

# BDVOX: BASE DE DONNEES POUR SYSTEMES DE RECONNAISSANCE DE LA PAROLE MULTILOCUTEUR

I. C SEARA; F. S. PACHECO; R. SEARA JR.; S. G. KAFKA; S. KLEIN; R. SEARA  
Université Fédérale de Santa Catarina – Florianópolis – Brésil  
izabels, fernando, ruijr, kafka, klein, seara@linse.ufsc.br

## RESUME

Le BDVOX est une base de données qui a été créé pour le développement d'un système de reconnaissance de la parole spontanée indépendant de locuteur et robuste aux variabilités acoustiques et linguistiques. C'est-à-dire le développement d'un système pour la compréhension par un ordinateur des mots énoncés par une voix humaine. Les performances d'un tel système ne dépendent pas seulement de la qualité des modèles acoustiques et linguistiques utilisés, mais aussi de la base de données conçue pour ce système.

Ainsi, pour aboutir à un système de reconnaissance de la parole qui soit indépendant du locuteur et de la tâche à remplir, il y a la nécessité d'une énorme quantité de voix, pour que le système puisse s'entraîner le mieux possible. D'autre part, la création de cette base de données doit être très bien conçue pour que tous les sons de la langue à reconnaître, les règles des constructions de syllabes, les tailles des mots, et d'autres contraintes linguistiques soient considérées dans cette base. La finalité de ces systèmes de reconnaissance de la parole est l'interaction homme/machine, par exemple, dans les services bancaires automatisés, téléphoniques, etc.

Le but principal de ce travail est que les systèmes puissent être utilisés par n'importe qui, c'est à dire qu'ils soient indépendant de locuteur (multilocuteur). Toutefois, pour arriver à un système de reconnaissance de la parole multilocuteur, il faut considérer les différents facteurs qui posent de complexité à la reconnaissance de la parole: d'abord la variabilité intra et interlocuteurs; après dans la parole continue, les mots dans les phrases sont fortement articulés, c'est-à-dire que la prononciation de chaque mot est affectée par le mot qui précède ainsi que par celui qui suit et, finalement, la variabilité concernat à la lecture ou spontanéité c'est-à-dire si le texte est lu ou parlé.

Ainsi nous présenterons ici la méthodologie utilisée pour la création des *corpora* pour l'enregistrement de voix qui constituerons la base de données des systèmes de reconnaissance de la parole (BDVOX), plus spécifiquement pour la langue portugaise parlée au Brésil. Ces *corpora* sont constitué de plusieurs *corpus* de la langue parlée contrôlée et spontanée, c'est-à-dire, une collection d'enregistrement de parole qui est accessible pour la lecture en ordinateur qui reçoit des notations et étiquetages suffisants pour permettre le re-usage de donnés par d'autres personnes et pour celles qui ont élaborées ces *corpora*.

Divers *corpora* ont été déjà créés avec des différents objectifs. Ils se distinguent par le nombre de locuteurs, par la forme et donnés enregistrés. Ce sont-ils:

1) EUROM\_1: parole contrôlée par la lecture - 60 locuteurs - salle spéciale. Cette base de données représente le travail conjoint de huit pays européen (Italie, Angleterre, Allemagne,

Hollande, Danemark, Suède, France, Norvège et postérieurement Grèce, Espagne et Portugal) – plusieurs langues [1];

2) BDFALA: parole contrôlée par la lecture - 10 locuteurs - salle à preuve de son – langue portugaise européenne [1];

3) BD-PUBLICO: parole contrôlée par la lecture - 120 locuteurs - salle à preuve de son. Base de données créée pour la langue portugaise européenne avec presque dix millions de mots [1];

4) SPEECHDAT - parole contrôlée par la lecture et spontanée - 5000 locuteurs - réseau téléphonique – plusieurs langues [2];

5) CORAL – dialogues spontanés – 32 locuteurs – salle à preuve de son [1];

6) SPEECON - collection de données de parole contrôlée par la lecture/spontanée - pour au moins 20 langues et 600 locuteurs par langue (adultes et enfants) [3].

Le BDVOX, développé par l'Université Fédérale de Santa Catarina, au Brésil, a été élaboré avec la parole contrôlée par la lecture et aussi spontanée, enregistrées par le réseau téléphonique. Cette base est composée par 1100 locuteurs brésiliens (portugais du Brésil), ce qui fait 180 heures de voix enregistrées. Dans ce moment, nous sommes en train de augmenter cette base, en incorporant plus 500 locuteurs.