
ALGORITHMES ÉVOLUTIONNAIRES POUR L'ÉTUDE DE LA ROBUSTESSE DES SYSTÈMES DE RECONNAISSANCE AUTOMATIQUE DE LA PAROLE

Anne Spalanzani

CLIPS-IMAG, Université Joseph Fourier
Bat. C, BP 53, F-38041 Grenoble cedex 9, France
Tel : (+33) 4 76 63 56 82
Mel : Anne.Spalanzani@imag.fr

Les systèmes de reconnaissance automatique de la parole (SRAPs) sont aujourd'hui bien connus dans le monde de l'informatique et suscitent l'intérêt d'un public de plus en plus large. De ce fait, ces systèmes doivent être opérationnels en situation réelle, dans un contexte de fortes variations de conditions acoustiques. L'environnement, l'utilisateur, les microphones utilisés, etc. sont autant de sources de variabilité qui peuvent faire chuter les performances de ces systèmes. Afin de garder des taux de reconnaissance acceptables, il est alors nécessaire d'imposer des contraintes sur leurs conditions d'utilisation. Le choix du microphone, le débit de parole ou la taille du vocabulaire font partie des paramètres à gérer correctement pour une bonne robustesse des SRAPs. Ainsi, un SRAP est robuste s'il est capable de garder un bon taux de reconnaissance même si la qualité du signal est dégradée, ou si les caractéristiques acoustiques du signal sont différentes entre la phase d'apprentissage du système et celle d'utilisation.

Cette notion de robustesse est l'un des problèmes majeurs qui limitent l'utilisabilité de ces systèmes. Comment rendre les systèmes existants adaptables aux conditions acoustiques et aux caractéristiques changeantes d'élocution en situation réelle, c'est-à-dire hors laboratoire ?

D'une manière générale, la stratégie prédominante adoptée par les chercheurs pour résoudre ces problèmes consiste à enregistrer des corpus d'apprentissage dans des conditions qui soient les plus proches de celles dans lesquelles les différents systèmes sont utilisés. Cela nécessite de définir des normes et des critères de comparaison des systèmes de reconnaissance, l'objectif étant d'estimer et de garantir leur bon fonctionnement lors de leur utilisation. Malheureusement les sources de variabilité qui peuvent affecter les performances d'un système de reconnaissance sont nombreuses et non prévisibles (changement de locaux, de bruit ambiant, etc.), et ces systèmes, utilisés alors comme des "boîtes noires" statiques, s'adaptent mal aux variations acoustiques.

Toutefois, des travaux mettant l'accent sur des méthodologies nouvelles, visant une meilleure utilisation des données existantes, reçoivent une plus grande attention de la communauté scientifique. L'adaptation aux conditions acoustiques a ainsi fait l'objet de nombreux travaux qui ont débouché sur des techniques de filtrage et d'adaptation au bruit ambiant qui améliorent sensiblement les performances des systèmes de reconnaissance. Quant aux problèmes d'adaptation aux locuteurs, ils sont de plus en plus considérés comme cruciaux pour les nouvelles technologies vocales. Une multitude d'algorithmes ont ainsi été élaborés afin d'améliorer la robustesse.

Dans le cadre de cette thèse, nous nous sommes intéressée à un type d'algorithme, inspiré de l'évolution naturelle formulée par Darwin, qui propose de faire évoluer des populations d'individus dans un environnement en favorisant la survie et la reproduction des individus les mieux adaptés à cet environnement. On connaît l'efficacité de ce type d'approche pour l'optimisation de fonctions complexes, pour la recherche de solutions à des problèmes ayant un nombre de variables très grand. Nous cherchons à savoir si ce type d'algorithmes appelés algorithmes évolutionnaires peut constituer une nouvelle approche pour l'adaptation des SRAPs aux changements d'environnement.

Afin d'étudier la possibilité d'incorporer des algorithmes évolutionnaires dans le domaine de la robustesse des SRAPs, nous proposons, dans le cadre de cette thèse, de nous limiter au niveau acoustique du traitement de la parole et plus particulièrement au niveau de l'adaptation acoustique des SRAPs. Souhaitant aborder le problème de la robustesse d'un point de vue global, nous ne faisons aucune supposition sur le type de bruit traité (par exemple téléphone, voiture, etc.) et les méthodes proposées ne dépendent d'aucun domaine spécifique.

Dans ce contexte, notre travail s'est situé à deux niveaux. Dans un premier temps, nous avons cherché à adapter le système de reconnaissance lui-même aux différents environnements. Nous avons étudié les capacités du système à s'adapter aux changements de conditions acoustiques, à l'aide d'une approche locale (par rétro-propagation du gradient) et d'une approche globale (par algorithmes évolutionnaires), en vue de trouver un SRAP optimal.

Dans un second temps, nous nous sommes placée dans le cadre du traitement des données en entrée du système. Partant d'une base de projection issue d'une analyse en composantes principales, nous avons cherché à trouver une base de projection adaptée à chaque environnement rencontré et permettant de retrouver les conditions acoustiques connues du système. L'étude que nous avons effectuée s'est alors portée sur l'optimisation de cet espace de représentation par algorithmes évolutionnaires. Nous avons étudié l'influence de ces algorithmes sur l'évolution de l'espace de représentation.

Nous avons mis en place une plate-forme de simulation permettant de faire évoluer des populations de systèmes de reconnaissance. Les résultats obtenus montrent qu'en moyenne l'hybridation des algorithmes évolutionnaires et des techniques de reconnaissance classiques améliore sensiblement, et de manière stable, les performances du système de reconnaissance, et ceci dans les deux types d'hybridation que nous avons testés.