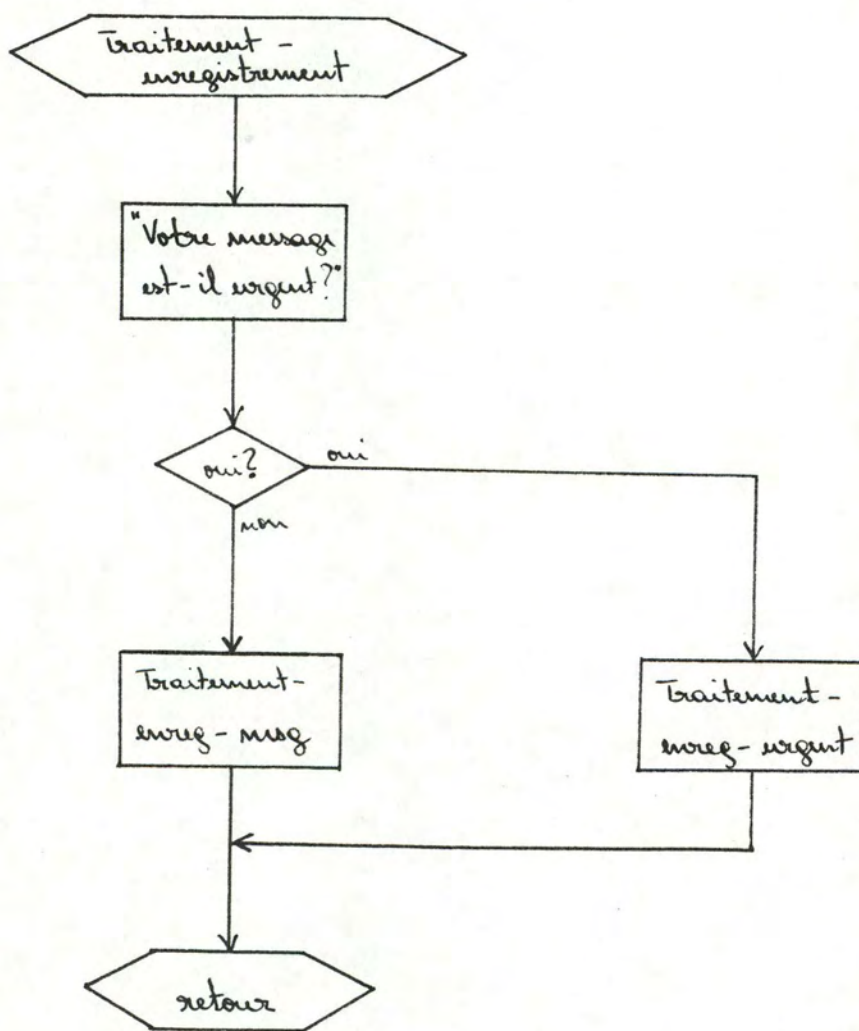
(Schéma suite)



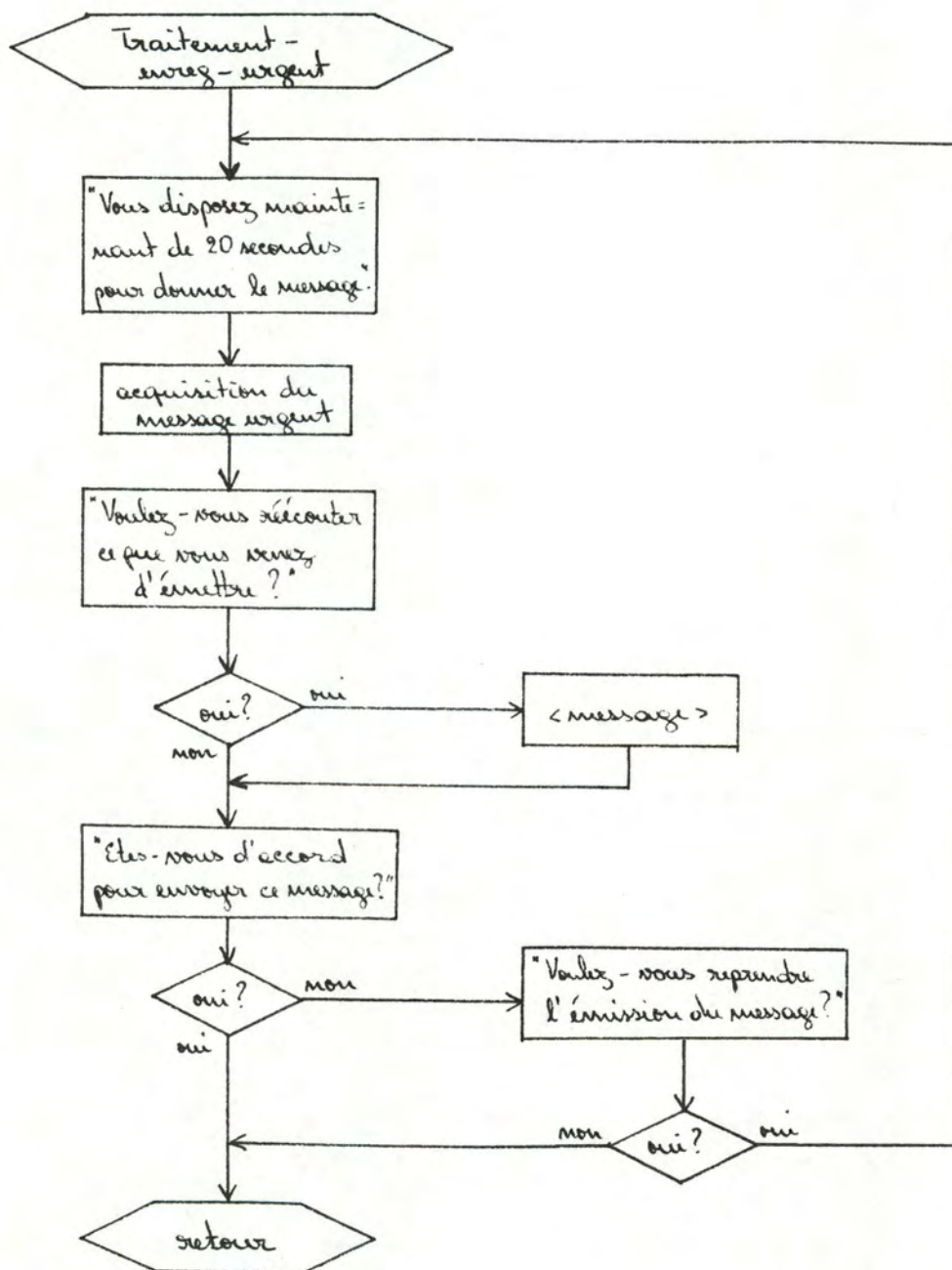
(Schéma suite)



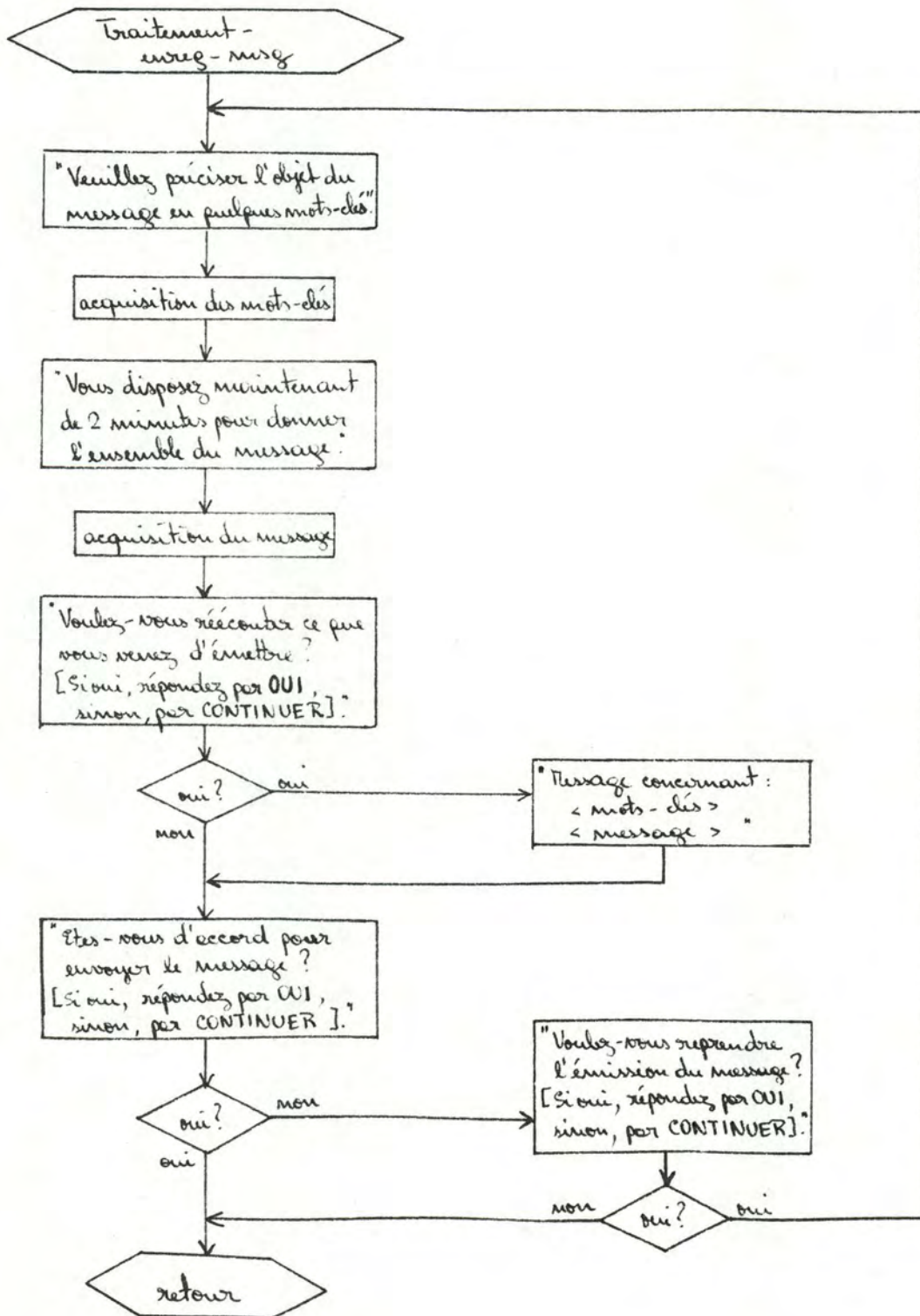
(Schéma suite)



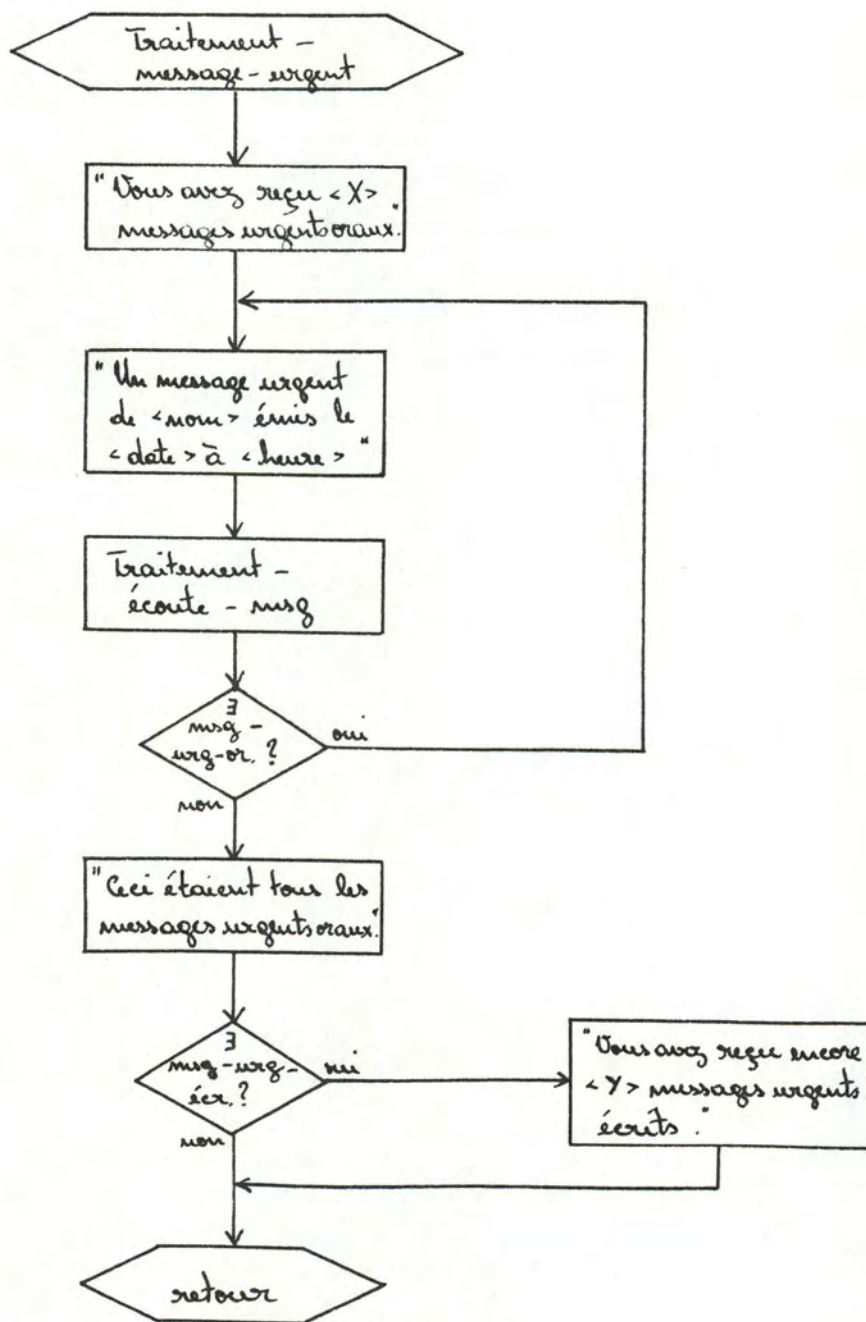
(Schéma suite)



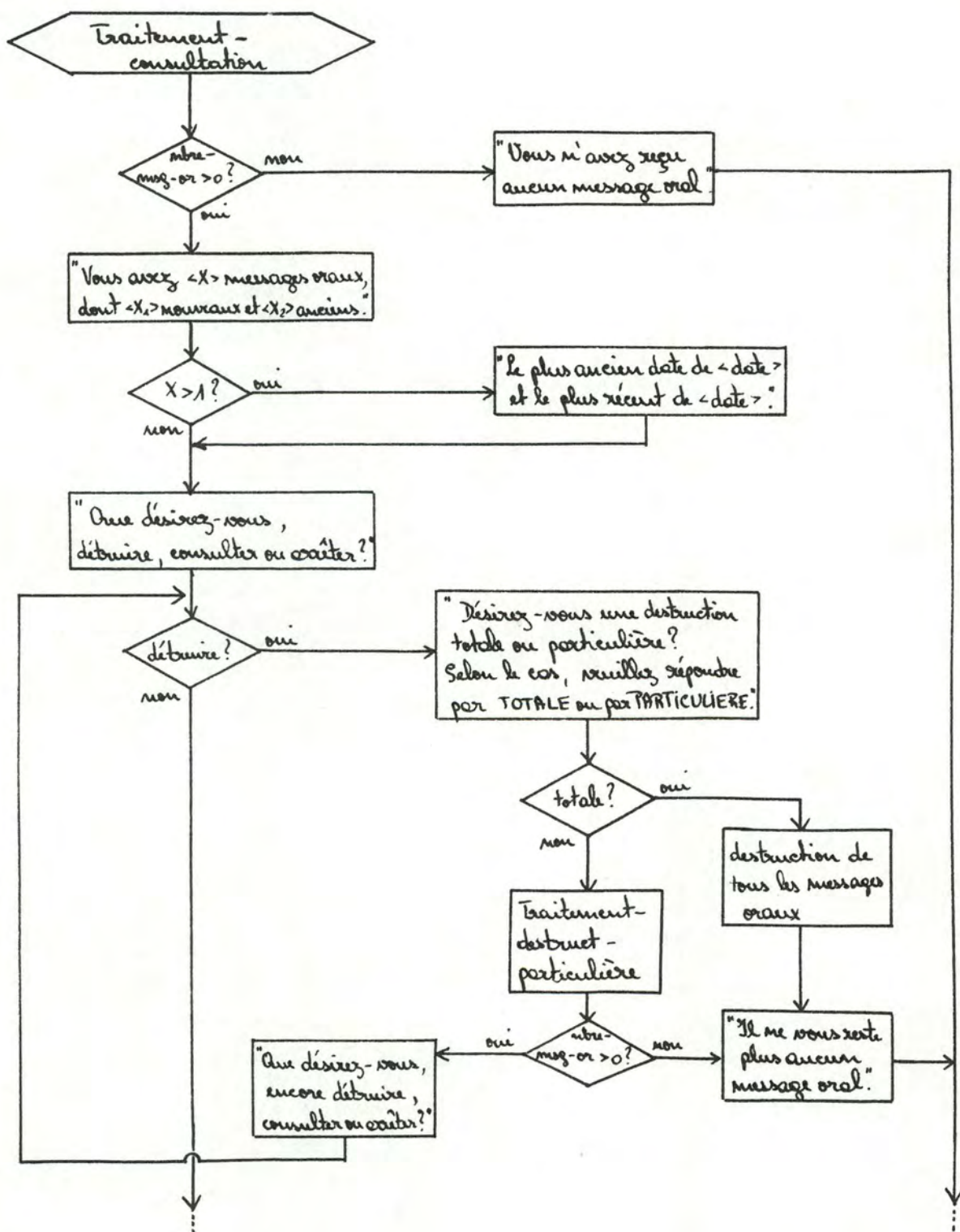
(Schéma suite)



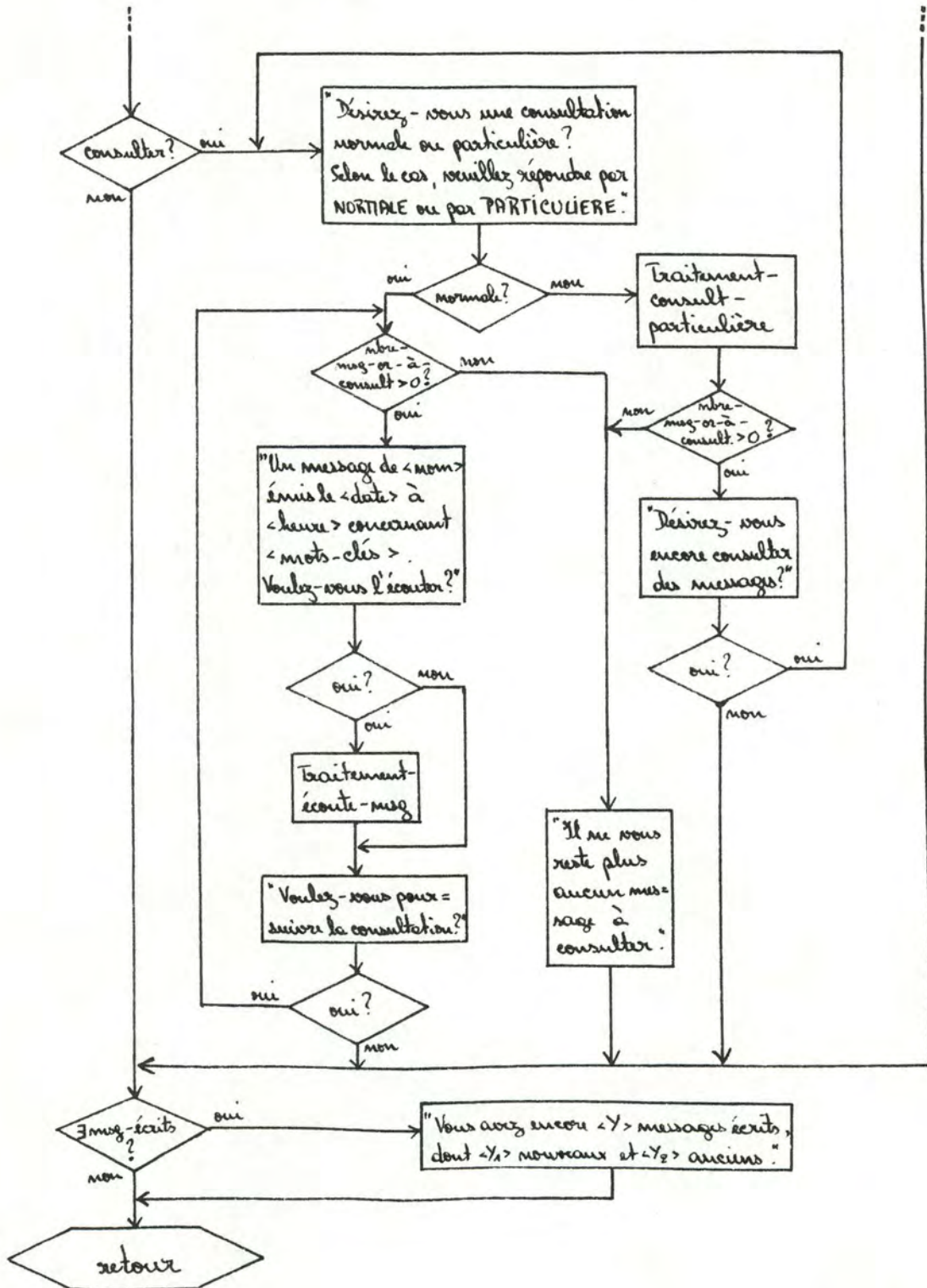
(Schéma suite)



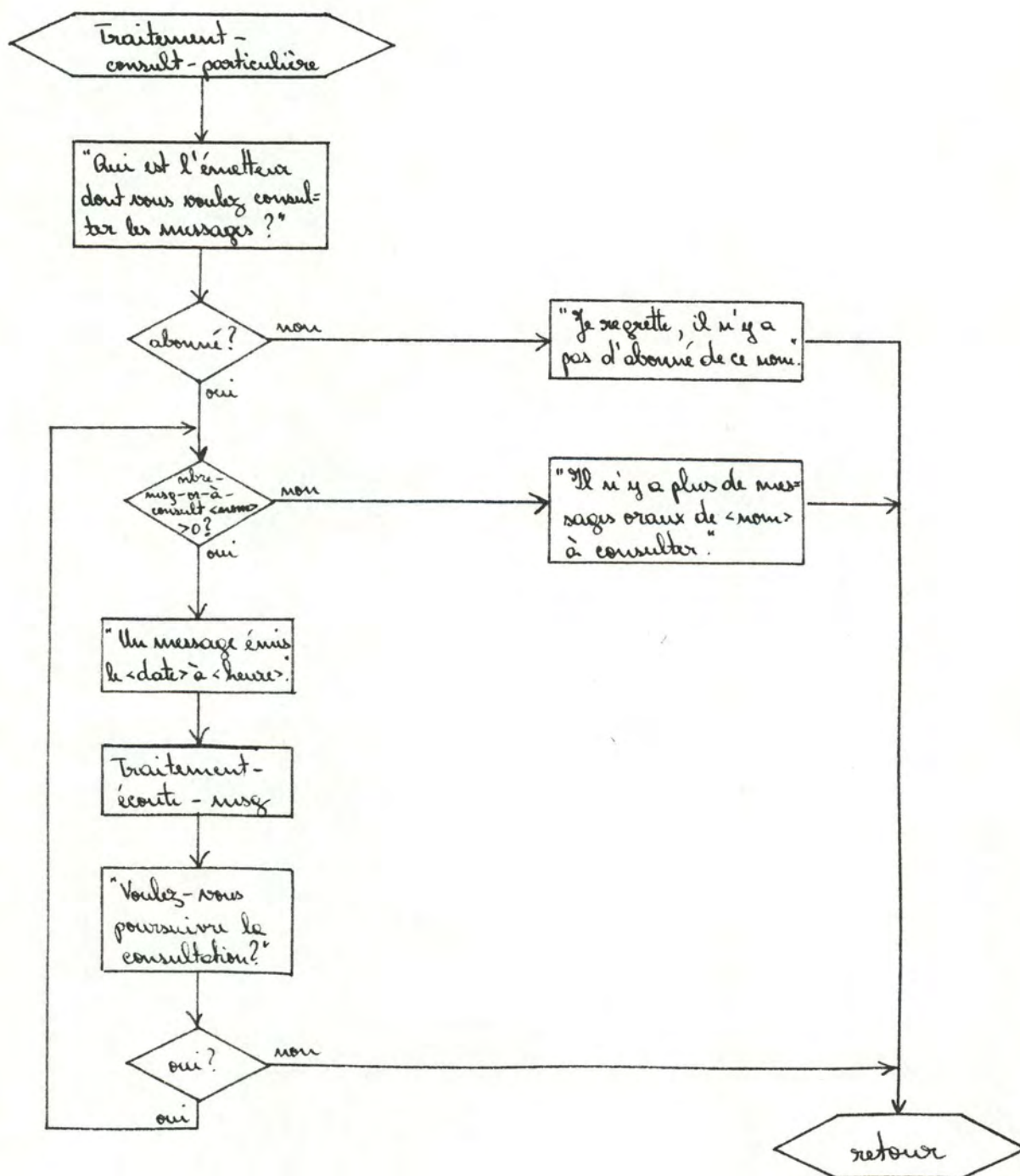
(Schéma suite)



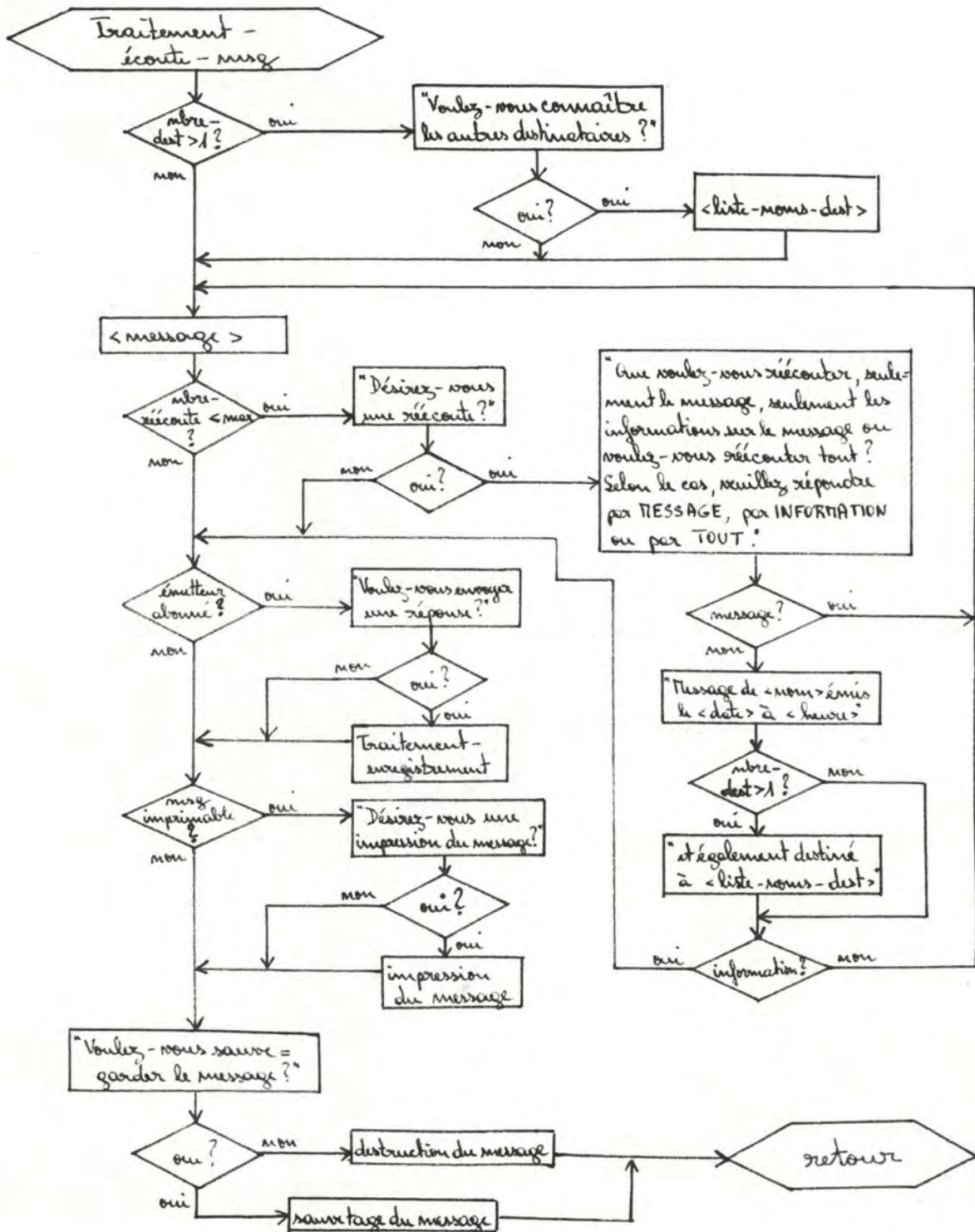
(Schéma suite)



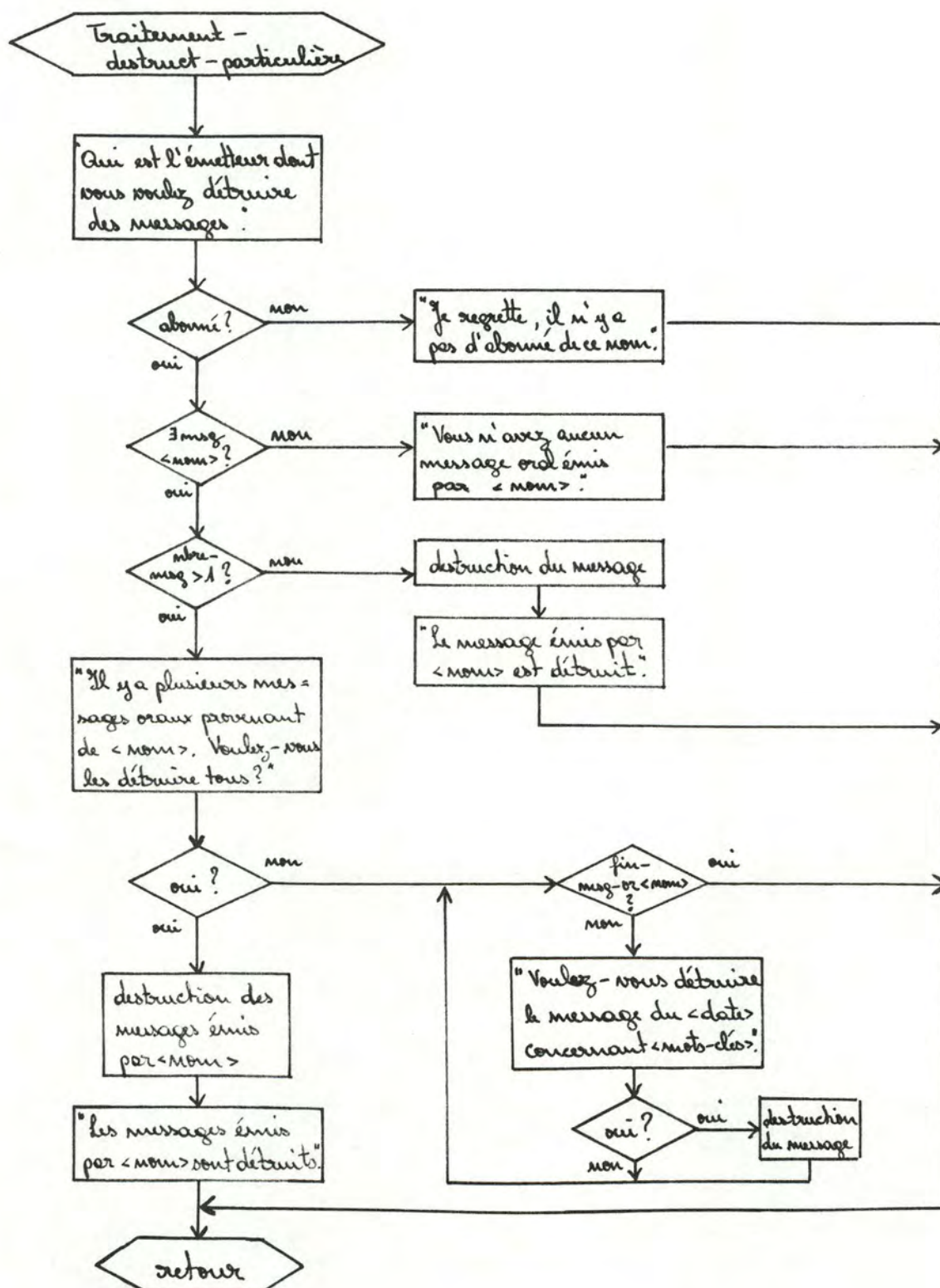
(Schéma suite)



(Schéma suite)



(Schéma suite)



ANNEXE B

COMPARAISON AVEC LE SYSTEME D'AUDIO MESSAGERIE D'IBM

Fig. 12. Tableau comparatif

Composants	Le système IBM éléments particuliers	Eléments communs	MVOC éléments particuliers
Utilisateurs	L'abonné peut définir un code pour: <ul style="list-style-type: none"> - sa secrétaire qui peut émettre des messages à n'importe quel abonné, consulter les informations sur les messages et changer certaines informations sur l'abonné; - sa famille et - un invité qui ont alors le droit d'émettre des messages destinés à l'abonné qui les a définis. 	L'abonné a accès à tous les services, à savoir: <ul style="list-style-type: none"> - la personnalisation; - l'enregistrement; - la réception; - la lecture; - la transmission. 	Le non-abonné, personne totalement inconnue du système peut émettre des messages à n'importe quel abonné.

(Tableau suite)

Composants	Le système IBM éléments particuliers	Eléments communs	MVOC éléments particuliers
Dialogue Utilisateur- Système	<p>L'utilisateur demande les différents services souhaités par l'intermédiaire de codes numériques, introduits à l'aide des touches du clavier téléphonique; il peut:</p> <ul style="list-style-type: none"> - régler le débit des messages; - choisir entre une version complète, une version de base et une version de base assistée; - préciser la langue à utiliser pour les messages-système. 		<p>Le dialogue (oral ou écrit) entre le système et l'utilisateur est un peu rigide du fait que le système a l'initiative et offre à l'utilisateur le choix entre différentes alternatives, par contre ce dialogue a l'avantage d'éviter à l'utilisateur de devoir apprendre par coeur un manuel d'utilisation.</p>
Messages	détails inconnus.		Trois types: <ul style="list-style-type: none"> - oral; - écrit; - mixte (oral et écrit).
Options en émission	Trois options: <ul style="list-style-type: none"> - urgent; - spécial; - normal. 		Deux options: <ul style="list-style-type: none"> - urgent (taille réduite); - normal.
Catégories en stockage	Quatre catégories: <ul style="list-style-type: none"> - nouveau (non écouté); - ancien (écouté); - en attente (à transmettre en différé ou attendant une réponse, une approbation ou un accusé de réception); - envoyé. 		Trois catégories: <ul style="list-style-type: none"> - urgent; - nouveau; - ancien. <p>(l'émetteur ne garde aucune trace des messages enregistrés, ceux-ci étant envoyés immédiatement)</p>

(Tableau suite)

Composants	Le système IBM éléments particuliers	Eléments communs	MVOC éléments particuliers
<p>Accès en lecture</p> <p>Options en protection</p>	<p>Comme a priori, l'abonné peut lire n'importe quel message, les messages peuvent être divisés en cinq classes:</p> <ul style="list-style-type: none"> - non-classifié; - réservé à l'usage interne; - confidentiel; - personnel; - personnel et confidentiel. <p>Comme l'abonné a le droit d'ajouter des commentaires à des messages reçus, l'émetteur peut protéger l'accès à ses messages, on distingue trois niveaux d'accès:</p> <ul style="list-style-type: none"> - non-réglémenté; - commentaire autorisé; - limité à l'écoute. 		<p>Tout abonné ne peut lire que les messages qui lui sont destinés.</p>
<p>Appels</p>	<p>Le système appelle l'abonné:</p> <ul style="list-style-type: none"> - si urgent, au numéro de téléphone ordinaire dès que possible; - si spécial, au numéro de téléphone ordinaire, temporaire ou de nuit; - si normal, au numéro de téléphone ordinaire à des intervalles réguliers. 	<p>L'abonné appelle le système.</p>	

(Tableau suite)

Composants	Le système IBM éléments particuliers	Eléments communs	MVOC éléments particuliers
Personnalisation	<p>L'abonné doit définir:</p> <ul style="list-style-type: none"> - sa position habituelle et éventuellement la position temporaire; - ses paramètres personnels: <ul style="list-style-type: none"> *numéro de téléphone ordinaire, temporaire et de nuit; *codes secrets personnels, de la secrétaire, de la famille et de l'invité; * caractéristiques du dialogue (débit des messages; version du dialogue: complète, de base ou de base assistée; langue des messages-systèmes) 	<p>L'abonné peut prédéfinir des listes d'abonnés destinataires. (listes de diffusion)</p>	<p>L'abonné doit prédéfinir son mot-de-passe oral et écrit.</p>
Enregistrement	<ul style="list-style-type: none"> - de changements dans l'une des parties du message en cours d'enregistrement; - d'un commentaire en tête ou en fin d'un message existant (sans effacement du message); - d'un commentaire après arrêt en cours d'écoute d'un message, à partir de la position d'arrêt et sans effacement. 	<ul style="list-style-type: none"> - de nouveaux messages; - d'un message réponse à celui qu'on vient d'écouter. <p>L'abonné peut en outre:</p> <ul style="list-style-type: none"> - réécouter le message enregistré; - revenir sur cet enregistrement en éliminant le message qu'il vient d'enregistrer. 	<p>L'abonné ou le non-abonné enregistrent d'abord l'objet du message en quelques mots-clefs avant d'enregistrer le message.</p> <p>L'abonné peut reprendre l'enregistrement à partir du début.</p>

(Tableau suite)

Composants	Le système IBM éléments particuliers	Eléments communs	MVOC éléments particuliers
Réception	<ul style="list-style-type: none"> - l'heure de transmission et de réception; - la classification du message; - l'action demandée par l'émetteur (réponse, approbation ou accusé de réception), avec une date limite ou, si l'action est déjà effectuée, l'heure correspondante. 	<p>La réception permet à l'abonné de choisir les messages qu'il veut écouter en lui fournissant des informations sur les messages, à savoir:</p> <ul style="list-style-type: none"> - l'émetteur; - les autres destinataires. 	<ul style="list-style-type: none"> - la date et l'heure d'émission; - l'objet du message (en quelques mots-clefs) en cas de consultation normale.
Gestion de la consultation	<p>Les messages sont stockés en 4 catégories, l'abonné peut:</p> <ul style="list-style-type: none"> - sélectionner une catégorie; - sélectionner la catégorie suivante; - commencer au début d'une catégorie; - sauter au dernier message d'une catégorie; - avancer ou reculer d'un message. 		<p>L'abonné a le choix entre:</p> <ul style="list-style-type: none"> - consultation normale: dans l'ordre croissant des dates d'émission (nouveaux avant anciens); - consultation particulière des messages provenant d'un abonné donné: même ordre; <p>En outre il y a la consultation des messages urgents qui est prioritaire et obligatoire: même ordre.</p>

(Tableau suite)

Composants	Le système IBM éléments particuliers	Eléments communs	MVOC éléments particuliers
Lecture	L'abonné peut: <ul style="list-style-type: none"> - lire les messages d'un fichier de messages quelconque (sauf si les messages sont protégés); - contrôler la vitesse et le volume de l'émission; - changer de position dans le message (revenir au début ou avancer vers la fin du message; reculer ou avancer par multiples de 3 secondes de parole environ; arrêter et relancer la lecture). 	L'abonné peut: <ul style="list-style-type: none"> - réécouter le message qu'il vient d'enregistrer; - réécouter le message consulté et l'information sur ce message; - renvoyer à l'émetteur un message réponse. 	L'abonné peut seulement consulter les messages qui lui sont destinés;
Transmission	Transmission différée et immédiate (avec les trois options urgent, spécial et normal). L'abonné doit définir: <ul style="list-style-type: none"> - la classification, le niveau d'accès et période de retention (durée de conservation) du message; - l'action qu'il demande au destinataire avec la date limite (réponse, approbation ou accusé de réception). 	Transmission de messages vers tout abonné, y compris soi-même. L'abonné doit définir les destinataires soit en les énumérant soit en donnant le nom d'une liste de diffusion prédéfinie.	Transmission immédiate (de messages urgents et normaux).

(Tableau suite)

Composants	Le système IBM éléments particuliers	Eléments communs	MVOG éléments particuliers
Gestion- message	<p>L'émetteur peut spécifier d'avance une durée de conservation du message.</p> <p>Le destinataire peut en spécifier également une qui doit être inférieure à celle spécifiée éventuellement par l'émetteur.</p> <p>L'émetteur peut détruire des messages en attente.</p>	Après consultation d'un message, le destinataire peut l'éliminer.	<p>Le destinataire peut procéder à une destruction totale, partielle ou particulière des messages reçus.</p> <p>A condition que le message consulté soit imprimable, le destinataire peut en demander une impression.</p>

BIBLIOGRAPHIE

- [ADANS-84] Jean-Paul Adans, *Vers une intégration des services bureautiques ?* Actes des deuxièmes journées de réflexion sur l'Informatique, Namur, 30-31 août et 1 septembre 1984
- [ADANT-83] Jean-Michel Adant, *Transmission numérique de la parole à 2400 bps*, Actes des journées ATHENA sur 'L'analyse, la synthèse et la reconnaissance de la parole', Louvain-la-Neuve, 17 mai 1983
- [ASEO-84] Joseph Aseo, *Speech Compression Brings Voice Messaging Down to Earth*, Computer Design, July 1984, pp35-37
- [BAKER-84] Janet M. Baker, *Voice-Store-and-Forward: The Voice Message Medium*, Speech Technology, Aug./Sept. 1984, p44
- [BALL-81] Eugene Ball, Philip J. Hayes et D. Raj Reddy, *Breaking the Man-Machine Communication Barrier*, Computer, vol.14 nr.3, March 1981, pp19-30
- [BENNANI-84] Mohammed Bennani, *Adjonction d'entrées-sorties vocales et graphiques en EAO: amélioration du dialogue et nouveaux champs d'applications*, Thèse de Troisième Cycle, Université de Nancy I, 1984
- [BERGER-84] Pierre Berger et Robert Mahl, *Les réseaux de la recherche*, TSI, vol.3 nr.6 1984, pp457-473
- [BERLEUR-83] Jacques Berleur, *Informatique et Rationalité*, Cours de Méta-informatique de deuxième Licence et Maîtrise en Informatique, Facultés Universitaires de Namur, 1983/84
- [BOITE-83] R. Boite, *Introduction au traitement de la parole*, Actes des journées ATHENA sur 'L'analyse, la synthèse et la reconnaissance de la parole', Louvain-la-Neuve, 17 mai 1983
- [COLE-82] Randy Cole, *Packet Voice: When it Makes Sense*, Speech Technology, Sept./Oct. 1982, pp52-61

- [DEAN-82] M. Dean, *How a Computer Should Talk to People*, IBM Systems Journal, vol.21 nr.4 1982, pp424-453
- [DIMARTINO-84] Joseph Di Martino, *Contribution à la reconnaissance globale de la parole: mots isolés et mots enchaînés*, Thèse de Docteur Ingénieur, Université de Nancy I, 1984
- [DRINKWATER-84] Larry Drinkwater, *Voice Processing: An Emerging Computer Technology*, Speech Technology, Aug./Sept. 1984, pp50-54
- [FAWE-83] Fawe, *Représentation de la parole par la densité des passages par zéro (ou fréquence instantanée) en vue de la transmission, du traitement et de la reconnaissance*, Actes des journées ATHENA sur 'L'analyse, la synthèse et la reconnaissance de la parole', Louvain-la-Neuve, 17 mai 1983
- [HANSON-84] Robert J. Hanson, *The DSC-2000 VoiceServer System*, Speech Technology, Aug./Sept. 1984, pp55-65
- [HATON-84-a] Marie-Christine Haton, Jean-Paul Haton et Jean-Marie Pierrel, *L'édition de sons et ses applications*, Rapport interne du CRIN, Nancy, 1984
- [HATON-84-b] Jean-Paul Haton, *Compréhension de la parole et intelligence artificielle: l'état des recherches*, Actes du séminaire GRECO-GALF sur le 'Dialogue homme-machine à composante orale', Nancy, 11-12 octobre 1984, pp1-60
- [HATON-84-c] Marie-Christine Haton, *Analyse et rééducation de la production vocale assistées par ordinateur*, Thèse d'Etat, Université de Nancy I, 1984
- [HAYES-83] Philip J. Hayes et D. Raj Reddy, *Steps Toward Graceful Interaction in Spoken and Written Man-Machine Communication*, International Journal of Man-Machine Studies, vol.19 1983, pp231-284
- [HELMREICH-81] Reinhard Helmreich, *Acceptance Research Strategies in Computer Message Systems*, in 'Computer Message System', R.P. Uhlig ed., North Holland Publishing Company, IFIP 1981, pp447-452
- [IBM-82] Compagnie IBM France, *Système de l'Audio-Messagerie IBM*, Guide de l'abonné, 1982

- [LEA-80-a] Wayne A. Lea, *Speech Recognition: Past, Present and Future*, in 'Trends in Speech Recognition', W.A. Lea ed., Prentice Hall, New Jersey 1980, pp39-98
- [LEA-80-b] Wayne A. Lea, *The Value of Speech Recognition Systems*, in 'Trends in Speech Recognition', W.A. Lea ed., Prentice Hall, New Jersey 1980, pp3-18
- [LEVINSON-80] S. Levinson et M. Liberman, *La reconnaissance de la parole par ordinateur*, Pour la Science, Juin 1981, pp88-101
- [LYMAN-84] Guy C. Lyman, *Voice Messaging Comes of Age*, Speech Technology, Aug./Sept. 1984, pp45-49
- [MAGUIRE-82] Martin Maguire, *An Evaluation of Published Recommendations on the Design of Man-Computer Dialogues*, International Journal of Man-Machine Studies, vol.16 1982, pp237-261
- [MICHELON-83] Jean-Luc Michelon, *Système de télématique vocale de la mairie de Blagnac*, Rapport interne, Thomson TITN (Traitement de l'Information Techniques Nouvelles), 1983
- [MOUSEL-83] Pierre Mousel, *Utilisation de contraintes syntaxico-sémantiques dans un système de compréhension du discours continu*, Mémoire de Licence et Maîtrise en Informatique, Facultés Universitaires de Namur, 1983
- [NEWMAN-81] Julian Newman, *Human Factors Requirements for Managerial Use of Computer Message Systems*, in 'Computer Message System', R.P. Uhlig ed., North Holland Publishing Company, IFIP 1981, pp453-465
- [PIERREL-81] Jean-Marie Pierrel, *Etude et mise en oeuvre de contraintes linguistiques en compréhension automatique du discours continu*, Thèse d'Etat, Université de Nancy I, 1981
- [PIERREL-85] Jean-Marie Pierrel, *Aspects of Man-Machine Voice Dialogue*, Actes du séminaire INRIA sur les 'Principes de la communication homme-machine: Parole, Vision et Langage Naturel', Versailles, 28 mai - 7 juin 1985, pp253-278
- [ROHLFS-81] Sabine Rohlfs et Hilary Williamson, *The User Interface Design Process*, in 'Computer Message System', R.P. Uhlig ed., North Holland

Publishing Company, IFIP 1981, pp427-445

- [SAMBUR-80] Marvin R. Sambur et Richard A. Smith, *Hypothesizing and Verifying Words for Speech Recognition*, in 'Trends in Speech Recognition', W.A. Lea ed., Prentice Hall, New Jersey 1980, pp139-165
- [SAMBUR-82] Marvin R. Sambur, *Speech Algorithm Advances Promise Toll-Quality Medium-Band Digitized Speech*, *Speech Technology*, Sept./Oct. 1982, pp22-34
- [SCAPIN-85] Dominique Scapin et Jean-Claude Sperandio, *Aspects ergonomiques de la communication homme-machine*, Actes du séminaire INRIA sur les 'Principes de la communication homme-machine: Parole, Vision et Langage Naturel', Versailles, 28 mai - 7 juin 1985, pp97-109
- [SCHALK-83] Thomas B. Schalk et Elizabeth L. Van Meir, *Terminals, Listen up, Speech Recognition Is a Reality*, *Computer Design*, Sept. 1983, pp97-102
- [SCHMANDT-84] Chris Schmandt, *Speech Synthesis Gives Vocal Access to an Electronic Mail System*, *Speech Technology*, Aug./Sept. 1984, pp66-68
- [VANLIEFFERINGE-83] Van Liefferinge, *Le synthétiseur de parole DIVA (Digital Vocal Announcer)*, Actes des journées ATHENA sur 'L'analyse, la synthèse et la reconnaissance de la parole', Louvain-la-Neuve, 17 mai 1983
- [WELLEKENS-83] Christian J. Wellekens, *Un aperçu général sur les algorithmes de reconnaissance de la parole*, Actes des journées ATHENA sur 'L'analyse, la synthèse et la reconnaissance de la parole', Louvain-la-Neuve, 17 mai 1983