

Groupement des Acousticiens de Langue Française

5^{èmes} JOURNEES D'ETUDE

du groupe

“ COMMUNICATION PARLEE ”

Avec la participation de l' AFCET

VOLUME I

Textes des Exposés

Organisées par
le LABORATOIRE D'INFORMATIQUE
POUR LA MECANIQUE
ET LES SCIENCES DE L'INGENIEUR
DU C.N.R.S.

ORSAY
15-17 Mai 1974

L'UNITE A REPOSE VOCALE ICOPHORE V

par D. TEIL *, M. CASTELLENGO **, J. SAPALY **

* Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingenieur, C.N.R.S., - Orsay

** Laboratoires d'Acoustique et d'Electronique, Institut de Mécanique Théorique et Appliquée, - Université Paris VI

Résumé

Dans le cadre de nos études sur la synthèse de la parole, nous avons conçu et réalisé un terminal à réponse vocale de dimensions réduites, au fonctionnement simple et à vocabulaire illimité.

Le message à synthétiser est donné sous forme orthographique. Il est transformé en une suite phonétique qui permet l'assemblage des diphonèmes mémorisés sous forme binaire.

Les règles de synthèse sont implicitement formulées dans la forme temps-fréquence des diphonèmes.

La synthèse est quasi-instantanée. L'appareil peut être relié à n'importe quel ordinateur muni d'une prise télétype.

Il peut également fonctionner de manière autonome.

THE VOCAL REPOSE UNIT ICOPHORE V

Summary

Our speech synthesis studies lead us to design and build a vocal reponse unit. With its small dimensions and its unlimited vocabulary, it is of particularly easy use.

The message to be synthesized is given in orthographic form. It is transformed into phonetic codes which permit to link the memorised diphones.

The synthesis rules are given by the time-frequency pattern of the diphones.

The synthesis is immediate. The device can be connected to any computer which has a teletype input-output. It can also work off-line.

L'UNITE A REPOSE VOCALE ICOPHONE V

par D. TEIL *, M. CASTELLENGO **, J SAPALY **

* Laboratoire d'Informatique pour la Mécanique et
les Sciences de l'Ingénieur, C.N.R.S., - Orsay

** Laboratoires d'Acoustique et d'Electronique, Institut de
Mécanique Théorique et Appliquée, - Université Paris VI

I La Synthèse par Diphonèmes

Des recherches commencées en 1966 ont conduit à la réalisation de deux générateurs de parole synthétique à commande optique: Les ICOPHONES I et II, qui ont permis de vérifier nos hypothèses sur l'utilisation des diphonèmes en synthèse de parole.(1).

La synthèse en temps réel a ensuite été abordée à l'aide de l'ICOPHONE III à commande numérique, couplé à un ordinateur IBM 1130 (2).

Cet ensemble nous a permis de terminer la mise au point du dictionnaire des diphonèmes de la langue française et de résoudre les problèmes informatiques (software et hardware) relatifs à la traduction automatique d'un message écrit en message parlé (3),(4).

Nous avons ensuite réalisé l'ICOPHONE IV qui, couplé à l'IBM 1130, nous a permis d'améliorer la qualité de certains éléments phonétiques par l'utilisation de générateurs de bande de bruits et de commencer une étude sur le timbre et l'intonation.

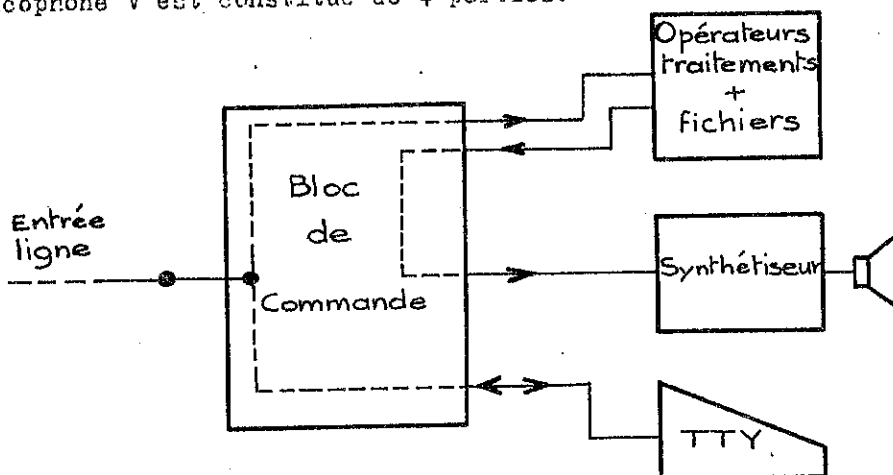
II L'unité a reponse vocale

L'appareillage actuel du L.I.M.S.I.* constitue un puissant outil de recherche en synthèse de la parole.

Parallèlement à la recherche fondamentale qui se poursuit avec cet appareillage, nous avons mis au point, avec l'aide de la Délégation à l'Informatique, une unité à réponse vocale exploitable industriellement: L'ICOPHONE V. Cet appareil transforme en parole parfaitement intelligible un texte fourni sous la forme alphabétique habituelle. Le vocabulaire est illimité, la réponse immédiate. Le débit requis sur la ligne de commande est minime (de l'ordre de 50 bauds), et l'appareil peut fonctionner soit de manière autonome en liaison avec un simple télétype, soit comme un terminal d'ordinateur. Dans ce dernier cas la connexion est effectuée au moyen de la sortie "télétype" de l'ordinateur.

III Description d'ensemble

L'Icophone V est constitué de 4 parties:



Le bloc de commande constitue la partie qui a pour rôle la gestion des différents modules. C'est l'unité centrale d'un mini-calculateur (ORDOPROCESSEUR), dont la microprogrammation a été réalisée en collaboration avec TITN.

Les opérateurs de traitement assurent les fonctions suivantes:

- Transformation du message alphabétique en une suite de symboles phonétiques codés,
- Reconstitution du squelette informatif du message par assemblage de diphonèmes schématisés extraits du dictionnaire,
- Edition de cette configuration binaire sur les oscillateurs de l'ICOPHON.

Le synthétiseur est une réplique simplifiée de l'ICOPHON II. La fréquence du fondamentale est fixe (100 Hz); la vitesse d'élocution et l'amplitude globale du message sont réglables manuellement.

Une machine à écrire du genre télétype assure le dialogue avec l'ordinateur, ou avec le synthétiseur en fonctionnement autonome.

IV Réduction du dictionnaire

Le dictionnaire de départ ayant servi à la synthèse de l'ICOPHON III est composé de 627 diphonèmes divisés dans le temps en 20 événements. (4). Le rangement en mémoire de ce lexique complet utilise 37680 mots de 16 bits.

La nécessité de faire fonctionner l'ICOPHON V en temps réel nous a conduit à utiliser des mémoires mortes. Pour minimiser le coût de l'appareillage, nous avons cherché à réduire au minimum l'encombrement du dictionnaire.

Les premiers essais de réduction ont porté sur la discrétisation du temps. Les éléments phonétiques ont été découpés selon divers pas de temps compris entre 4 et 40 ms par événement. Les tests d'intelligibilité ont montré que l'optimum se situait aux environs de 10 ms. Notre choix s'est donc porté sur une discrétisation de 8 événements par diphonème ce qui correspond sensiblement à cet optimum.

Les essais suivants ont consisté à regrouper des commandes d'oscillateurs pour réduire l'information fréquentielle. Pour que l'opération soit rentable il fallait passer de 46 à 32 commandes. Un tel résultat ne pouvait être obtenu qu'en changeant la distribution fréquentielle des oscillateurs ce qui conduirait à refaire une étude complète des diphonèmes et du synthétiseur.

Dans l'étape suivante, nous avons classé les diphonèmes en fonction de leur propriété de réversibilité pour ne ranger en mémoire qu'une transition sur deux. Certains diphonèmes sont entièrement réversibles, c'est le cas des éléments voyelle - voyelle et des éléments consonne - voyelle où la consonne est /m/, /n/, /l/, /j/.

Pour les autres éléments consonne-voyelle nous avons considéré que la transition vocalique est réversible ce qui a donné des demi-diphonèmes. Les consonnes sont reconstituées dans l'autre moitié. Pour simplifier nous utilisons les mêmes consonnes fricatives quelle que soit la voyelle. Les consonnes occlusives ont été classées en trois catégories suivant qu'elles sont associées à une voyelle qualifiée de grave (/o/, /ɔ/, /u/, /ɔ̃/), moyenne (/a/, /œ/, /ã/, /ẽ/), ou aigüe (/i/, /y/, /e/, /ɛ/). Dans les consonnes occlusives nous n'avons mis en mémoire que l'attaque définie sur l'événement. Le "silence" qui précède (plus ou moins long, voisé ou non suivant la consonne) est introduit au moment de l'émission sur le synthétiseur.

Les éléments voyelle-consonne occlusive sont formés par la transition vocalique suivie d'un "silence" voisé ou non de 4 événements.

Tous les événements ont été comparés entre-eux et nous n'avons conservé en mémoire que les événements différents.

Le lexique des formes phonétiques, composé d'un dictionnaire de phonèmes (12 voyelles, 3 fois 6 consonnes occlusives et 6 consonnes fricatives), d'un dictionnaire de diphonèmes entiers et d'un dictionnaire de demi-diphonèmes, occupent 4111 mots de 16 bits dans l'élément mémoire de l'ICOPHONE V, et la table des identificateurs, 371 mots de 16 bits.

V Réalisation de l'ICOPHONE V

1 Les opérateurs de traitements et les fichiers:

Les fonctions à réaliser les plus importantes en volume sont la traduction phonétique, le rythme et l'adressage des formes phonétiques. La logique d'adressage assure la reconstitution des spectres à partir d'un algorithme de pointage dans une table et du lexique des formes phonétiques.

L'analyse de ces fonctions nous a conduit à utiliser un mini-ordinateur à programme figé du type ordoprocasseur qui répond parfaitement au problème:

- Par sa structure "bus" qui permet de connecter facilement toutes sortes d'opérateurs, en particulier l'opérateur "table" contenant le lexique des formes acoustiques, les opérateurs d'entrée (contrôleur ligne et contrôleur clavier) et de sortie (un ou plusieurs contrôleurs de synthétiseur).

- Par sa souplesse de code qui permet d'introduire des mini-instructions spécialisées en particulier pour l'algorithme de recherche dans la table.

- Par son coût économique qui permet d'obtenir un faible prix de revient.

2 Le Synthétiseur:

Il comprend sous forme de modules miniaturisés:

- 44 oscillateurs sinusoidaux couvrant linéairement la gamme 100 Hz 4400 Hz,

- 3 générateurs de bande de bruit répartis dans la bande 1500 à 6000 Hz,

- 1 mélangeur

- 1 chaîne d'amplification et d'écoute.

Les oscillateurs fonctionnent en tout ou rien sur ordre de la logique de sortie. La fréquence et l'amplitude du signal de sortie sont réglés une fois pour toute, la constante de temps de la commande est ajustable en modifiant des composants discrets accessibles sur la carte, ce qui permet d'atténuer les fronts d'attaque des signaux. Les oscillateurs ne sont pas modulables en fréquence: on se limite à une voix voisée sans intonation. Les générateurs de bruit fonctionnent également en tout ou rien. On peut ajuster une fois pour toutes la fréquence centrale, la largeur de bande et l'amplitude du signal de sortie.

VI Les caractéristiques du terminal ICOPHONE V

Caractéristiques physiques:

- dimensions approximatives du terminal: 50 x 60 x 40 cm,
- organe de dialogue: machine à écrire,
- software intégré,
- réglages manuels de la vitesse d'élocution et de la puissance de sortie.

Caractéristiques de fonctionnement:

- autonome, en liaison avec un autre terminal du même type ou en liaison avec un ordinateur à la place d'un télétype,
- messages transmis en clair en français ou en codes phonétiques par bloc de 60 caractères maximum,
- réponse quasi immédiate après la réception du dernier caractère,
- décodage automatique des nombres entiers positifs ou négatifs inférieurs à 1 milliard,
- possibilité de répétition programmée de l'édition de 2 à 9 fois,
- commutateur pour fonctionnement en test (autonome),
- commutateur de sortie: vocale, imprimée, ou vocale plus imprimée.

REFERENCES

- (1) - E. LEIPP, J.S. LIENARD, M. CASTELLENGO, J. SAFALY, D. TEIL, A. CALINET, M. MLOUKA - Colloque sur la parole. Bulletin du Groupe d'Acoustique musicale de l'Université de Paris VI n° 53, janvier 1971.
- (2) - J. QUINIO, D. TEIL - La synthèse de la parole par ordinateur à partir de digrammes phonétiques. Revue d'Acoustique t.3, n° 9, 1970.
- (3) - E. LEIPP, M. CASTELLENGO, J.S. LIENARD, J. QUINIO, J. SAFALY, D. TEIL - Générateur synthétique de parole. Brevet ARVAN n° 1502936, février 1971.
- (4) - J.S. LIENARD, D. TEIL - Les éléments phonétiques et la traduction automatique du message écrit en message parlé - Automatisation n° 10. Octobre 1970.