

Vers un laboratoire artificiel pour le traitement de la parole

Towards an Artificial Laboratory for Speech Processing

Jean-Claude JUNQUA, Speech Technology Laboratory 3888 State Street, Santa Barbara
California 93105 USA

Résumé

Le travail présenté dans cette correspondance s'appuie sur l'idée que le développement d'algorithmes pour la reconnaissance automatique de la parole nécessite des connaissances variées qui doivent coopérer. Afin d'aider au développement de tels algorithmes, appelés algorithmes hybrides, et de façon plus générale pour le traitement automatique de la parole, des architectures avancées et des environnements intégrés sont souhaitables. Il est aussi souvent nécessaire de représenter les connaissances du domaine parole étudié par l'intermédiaire de structures abstraites, de manipuler des bases de données importantes, et de modéliser, simuler, et évaluer des systèmes de reconnaissance automatique de la parole. Afin de prendre en compte ces différents aspects, nous proposons dans cette correspondance le nouveau concept de laboratoire artificiel

pour le traitement automatique de la parole. Un tel système simule un laboratoire réel et permet de manipuler des données. Il fournit aussi à l'utilisateur un ensemble de fonctions préprogrammées qui facilitent la modélisation et la simulation d'algorithmes de traitement automatique de la parole. Dans cette correspondance nous décrivons brièvement les principaux concepts du prototype de laboratoire artificiel en cours de développement.

Mots clés

Reconnaissance automatique de la parole, laboratoire artificiel, intégration, réutilisation, modélisation, classes, programmation objet.

Summary

It is our belief that speech recognition algorithms can best be handled through the effective and efficient cooperation of multiple knowledge sources. For the development of various types of such algorithms, also called hybrid speech algorithms and, more generally, for speech processing, we need some advanced architectures and speech processing environments. There is also a need to manipulate speech knowledge, through the use of abstract data structures, to process data bases, and to help the modeling, simulation, and evaluation of automatic speech recognizers. To tackle these problems, we propose the new concept of an artificial

laboratory for speech processing. Such a system simulates a real laboratory and allows analysis of data. It also provides a large range of computing facilities which can be used with ease to perform modeling and simulation. In this correspondence, the main concepts of the system are briefly described.

Key-words

Automatic speech recognition, artificial laboratory, integration, reusability, modeling, object oriented classes.

De nos jours un ensemble d'outils mathématiques, d'outils statistiques, et de logiciels d'intelligence artificielle (IA) sont disponibles. Les possibilités permettant de stocker des données ou de les représenter graphiquement évoluent rapidement. Le but principal du travail présenté dans cette correspondance est de rassembler ces outils dans un

environnement intégré en utilisant des techniques d'intelligence artificielle. En d'autres termes, notre objectif est de représenter nos connaissances sur la parole par l'intermédiaire de composants réutilisables (objects), et de faciliter l'intégration et la coopération de différentes techniques dans un environnement intégré unique. Ceci est réalisé par

le couplage de techniques d'intelligence artificielle avec une méthodologie de développement logiciel et de représentation des données, dans le cadre de la modélisation et de la simulation d'algorithmes de traitement automatique de la parole. Compte tenu de l'ensemble des facilités à la disposition du chercheur sur le traitement de la parole et de la souplesse avec laquelle ces facilités peuvent être utilisées, un tel environnement est appelé *laboratoire artificiel* [1]. Un laboratoire artificiel est un concept abstrait que nous avons appliqué, dans le travail présenté, au traitement automatique de la parole. Le système réalisé, appelé StanisLAX (X window based artificial laboratory for the simulation and interactive speech processing), permet :

- de développer un modèle incrémentalement pour un problème donné,
- de simuler et raffiner ce modèle,
- d'évaluer le modèle par comparaison à d'autres modèles existants,
- d'interagir avec la simulation,
- de représenter le modèle graphiquement,
- d'étendre les possibilités de modélisation et de simulation par l'addition de nouveaux outils,
- de manipuler des bases de données,

- de faciliter l'acquisition et la représentation de connaissances sur la parole, et
- de développer des systèmes hybrides de reconnaissance automatique de la parole qui nécessitent la coopération entre plusieurs sources de connaissances.

Grâce à un interface utilisateur élaboré (e.g. fig. 1) StanisLAX peut être utilisé par des experts sur le traitement automatique de la parole mais aussi par des débutants. Il permet de se concentrer sur la définition d'un modèle particulier plutôt que sur l'implémentation de celui-ci. Le développement d'un tel système est né des réflexions suivantes :

- la nécessité d'utiliser des composants réutilisables et des interfaces logiciels standardisés afin d'améliorer la souplesse et l'efficacité dans le développement de logiciels,
- la modularité du signal de parole (décomposition en phrases, mots, phonèmes) et des traitements automatiques de la parole,
- le besoin d'outils logiciels pour traiter automatiquement des bases de données et évaluer des systèmes de reconnaissance automatique de la parole,
- la conviction qu'afin d'obtenir des résultats satisfaisants en reconnaissance automatique de la parole il est nécessaire d'utiliser des connaissances variées qui coopèrent,
- la nécessité d'utiliser des environnements adaptés intégrant nos connaissances sur la parole et permettant d'en acquérir d'autres facilement, et
- la conviction qu'un environnement intégré est nécessaire pour capturer nos connaissances et faciliter leur utilisation.

StanisLAX n'est pas une collection de bibliothèques. C'est plutôt une station de travail experte destinée à faciliter le transfert de connaissances du chercheur sur la parole vers l'outil informatique utilisé.

1. Vue générale du système

Comme le montre la figure 2, StanisLAX se compose d'un ensemble de modules :

- le noyau de StanisLAX,
- un module scientifique,
- un module d'interface des données,
- un module de manipulation des données, et
- des boîtes à outils.

Le module scientifique contient la plupart des connaissances sur la parole. Ce module ainsi que le noyau du système, qui permet une abstraction des données et une représentation objet, constituent les modules critiques de StanisLAX. Le module interface des données réalise la liaison entre le monde réel et le laboratoire artificiel, alors que le module de manipulation des données s'occupe de l'interactivité et de la présentation des données. Enfin, les boîtes à outils sont des bibliothèques de programmes qui sont interfacées avec le noyau de StanisLAX afin de faciliter leur utilisation.

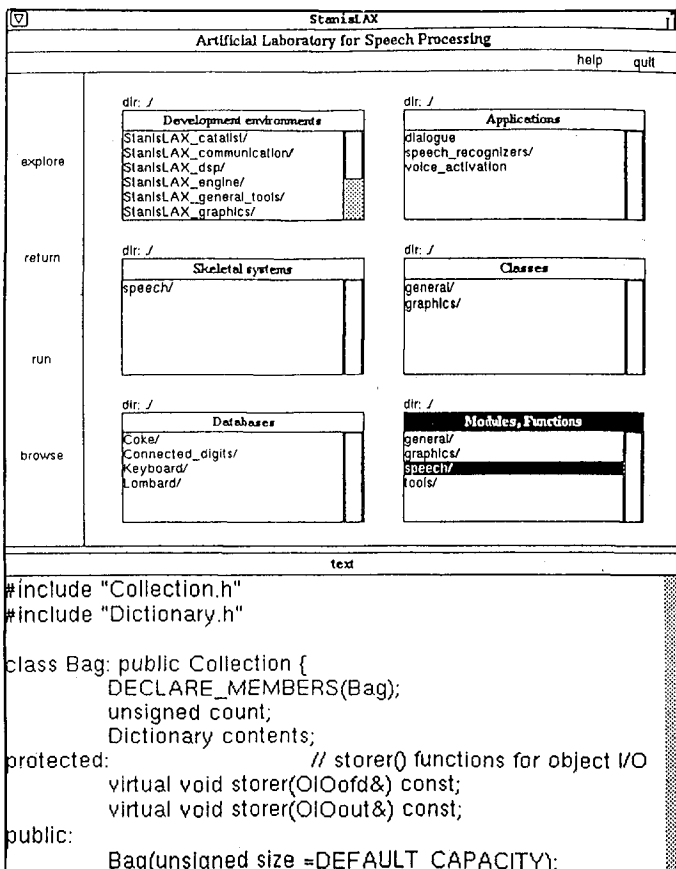


Fig. 1. — Exemple d'interface utilisateur.

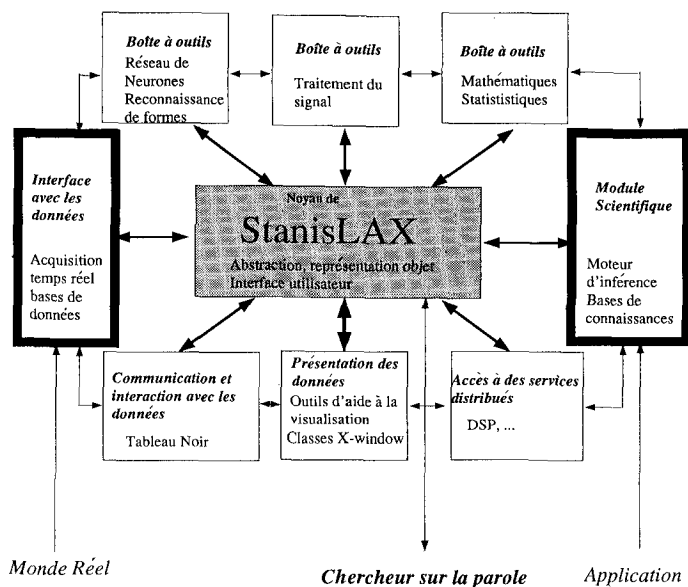


Fig. 2. — Vue générale de StanisLAX.

Deux types de représentation des données sont utilisés dans le système. Grâce à un ensemble de structures symboliques définies dans les différents modules, StanisLAX permet d'établir la connexion entre des fonctions bibliothèques et le noyau du système qui fournit un niveau d'abstraction plus important. L'utilisation de structures symboliques permet de bénéficier de ressources structurées.

Le deuxième type de données structurées est le concept de classe. Chaque structure de données est associée avec des méthodes permettant de manipuler les données. Ceci donne la possibilité au chercheur sur la parole de manipuler des niveaux d'abstraction plus importants que ceux fournis par le premier type de représentation.

2. Conclusions

Le but de StanisLAX est de permettre l'utilisation aisée de composants réutilisables et de fournir une organisation permettant de développer des algorithmes spécifiques au traitement de la parole. En utilisant cet environnement nous avons réalisé : 1) un système monolocuteur de commande vocale dont la particularité est de contrôler un sous-ensemble de l'interface à menus de StanisLAX et, 2) un système multilocuteur de dialogue permettant d'accéder à une base d'informations sur Santa Barbara. Le système à commande vocale et le système de dialogue que nous avons développés ont montré qu'un tel environnement permettait d'accélérer le prototypage de systèmes et le test d'applications. StanisLAX a permis d'implémenter ces systèmes rapidement (une semaine par application). De plus, comme les deux applications furent développées dans le cadre d'un système intégré, les extensions se sont avérées faciles. En particulier, StanisLAX nous a permis de prendre en compte des facteurs humains (utilisation d'un retour visuel pour indiquer à l'utilisateur avec quel succès une commande a été reconnue). Cette tâche aurait été plus difficile en dehors de l'environnement intégré.

Remerciements

L'auteur tient à remercier Qing Yu pour son aide dans la réalisation de l'interface utilisateur.

Manuscrit reçu le 28 mars 1991.

BIBLIOGRAPHIE

- [1] M. E. LACY, Artificial laboratories. *AI Magazine*, Summer-89 : 43-47, 1989.