

RÉSUMÉ

Le projet Lipcom est un effort de recherche et de développement d'IBM France. Son but est de faciliter la réception de la parole par les personnes sourdes face à un interlocuteur entendant. Pour ce faire, Lipcom effectue un sous-titrage phonétique de la parole du locuteur entendant. Cette information phonétique, combinée à d'autres sources d'informations (comme le mouvement des lèvres et l'expression du visage du locuteur ainsi que celles apportées par les prothèses auditives) devrait permettre à la personne sourde de comprendre sans ambiguïté son interlocuteur. Lipcom cherche à reconnaître la parole spontanée en temps réel, sans aucune contrainte de vocabulaire ni de syntaxe. De ce fait, la technique informatique employée est très différente de celle des systèmes de dictée automatique. Depuis 5 ans, Lipcom est expérimenté dans une école d'enfants sourds. La première phase de l'expérimentation a déjà montré l'intérêt de cette recherche : le sous-titrage effectué par Lipcom améliore les performances de compréhension des enfants dans des situations de lecture labiale ambiguës. La deuxième phase de l'expérimentation consiste aujourd'hui à améliorer les résultats de reconnaissance du système et à passer d'un prototype monolocuteur à un système multilocuteur afin de permettre l'utilisation de Lipcom en continu dans le cadre d'une salle de classe. Cet objectif est ambitieux et il faudra encore du temps avant que Lipcom puisse être utilisé plus largement.

MOTS CLÉS

Surdité - Réception de la parole - Aide technique - Reconnaissance de la parole - Sous-titrage - Recherche.

LIPCOM, prototype d'aide automatique à la réception de la parole par les personnes sourdes

par Audrey COURSAULT-MOREAU et Francis DESTOMBES

Audrey COURSAULT-MOREAU
Francis DESTOMBES
IBM-France, Scc 2485
RDS 5ème étage
68/76 quai de la rapée
75592 Paris cedex 12
e-mail :
audrey_coursant@fr.ibm.com

SUMMARY : LIPCOM, a prototype to help the Deaf receive speech.

The Lipcom project is a research and development effort by IBM France. Its goal is to make speech reception easier for the deaf. Lipcom performs a phonetic subtitling of a hearing speaker. This phonetic information, combined with other sources of information (such as lip movements, facial expression, and hearing aids) should help a deaf person to better understand a hearing speaker without any ambiguity. Lipcom performs real time recognition of spontaneous speech. The vocabulary is wholly unlimited and syntax is unconstrained. The recognition technique used by Lipcom is indeed totally different from dictation systems. Lipcom has been experimented for 5 years in a school for deaf children. The first stage of this experimentation has already proven the interest of this research : Lipcom's subtitling improves children's performances in understanding sentences which are very ambiguous in speechreading alone. The second stage of this research consists in improving speech recognition results. For the time being, the system is a single-speaker one, requiring adaptation to the voice of the speaker. In order to make the system more user friendly, Lipcom should be speaker-independent. This would also allow to better evaluate the benefit for the reception of speech using the system all day long in a class. This objective is ambitious and more research time is required before Lipcom may become commercially available.

KEY WORDS

Deafness - Speech perception - Technical Aid - Speech recognition - Subtitling - Research.

INTRODUCTION

Les difficultés de réception de la parole par les personnes déficientes auditives profondes ont motivé, depuis fort longtemps, de nombreuses recherches pour faciliter cet aspect de la communication. L'élaboration du Cued Speech (en français L.P.C., Langage Parlé Complété) par Cornett* en témoigne. Ce code manuel ayant fait ses preuves, son inventeur a voulu poursuivre ses investigations en vue de développer un système de Cued Speech automatique : l'Autocuer*, mais cette recherche, très ambitieuse pour l'époque, n'a jamais abouti.

L'idée de créer une aide automatique comme aide à la réception de la parole a été reprise un peu plus tard par Krasner & al.*, Huggins & al.** et Russel*** avec le projet VID-VOX, qui devait être un système de reconnaissance de la parole continue dont le principe était de transcrire phonétiquement et en temps réel la parole d'un locuteur. VID-VOX, tout comme l'Autocuer, était sans doute trop en avance sur les possibilités technologiques de l'époque, c'est pourquoi ces travaux n'ont pas été poursuivis.

Plus récemment, le CNET (Centre National d'Etudes des Télécommunications) a tenté de développer un système combinant le visiophone à la visualisation d'indices consonantiques de type acoustique permettant de lever certaines ambiguïtés de la lecture labiale*. Le champ d'application de ce système serait limité à l'utilisation des techniques de télécommunication.

Ces différents projets n'ayant pas abouti, force est donc de constater qu'il n'existe aujourd'hui aucune aide technique entièrement automatique pour faciliter la réception de la parole par le canal visuel ; c'est précisément dans cette perspective que s'inscrit le projet Lipcom.

*1967

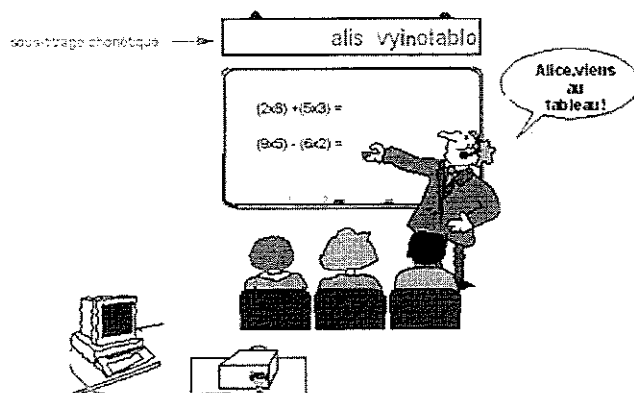
*Cornett & al. 1977

*1985 ** 1986 *** 1986

* Sokol & al. 1995 et Sokol et Mercier 1995

PRINCIPE DE LIPCOM

Une des premières applications de Lipcom est son utilisation dans le cadre de l'enseignement. Typiquement, on imagine une classe, dans laquelle il y a des enfants sourds. Le professeur entendant est équipé d'un micro HF. Tout ce qu'il dit est analysé par Lipcom qui effectue une transcription phonétique en temps réel. Le sous-titrage résultant est projeté sur un écran situé au dessus du tableau, ceci, au moyen d'un vidéo-projecteur relié à l'ordinateur (cf. schéma ci-après).



POURQUOI UN SOUS-TITRAGE PHONÉTIQUE ET NON UNE RECONNAISSANCE ORTHOGRAPHIQUE ?

Il existe en effet des systèmes de dictée automatique qui effectuent une reconnaissance orthographique de la parole, pourquoi ne pas les utiliser ?

Tout simplement parce que ces systèmes présentent un certain nombre de caractéristiques qui ne conviennent pas à l'application que nous visons avec Lipcom.

En effet, les dictées automatiques exigent, comme leur nom l'indique, que l'on dicte des phrases et que l'on énonce toutes les marques de ponctuation pour que la reconnaissance fonctionne correctement. Ces systèmes comportent en fait des modèles statistiques de langage fondés sur des probabilités d'occurrence de séquence de trois mots et cette caractéristique contraint, d'une certaine façon, la syntaxe employée par le locuteur. Par ailleurs, les dictées automatiques utilisent toutes un lexique ou vocabulaire limité aujourd'hui à environ 60.000 "mots" qui sont en réalité 60.000 formes fléchies (ex. : les mots "parle", "parler", "parlons"... sont autant de formes fléchies). Le vocabulaire est donc parfaitement adapté pour des champs lexicaux spécifiques (correspondance commerciale, radiologie...), mais non pour un discours totalement ouvert et spontané tel que nous l'envisageons avec Lipcom.

Nous voulons un système capable de faire une reconnaissance de la parole dans une situation banale de communication : un professeur qui s'adresse à ses élèves. Pour cela Lipcom doit reconnaître la parole spontanée (avec des hésitations, des phrases inachevées...), c'est-à-dire sans ponctuation, sans aucune contrainte syntaxique et avec un vocabulaire illimité. Le meilleur moyen de satisfaire ces exigences est d'effectuer une reconnaissance phonétique qui s'appuie essentiellement sur une analyse acoustique de la parole.

Lipcom effectue simplement un calcul de spectres acoustiques et compare ceux-ci à de très nombreux spectres stockés en mémoire pour décider d'afficher tel ou tel "phonème"*. Il n'utilise aucun lexique, ni aucune statistique sur les probabilités d'occurrence de mots.

Nous souhaitons revenir sur cette notion de "phonème" et de sous-titrage phonétique pour souligner le fait que la transcription de Lipcom n'est pas celle de l'alphabet phonétique international. Le sous-titrage de Lipcom est le résultat de nombreuses expériences auprès des enfants sourds qui ont contribué très activement à son élaboration. En voici quelques exemples :

vwasileisoutitrajéfèktuéparlipkôm

onprononstoutléétrkiaparèstisiânwar*

lorskeudeulètraparèstisiangri* èiseprononskom|nseulson

pourelzânfan

lantrénmenkikonsistalirleisoutitraj

étinvéritablejeû

* en réalité, Lipcom les affiche en vert, et en bleu, les digraphes qui doivent être prononcés comme un seul son.

La familiarisation des enfants à cette notation est très rapide : quelques séances d'entraînement suffisent (représentant au total entre 4 et 6 heures). Cet entraînement comporte essentiellement des exercices de correspondance sous-titrage/phonie/graphie qui font appel à la conscience phonologique. Le caractère ludique de ces exercices "devinettes" est un élément moteur de cette phase d'entraînement.

Par ailleurs, nous avons constaté que le sous-titrage de Lipcom était très accessible à toute personne sourde ou entendante n'ayant aucune connaissance en phonétique, et cela sans apprentissage particulier.

EXPÉRIMENTATION DU PROTOTYPE

Lipcom est expérimenté dans une école d'enfants sourds (CEOP, Centre Expérimental Orthophonique et Pédagogique, Paris 15ème) depuis près de cinq ans.

Une quinzaine d'enfants ont déjà participé à l'expérimentation : dix sourds profonds

* cf. Coursant-Moreau A. et Destombes F., 1996

de naissance et cinq sourds sévères. Ils étaient âgés de 8 à 13 ans. Pour des raisons d'organisation scolaire, le temps d'expérimentation était limité à une heure par semaine seulement.

Le but de l'expérimentation était de mesurer l'efficacité du système. La méthodologie employée se décomposait généralement en deux temps : la phase d'entraînement, et la phase d'évaluation.

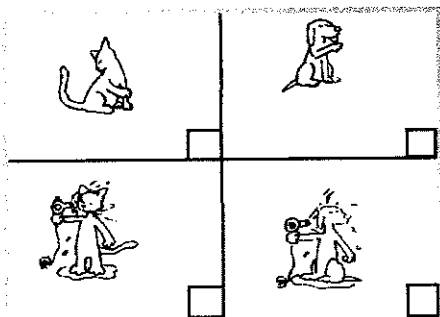
Phase d'entraînement :

- au décodage du sous-titrage effectué par Lipcom
- à l'utilisation des codes phonologiques sous la forme d'exercices de correspondances sous-titrage/phonie/graphie.
- au protocole expérimental (si nécessaire).

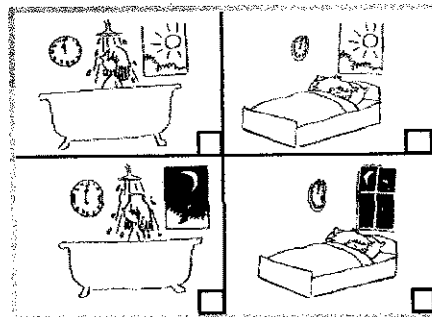
Phase d'évaluation :

- tests en lecture labiale + prothèses auditives
- tests en lecture labiale + prothèses auditives + Lipcom

Les corpus de tests de la phase d'évaluation portaient sur des mots ou des phrases présentant divers degrés d'ambiguïté en lecture labiale. Les protocoles employés étaient des questionnaires à choix multiple, de phrases orthographiées ou de dessins. En voici quelques exemples : l'enfant doit cocher le dessin correspondant à une des phrases prononcée par le locuteur.



Le chat se lèche.
Le chien se lèche.
Le chat se sèche.
Le chien se sèche.



Elle se douche à midi.
Elle se couche à midi.
Elle se douche à minuit.
Elle se couche à minuit.

RÉSULTATS

Les résultats ci-dessous sont des résultats globaux de diverses expériences conduites auprès des différents groupes d'enfants* :

Lecture labiale + prothèses	40 à 60 % d'identifications correctes
Lecture labiale + prothèses + Lipcom	70 à 90 % d'identifications correctes

Ces résultats montrent une amélioration sensible des pourcentages d'identifications obtenus par les enfants et confirment l'intérêt d'une aide telle que Lipcom pour la réception de la parole. Il est important de préciser qu'au moment de ces expériences les taux de reconnaissance phonémique du système variaient entre 75 et 85 %. Ce qui signifie

*pour plus de détails de ces expériences cf. A. Coursant-Moreau 1997.

que même avec un système de reconnaissance encore imparfait, les performances des enfants dans ce type de tâche sont loin d'être négligeables.

PERSPECTIVES

Aujourd'hui, nos efforts portent sur l'amélioration de la reconnaissance de la parole, en terme de taux mais aussi de fiabilité, de résistance au bruit... Parallèlement, nous cherchons à faire une reconnaissance multilocuteurs. Actuellement, le système est monolocuteur et nécessite une phase d'apprentissage contraignante pour la personne qui souhaite l'utiliser. Cette phase d'apprentissage, qui consiste à enregistrer des modèles de la voix du locuteur, est encore très coûteuse en temps. Le passage à un système multilocuteurs devrait permettre de réduire considérablement, voire d'éliminer l'apprentissage et favoriser une plus large utilisation de Lipcom.

QUELLES SONT LES APPLICATIONS POSSIBLES DE LIPCOM ?

Une des premières applications que l'on envisage avec Lipcom est celle de l'enseignement. L'utilisation d'un tel support à la réception de la parole dans ce cadre semble s'imposer. De plus, nous avons de bonnes raisons de croire que cet accès privilégié à la langue orale que représente Lipcom, peut jouer un rôle déterminant dans l'appropriation du langage écrit et oral de l'enfant sourd. Dans une perspective plus large, cette aide technique pourrait favoriser l'insertion professionnelle des adultes sourds.

Autres applications possibles de Lipcom :

- aux guichets des établissements recevant du public (ex. : gares, hôpitaux, ANPE...),
- pour le sous-titrage en direct des émissions télévisées,
- en rééducation orthophonique, avec des personnes ayant un implant cochléaire par exemple.

Le champ des applications de Lipcom est donc très ouvert et toutes ces perspectives sont autant de motivations pour poursuivre cette recherche. Rappelons que Lipcom est encore un prototype et qu'il est nécessaire de le perfectionner avant de diffuser son utilisation.

BIBLIOGRAPHIE

- CORNETT R.O. BEADLES R. et WILSON B. (1977). *Automatic Cued Speech*. In J.M. PICKETT (Ed.), *Research Conference on Speech-Processing-Aids for the Deaf*, Washington, DC, 1982, 224-239.
- CORNETT R.O. (1967). *Cued Speech*, *American Annals of the Deaf*, 112, 3-13.
- COURSAINT-MOREAU A. et DESTOMBES F. (1996). *Lipcom, une aide automatique à la lecture labiale*, XXI^{èmes} JEP, GFCEP, Avignon, 235-238.
- COURSAINT-MOREAU A. (1997). *Un système d'aide automatique à la lecture labiale pour les personnes sourdes profondes : Lipcom. Elaboration et évaluation*, Thèse de doctorat de sciences du langage, Strasbourg.
- HUGGINS A.F.W., HOUDE R.A. ET COLWELL E. (1986). *Vidvox human factors investigation*, Report, BBN Laboratories Incorporated, Cambridge Massachusetts.
- KRASNER M.A., SCHWARTZ R.M., KIMBALL O.A. et PRICE P. (1985). *Development of a Vidvox feasibility system*, Bolt Benarek and Newman technical report 5906, BBN Laboratories, Cambridge MA.
- RUSSEL Y. (1986). *Vidvox final report*, Sensory Aids Foundation, Palo Alto.
- SOKOL R. et MERCIER G. (1995). *Recognition of phonetic features using a neural-fuzzy network*, *NATO Advance Study Institute*, sept. 1995, 137-139.
- SOKOL R., HUIBAN J.Y. et MERCIER G. (1995). *Système d'aide automatique à la lecture labiale*, *Technique Orthopédique Internationale*, 67, 42-44.