

La Vie Artificielle d'une société de robots parlants : émergence et changements du code phonétique.¹

Hervé Glotin¹, Rafael Laboissière²

^{1&2} Institut de la Communication Parlée

INPG, 46 avenue Félix Viallet 38031 Grenoble Cedex, France.

¹E-mail : glotin@icp.grenet.fr ²E-mail : rafael@icp.grenet.fr

Résumé

La parole est un parcours optimisé d'une séquence d'objectifs perceptifs négociés entre un locuteur et un auditeur. Comment sont négociés les objectifs et optimisés les parcours ? Ces questions sont à la base de l'élaboration de futurs systèmes de reconnaissance automatique et de synthèse adaptative de la parole. A partir d'un modèle articulatoire nous avons simulé l'émergence d'un code phonétique dans le cadre d'interactions acoustiques entre robots parlants nommés carls. La seule contrainte de convergence des lexiques des carls porte sur l'évaluation de la coopération entre les robots et sur leurs contraintes articulatoires. Les systèmes qui émergent s'approchent des systèmes les plus fréquents dans les langues du monde. Les algorithmes de type « vie artificielle » ont été mis en oeuvre pour simuler l'évolution d'une telle société d'agents sensori-moteurs.

mots clés : phonétique, autoorganisation, vie artificielle, émergence, communication, coopération, production de la parole.

1. Introduction

Supposez qu'un robot parlant soit capable d'apprendre à produire des gestes articulatoires à partir de sons entendus. Imaginez qu'une assemblée de ces robots essaie d'établir un code de communication orale. Nous serions alors en mesure de trouver des réponses à des questions qui sont au coeur de la recherche en communication parlée. Quelles unités phonétiques choisiront les robots pour communiquer ? En supposant que les simulations soient assez sophistiquées, comment le code phonétique changera au cours du temps et comment apparaîtront les dialectes ? Plusieurs études ont été consacrées à la compréhension de l'émergence du code linguistique à partir des connaissances articulatoires et acoustiques de la production de la parole (Boë et al., 1994). Elles partent toutes de la même question: comment via des interactions sensori-motrices, émergent les systèmes linguistiques? Ces études sont fondées sur des principes d'optimisation de critères liés à la fois à l'efficacité perceptive et à l'économie motrice. Lindblom (1995) conclue même que : « La parole est au fond un phénomène biologique, et, par conséquent, on ne peut pas regarder la variabilité de la parole comme un accident. Notre examen des données expérimentales nous mène [...] à proposer qu'elle est une conséquence inévitable de son organisation adaptative. »

Notre étude porte sur les changements phonétiques résultants d'échanges entre instances de robots parlants (des modèles articulatoires) au sein d'interactions en production et en perception. Nous avons conçu un cadre de modélisation suffisamment puissant pour que les phénomènes de changements phonétiques soient mis en évidence, ainsi que les mécanismes d'émergences des systèmes phonologiques. Dans cette modélisation, les ensembles de recherche sont purement continus et les interactions sont locales². Ces deux aspects sont novateurs dans ce domaine de recherche en phonétique.

¹ Publié dans les Actes de la conférence modélisation des systèmes naturels complexes : *du collectif au social ROCHEBRUNE 1996*. ECAL&Ecole Nationale Supérieure des Télécommunications-Paris.

² Contrairement aux travaux de Boë et al., 1994 (voir fin de la section 2).

Le robot parlant et sa fonction d'inversion permettant d'inférer les séquences articulatoires en fonction des buts acoustiques donnés, étaient déjà élaborés à l'Institut de la Communication Parlée (ICP) (Laboissière, 1992). Nous nous sommes donnés des critères et des mécanismes d'une bonne communication (meilleure coopération entre agents), en modélisant l'interaction entre production et perception dans le contexte d'une robotique anthropomorphique de la communication parlée. Nous avons construit un cadre d'interaction entre deux robots qui se « parlent » et déterminé comment les robots adaptent en conséquence leurs codes internes avec un nombre fixe d'items dans leur lexique. Nous avons structuré une société de robots, agents sensori-moteurs, en forçant les échanges entre tous, et en forçant l'utilisation de tous leurs items. Nos algorithmes sont du registre de la «vie artificielle». Nous présenterons tout d'abord les recherches en phonétique afin de montrer comment nous pouvons les transposer dans notre approche. Puis nous aborderons notre monde AGORA et ses algorithmes de vie artificielle ainsi que les résultats des simulations. Ensuite nous concevrons un système d'émergence de dialectes basé sur la théorie des niches écologiques en vie artificielle. Enfin nous ouvrirons une perspective sur l'étude des changements phonétiques à travers la théorie de la variabilité adaptative (Lindblom, 1987).

2. Les systèmes phonétiques

Lors de la production des sons de la parole, le conduit vocal peut être excité selon un mode d'excitation quasi-périodique. L'appareil articulatoire renferme des cavités résonnantes et à ce titre vient renforcer certaines fréquences, appelées formants. On caractérise les voyelles par leurs formants classés par ordre croissant de fréquences (F1, F2, F3 etc ...). Les contraintes articulatoires, dépendantes de la géométrie du conduit vocal humain, restreignent les formants dans un certain domaine clos. Dans le plan F1 et F2 ce domaine s'approche d'un triangle, les voyelles [a i u] occupant ses sommets. Chaque langue, possède un certain ensemble de voyelles, dites canoniques qui constituent un système vocalique. Le critère de qualité des systèmes vocaliques repose sur la maximisation des distances perceptives inter-voyelles (Liljencrants & Lindblom, 1972). Mais ce contraste est limité par une exigence d'économie articulatoire qui est antagoniste à la demande en contraste perceptif. Les données expérimentales en production montrent l'existence de deux « forces » antagonistes spécifiant le contrôle du système moteur de la parole : l'exigence de distinction perceptive (du point de vue de l'auditeur) et de la demande d'économie articulatoire (du point de vue du locuteur) (Schwartz, 1995). Dans le modèle de prédiction de Bosch (1986), *les voyelles préférentielles dans les systèmes sont celles qui au niveau articulatoire sont produites avec une déviation minimale par rapport à une position neutre, tout en conservant une distinction perceptive maximale.*

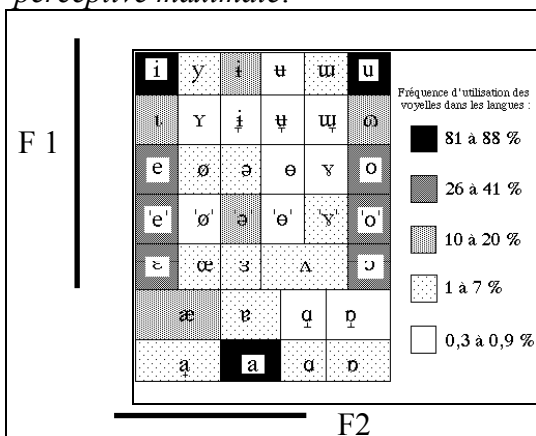


Figure 1 : occurrences des voyelles dans les langues (Vallée, 1994), placées suivant leur deux premiers formants.

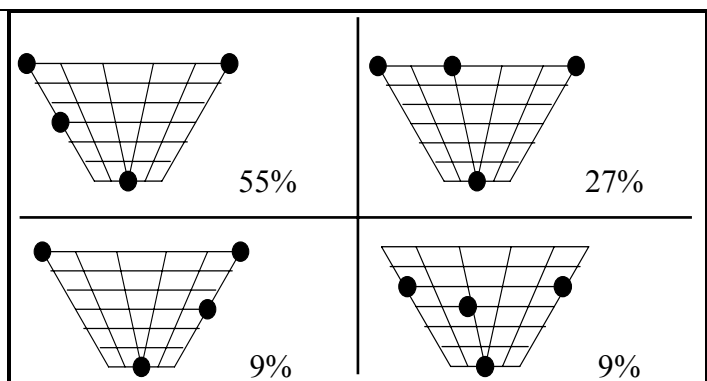


Figure 2 : les systèmes vocaliques à 4 voyelles relevés dans la base de 317 langues UPSID avec leur occurrence (Boë et al., 1994)

Par exemple la plupart des langues du monde qui sont composées de trois voyelles, ont leurs voyelles qui sont placées aux extrêmes du triangle vocalique. Mais les langues aborigènes à trois voyelles ne suivent pas cette règle : leurs voyelles sont plus centrées dans le triangle vocalique (Butcher, 1994). Néanmoins les grandes tendances des systèmes vocaliques (cf. Fig. 1 & 2), sont expliquées par des critères de perception et de production.

Liljencrants et Lindblom (1972) proposent un modèle de prédiction des systèmes vocaliques suivant le critère global de maximisation de la somme de toutes les distances inter-voyelles dans l'espace perceptif, ce qui assure une efficacité perceptive maximum. Cette théorie de dispersion maximale, améliorée par un coût intra-vocalique (Boë et al., 1994) donne des prédictions adéquates.

3. Un modèle d'évolution de coopération

MacLennan et Burghardt (1994) présentent un modèle d'évolution de coopération (Axelrod, 1984). Ils étudient le comportement de créatures artificielles appelées *simorgs*. Ils montrent comment à partir d'interactions locales entre les *simorgs* dans un environnement simple, un langage propre à la communication peut émerger, sur le seul critère de coopération.

Dans ces simulations, la coopération prend la forme d'une action pertinente sur un environnement régi par un automate interne. Dès lors, il est possible de comptabiliser les bénéfices d'une bonne coopération entre deux créatures. Ce qui donne lieu au calcul d'une fitness qui est utilisée pour choisir les créatures qui seront élevées et celles qui seront tuées. Une simulation consiste en des «cycles d'élevage», qui sont constitués de «cycle d'environnement», eux mêmes composés de «cycles d'action». Après un certain nombre de cycles «d'environnement et d'action», selon la valeur de leur fitness, les créatures sont tuées, ou choisies comme parents pour créer une nouvelle population en croisant (combinaison des chaînes de symboles), les tables de transitions de leur automate par un algorithme génétique (Goldberg, 1994). Donc plus la table de transition d'un *simorg* est adaptée, plus celui-ci se reproduit par le biais de l'algorithme génétique que nous décrivons dans le paragraphe suivant. Les enfants possèdent des propriétés issues de la combinaison de celles des parents, et sont plus ou moins bien adaptés. Finalement toute la société atteint un fort niveau de coopération grâce à un langage émergent.

Ce travail nous a inspiré pour nos recherches, même si la notion de communication par langage n'est pas explicitement introduite dans nos simulations. Les organismes coopérants de notre société, à l'instar des *simorgs*, n'attachent pas de sens aux symboles produits. Nos recherches ne portent pas sur la genèse du langage, mais seulement sur le choix des symboles phonétiques, en supposant le langage pré-existant.

4. AGORA et les carls

AGORA est un ensemble de robots parlants, chacun nommé *carl*³, qui interagissent les uns avec les autres. Le but est de les faire converger vers des symboles phonétiques communs propres à la communication en simulant les mécanismes de base de perception et de production de la parole.

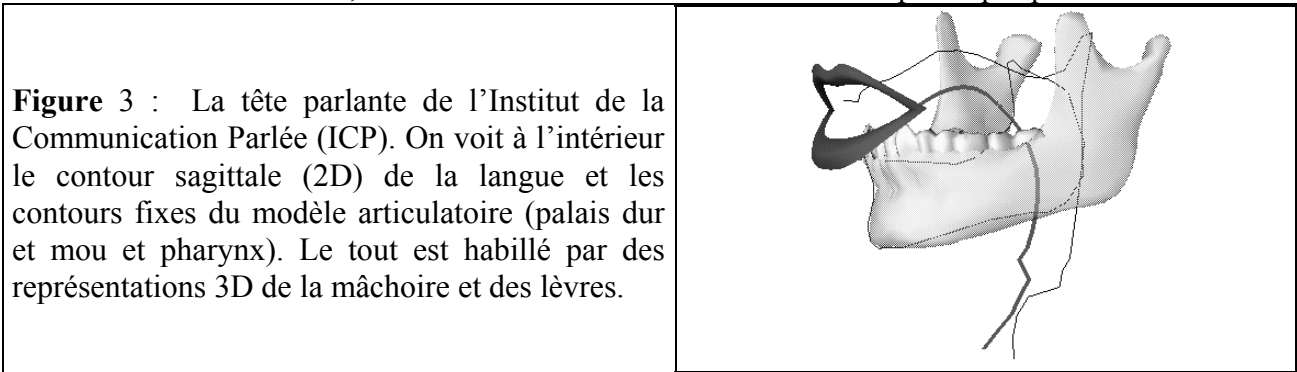
4.1. Le modèle articulatoire

Un *carl* est une instance d'un modèle articulatoire issu d'une analyse statistique de la forme du conduit vocal (Maeda, 1988). Avec huit paramètres de commande (écartement de la mâchoire, le dos et le corps de la langue, le déplacement horizontal et vertical du bout de la langue et celui des lèvres et la hauteur du larynx), le modèle articulatoire génère la coupe sagittale du conduit vocal (cf. Fig. 3). A partir de celle-ci, on calcule la fonction d'aire et la fonction de transfert du conduit vocal qui fournit les formants correspondants. Un *carl* est donc défini par la matrice des paramètres

³ Cerveau Analytique de Recherche en Linguistique / Cooperative Agent for Research on Linguistic

articulatoires (état interne), et des formants associés à chaque item vocalique (état externe accessible à chaque *carl*) donnés par la fonction directe du modèle.

Certes, le travail décrit ici pourrait être étendu à de vrais robots utilisant des médias de communication arbitraires, mais nos *carls* sont manifestement anthropomorphiques.



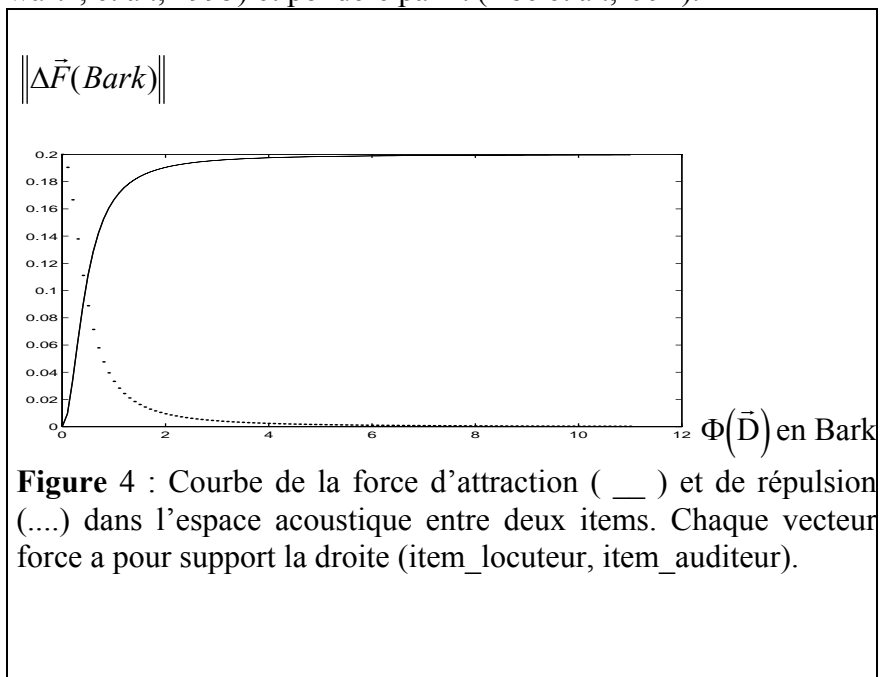
4.2. Les interactions phonétiques des carls

Les *carls* établissent des communications orales deux par deux en utilisant leurs symboles internes, chaque *carl* ayant son lexique d'items vocaliques, initialement tirés au hasard autour de la voyelle neutre (obtenue par le relâchement du conduit vocal, paramètres articulatoires nuls). Lors d'une interaction (cf. Fig. 5), le *carl* émetteur produit un symbole interne choisi au hasard qui est perçu par le *carl* récepteur. Le *carl* récepteur compare au niveau acoustique ce symbole perçu avec ceux de son lexique et il les adapte de la façon suivante: l'item le plus proche est approché, les autres sont éloignés, suivant des forces d'attraction et de répulsion (cf. Fig. 4). Cette règle simple d'interaction locale peut faire émerger des codes phonétiques comme nous le montrerons plus tard. Notre espace perceptif a pour unité le bark qui est l'échelle physiologique des fréquences. Dans cet espace la distance perceptive Φ , entre deux items vocaliques d'un *carl* A et un *carl* B est (Boë et al.,1994):

$$\Phi = \sqrt{\left((F1_a - F1_b)^2 + \lambda^2 (F2'_a - F2'_b)^2 \right)}$$

$F1_x$ et $F2'_x$ sont les formants en bark (Schroeder et al., 1979). $F2'$ le deuxième formant effectif est calculé d'après $F2$ $F3$ et $F4$ (Schwartz, et al., 1995) et pondéré par λ (Boë et al.,1994).

Chaque item du lexique du *carl* récepteur subit donc un changement acoustique. Puisque les symboles internes sont représentés en terme de position articulatoire il faut que le *carl* soit capable d'inverser une variation formantique en une variation articulatoire. Or la fonction inverse formants vers articulatoires n'est pas unique, car il s'agit d'un problème à excès de liberté (4 entrées pour 8 sorties). Nous implémentons une fonction d'inversion qui assure un minimum de changement articulatoire.



Dans le cadre de notre travail, étant donné que tous les *carls* sont identiques au niveau des capacités de production, nous ne traitons pas le problème de normalisation (Payan, 1992), à savoir comment l'auditeur ramène les symboles produits par le locuteur à son propre système de production.

4.3. L' algorithme d'apprentissage

Nous définissons une ronde d'interactions comme une suite aléatoire de couples de tous les *carls* qui vont communiquer successivement. On tire aléatoirement autant de rondes que de nombre de cycles de la phase d'apprentissage. Le cumul des normes des vecteurs de variation articulatoire qu'un *carl* effectue à chaque interaction nous donne son coût d'adaptation. Moins un *carl* change son lexique au cours des interactions, plus son langage est adapté pour communiquer avec la société. Ceci est le coeur de notre approche. Nous pouvons dès lors calculer une fonction rendant compte du degré de coopération d'un *carl* dans la société : sa *fitness*.

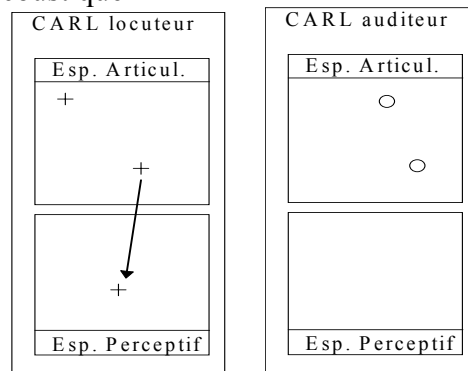
$$\text{fitness}(\text{carl}) = \frac{1}{1 + \text{cout}(\text{carl})} \quad \text{avec}$$

$$\text{cout} = \sum_{\text{interactions}} \sum_{\text{items}} \|\Delta \vec{\text{Articu}}\|$$

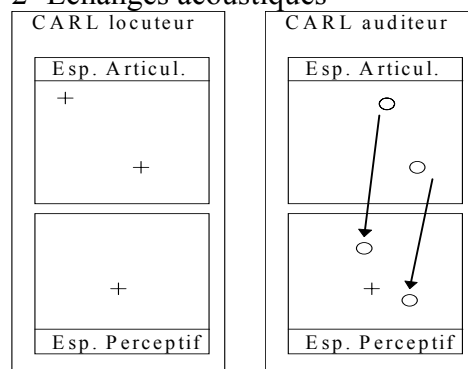
Nous infligeons à notre société une pression de sélection, pour une convergence vers une coopération croissante entre tous les *carls*. On favorise la perpétuation des *carls* les plus adaptés dans une nouvelle population de *carls* pour un nouvel apprentissage (cf. Fig. 6). L' outil d'exploration pour notre modèle, sera un algorithme génétique ou pseudo-génétique, en ce sens que c'est un algorithme qui ne contient pas de notion de mutation ni d'autre type d'opérateurs aléatoires et que la taille des populations reste faible. Cet outil ne reflète aucun aspect anthropologique dans nos simulations, et notre technique génétique qui est du type Lamarckien n'a pas pour vocation de refléter un aspect évolutif, mais seulement prospectif.

Figure 5 : les interactions entre carls

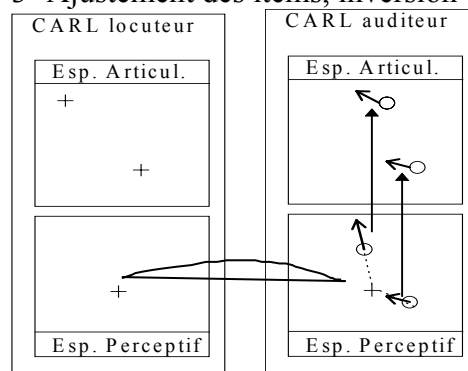
1-Choix d'un item et sa production acoustique



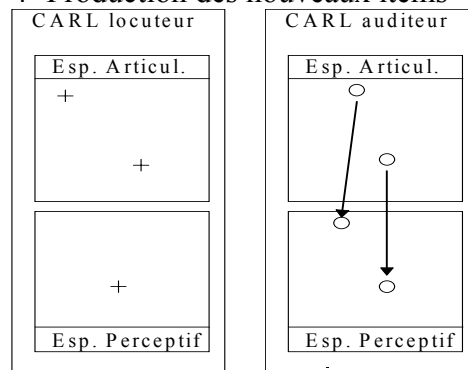
2- Echanges acoustiques



3- Ajustement des items, inversion



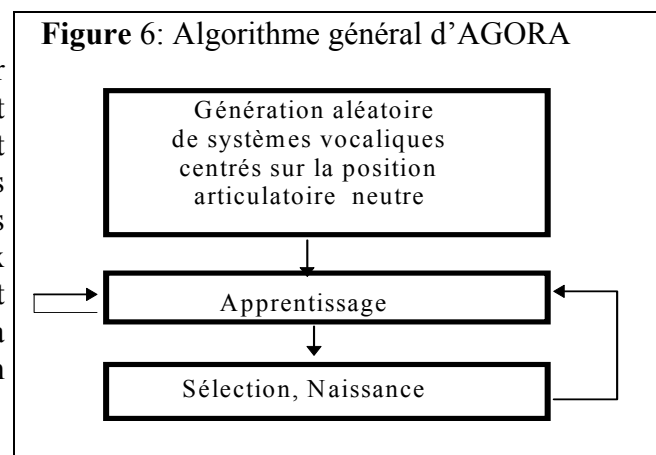
4- Production des nouveaux items



Une fois l'apprentissage terminé, les *carls* sont classés par ordre croissant de *fitness* et sont dupliqués dans la nouvelle génération. Notre fonction de duplication des meilleurs *carls* est appliquée en traitant les cas par ordre décroissant de *fitness*, et tant que le nombre d'individus est insuffisant pour obtenir une population de taille identique à la précédente. La règle est qu'un individu est dupliqué en un nombre de descendants proportionnel à sa *fitness* (Baker 1985).

Puis une moitié de cette sous population de meilleurs *carls*, est croisée entre elle, l'autre moitié reste intacte, suivant la technique de l'élitisme qui accélère la convergence (Goldberg, 1994). Le croisement consiste en un échange de la moitié des items des lexiques de deux *carls*. Par exemple, pour quatre items on fixe le point de croisement entre le second et le troisième item. Une fois née, la nouvelle génération reprend le même cycle d'apprentissage que celui de ses parents. On arrête la simulation lorsque l'on juge la convergence suffisante.(cf. Fig. 6).

De façon à assurer que l'apprentissage acquis par les parents sera perpétué chez les enfants, il faut que les régions du système vocalique qui sont occupées par les parents apparaissent aussi dans le lexique des enfants. Pour cela nous avons conçu un algorithme d'appariement de deux lexiques avant leur croisement. L'appariement produit les couples d'items inter-lexiques, la distance perceptive entre chaque item d'un même couple étant minimale.



4.4. Validation de l'algorithme d'apprentissage

Il nous est apparu nécessaire de démontrer la cohérence de notre algorithme en supprimant les fonctions articulatoires pour tester le sous-problème « acoustique seule ». Nous avons donc réalisé une version d'AGORA où les *carls* n'évoluent que dans le seul espace des formants, leur coût d'adaptation étant dans ce cas particulier la somme des variations acoustiques pour chaque item à chaque échange, ce qui garantit la coopération (convergence des lexiques) des individus dans l'espace acoustique. En effet, on minimise ainsi le déplacement des voyelles à chaque génération car on sélectionne les *carls* dont le lexique a peu bougé, donc ceux dont le niveau de coopération est élevé car leur demande d'adaptation a été faible.

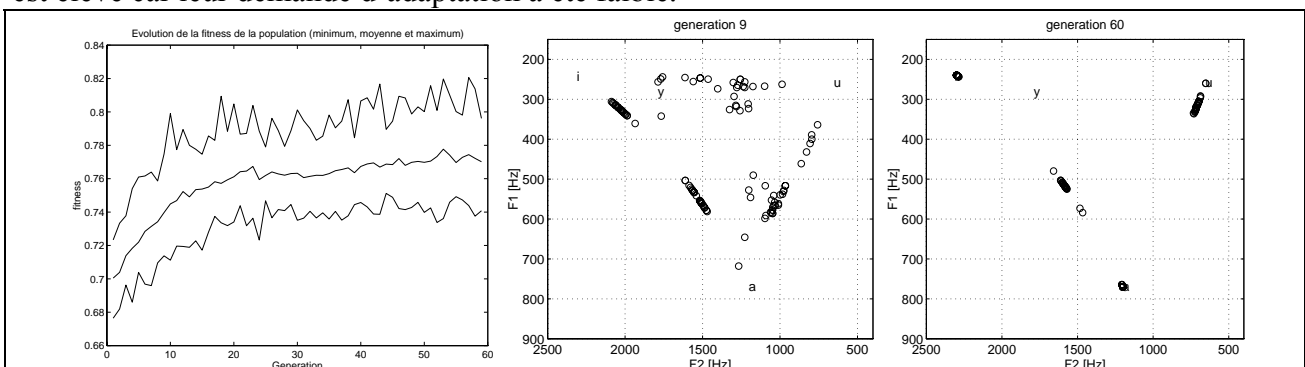


Figure 7: Simulation « acoustique seule », avec 30 *carls* et un cycle de quatre apprentissages. *A gauche:* les fitnesses maximum, moyenne et minimum de la population au cours des générations. *Au centre et à droite:* évolution des lexiques confondus de la population des 30 *carls* aux générations 9 et 60. **(nb: les items sont représentés par un rond, les prototypes [i y u a] servent de repère)**

Dans cette version d'AGORA l'espace acoustique a deux dimensions, les réalisations acoustiques sont bornées a priori par un triangle dont les sommets sont les formants F1 et F2 pour [i a u]. Durant l'apprentissage tout point qui sort du triangle est projeté orthogonalement sur le bord, ce qui simule les contraintes de production. La fitness (cf. Fig. 7) suit l'évolution prédite dans la section 4.3 : la coopération entre les agents *carls* croît. La convergence des items est observée pour *chaque* lexique des individus vers les voyelles [i a u] et ['e'] dès la soixantième génération (cf. Fig. 7), ce qui correspond au système vocalique à quatre voyelles le plus fréquent dans les langues du monde (cf. Fig. 2). Nous pouvons maintenant explorer l'espace articulatoire.

5. Résultats

Nous passons maintenant à des simulations perceptuo_motrices, c'est à dire que nous nous plaçons dans les espaces acoustique et articulatoire comme le décrit la section 4.2. La simulation réalisée comporte 4 *carls* possédant chacun 4 items, pour 6 cycles (6 rondes de dialogues durant l'apprentissage). Nous avons filtré la moyenne de la fitness au cours des générations (cf. Fig. 8) pour ne rendre compte que des variations significatives des populations. Aux générations correspondantes à un pic de la fitness moyenne filtrée, les *carls* convergent vers un lexique commun.

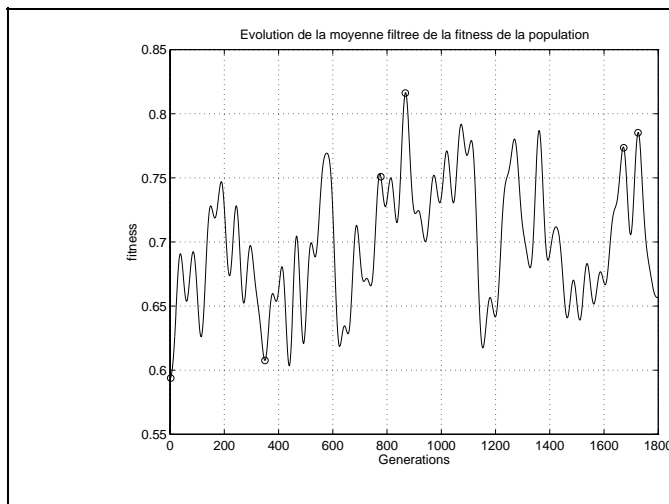


Figure 8: Evolution du degré de coopération des *carls*: la fitness moyenne filtrée de la population sur les 1800 premières générations. Les ronds indiquent les fitness moyennes des générations détaillées par la suite. Le filtre utilisé est du type Butterworth d'ordre 10 avec une fonction de coupure de $0.05 \text{ génération}^{-1}$.

Les chutes de la fitness peuvent être attribuées à une « catastrophe génétique » (MacLennan et al. 1994), ou à une instabilité des fonctions articulatoires. La dispersion acoustique des items (cf. Fig. 9) est observée à la génération 351. La fitness à cette génération est faible (cf. Fig. 8), les *carls* coopèrent pas. Le pic de fitness moyenne pour la 777 ième génération correspond déjà à l'agrégation des lexiques qui tendent vers [i] [a] [u] [y] (cf. Fig. 9). Finalement à la génération 867, qui a une forte fitness moyenne, chaque *carl* a un lexique qui a convergé vers des voyelles adéquates : le [a] et le [u] y sont approchés et les deux autres voyelles s'approchent du [i] et d'une voyelle centrale assez avancée. Ceci est proche du système [i, a, u, ε] qui est un système vocalique existant dans les langues du monde (cf. Fig. 2).

La chute de la fonction fitness vers la 1150 ième génération correspond à une réorganisation des systèmes (cf. Fig. 9). Nous observons ce phénomène sur l'ensemble de la simulation. Le système de la génération 1670 (cf. Fig. 10), est sensiblement identique à celui de la 867 ième, mais le [ε] a basculé vers le [o]. En mettant en évidence les coupes sagittales des *carls* (cf. Fig. 3) à la génération 867 (cf. Fig. 11), nous montrons que chaque *carl* a développé sa propre stratégie articulatoire: par exemple une même voyelle peut être plus ou moins labialisée. A la figure 12, nous

montrons que les *carls* 1 et 3 de la génération 1725 coopèrent sur le système [i a u 'o'] (d'après Fig. 1).

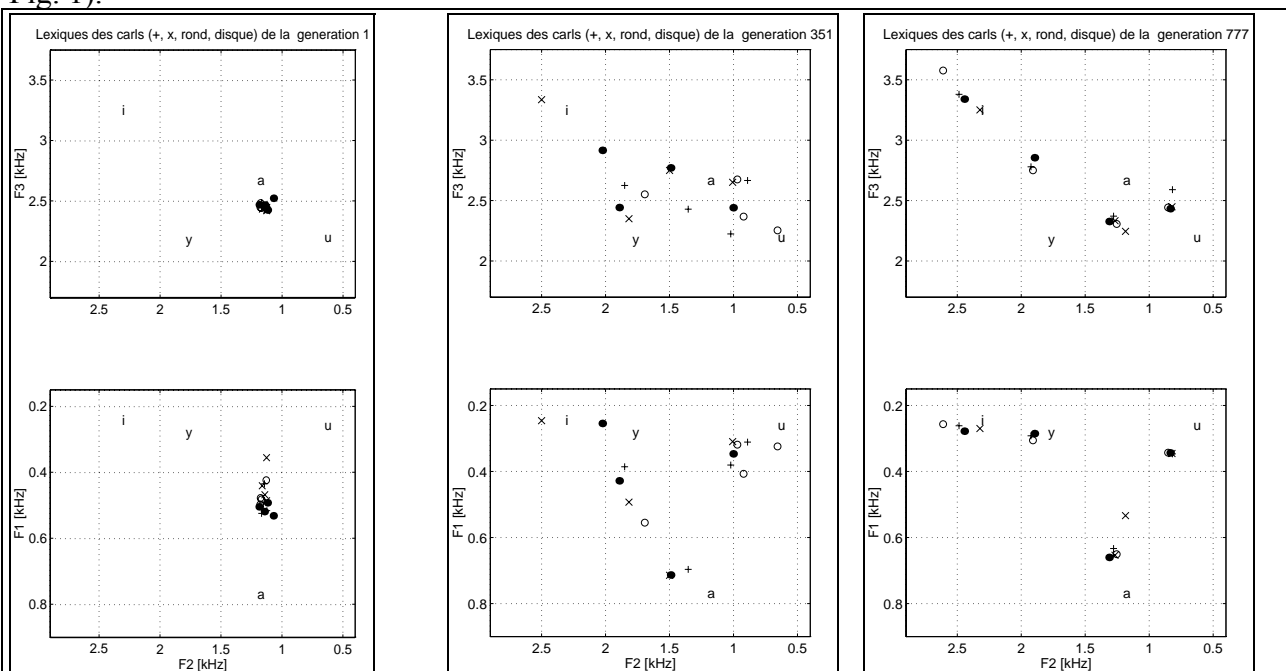


Figure 9: évolution des lexiques confondus de la population aux générations 1, 350 et 777. Les différents lexiques sont représentés par +, x, o, et un disque. Les prototypes [i y u a] servent de repère.

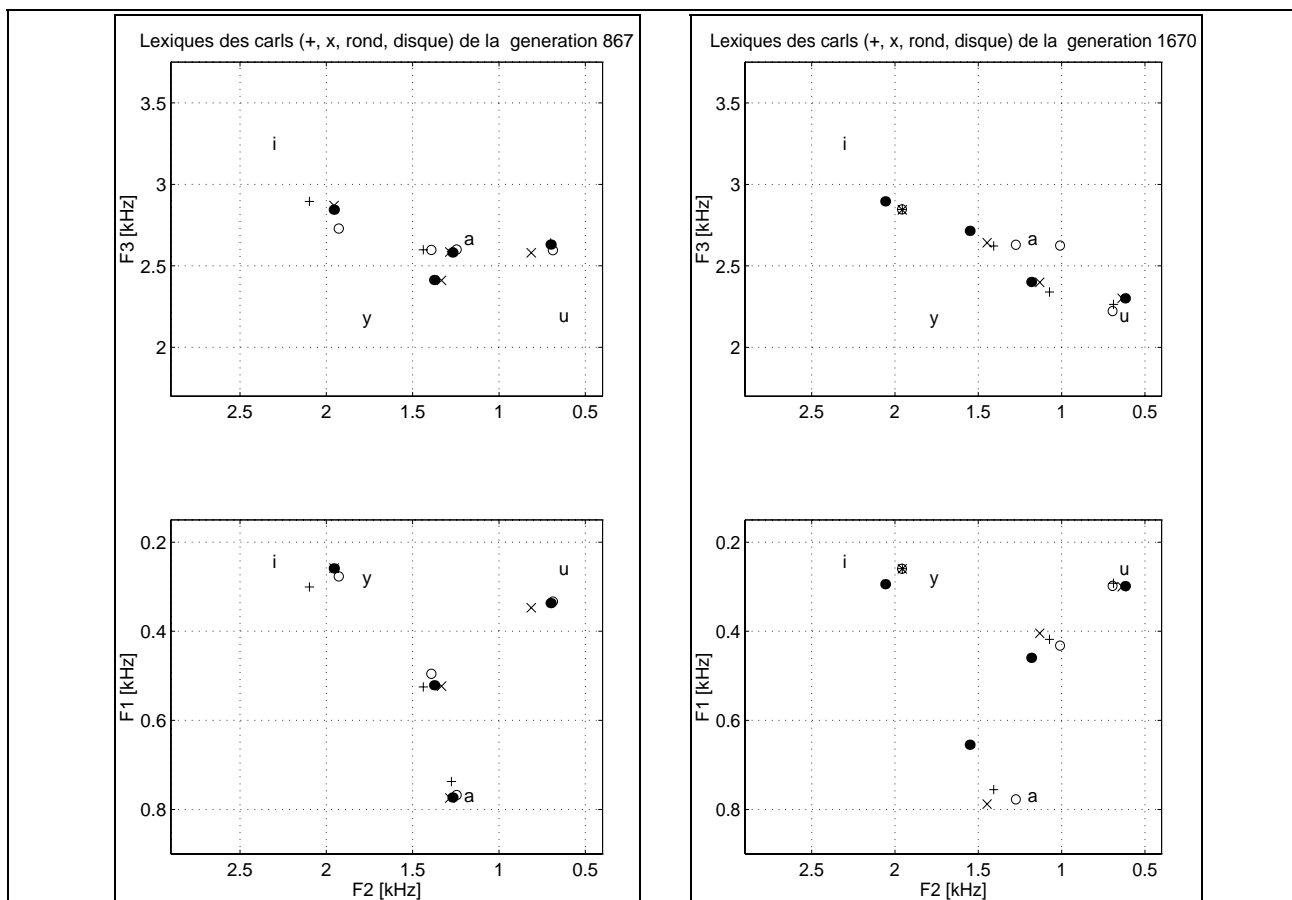
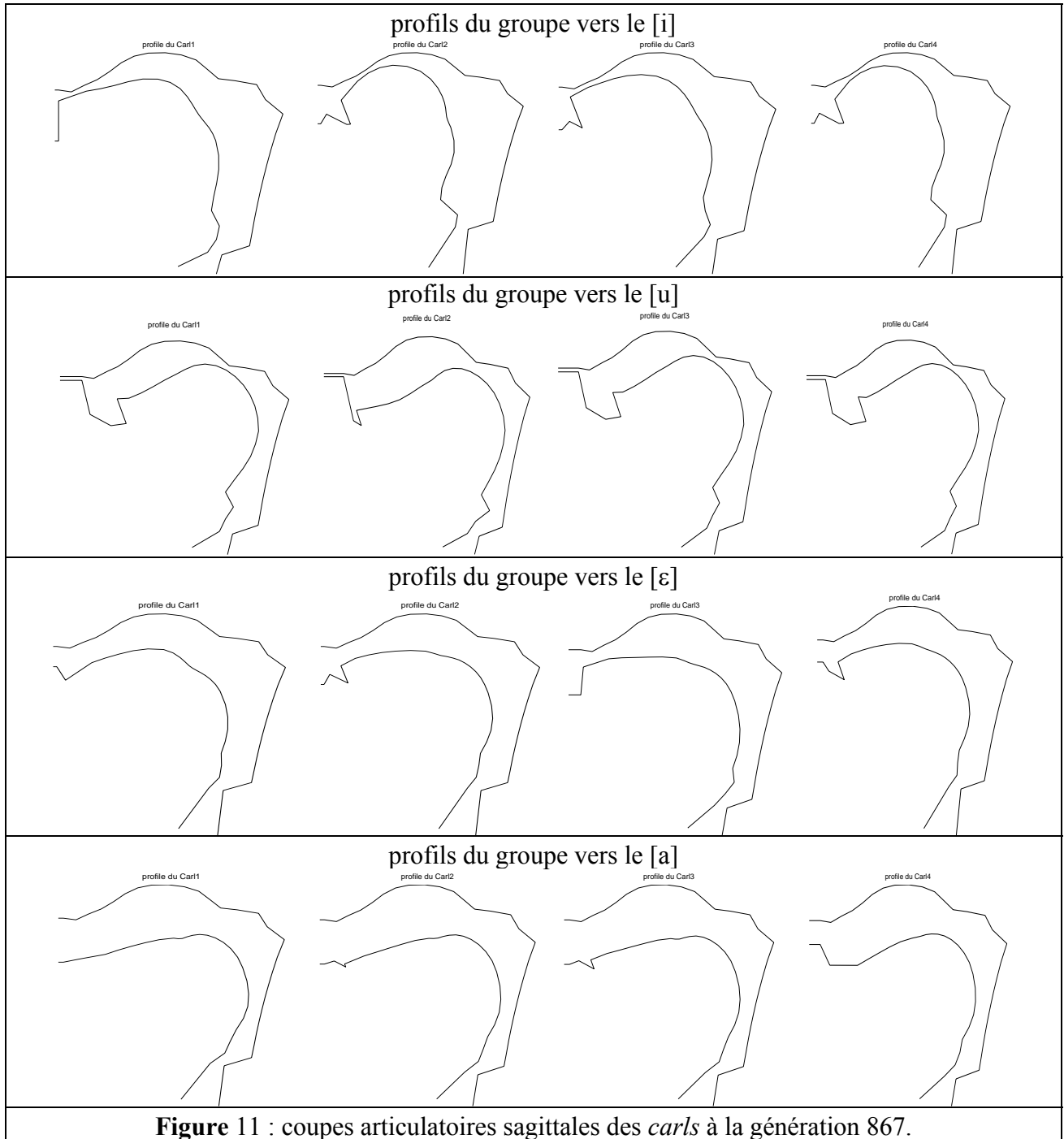


Figure 10: lexiques confondus de la population à la génération 867 à gauche. A droite ceux de la génération 1670. (Dans F1 F2 il y a superposition de quelques points).

Ce dernier est proche d'un des systèmes des langues à quatre voyelles. Les coupes sagittales articulaires de ces deux *carls* (cf. Fig. 13) montrent qu'ils ont choisi une stratégie articuloire différente. Les lexiques des deux autres *carls* sont inadéquats sur deux voyelles, elles n'ont pas été émises par des locuteurs lors de la phase d'apprentissage qui est courte.



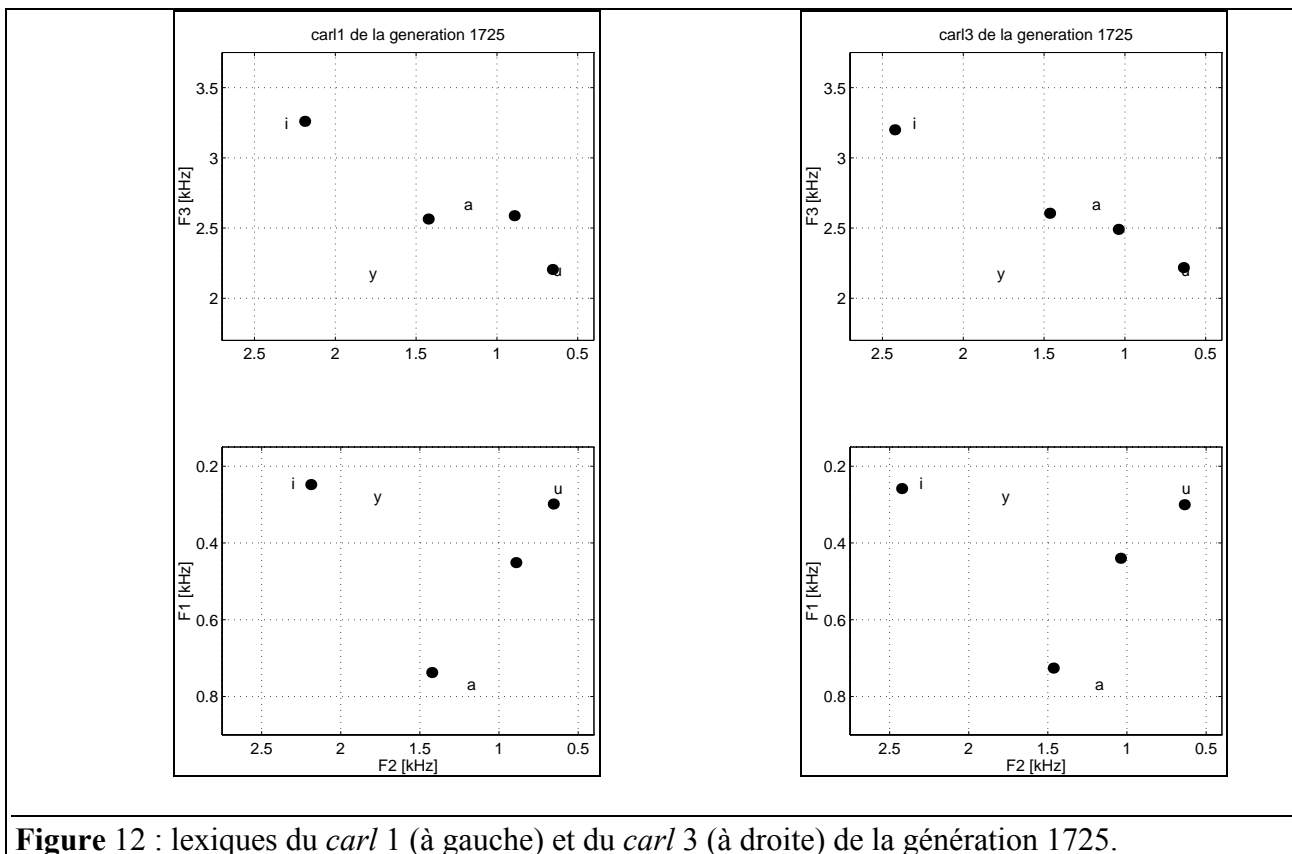


Figure 12 : lexiques du *carl* 1 (à gauche) et du *carl* 3 (à droite) de la génération 1725.

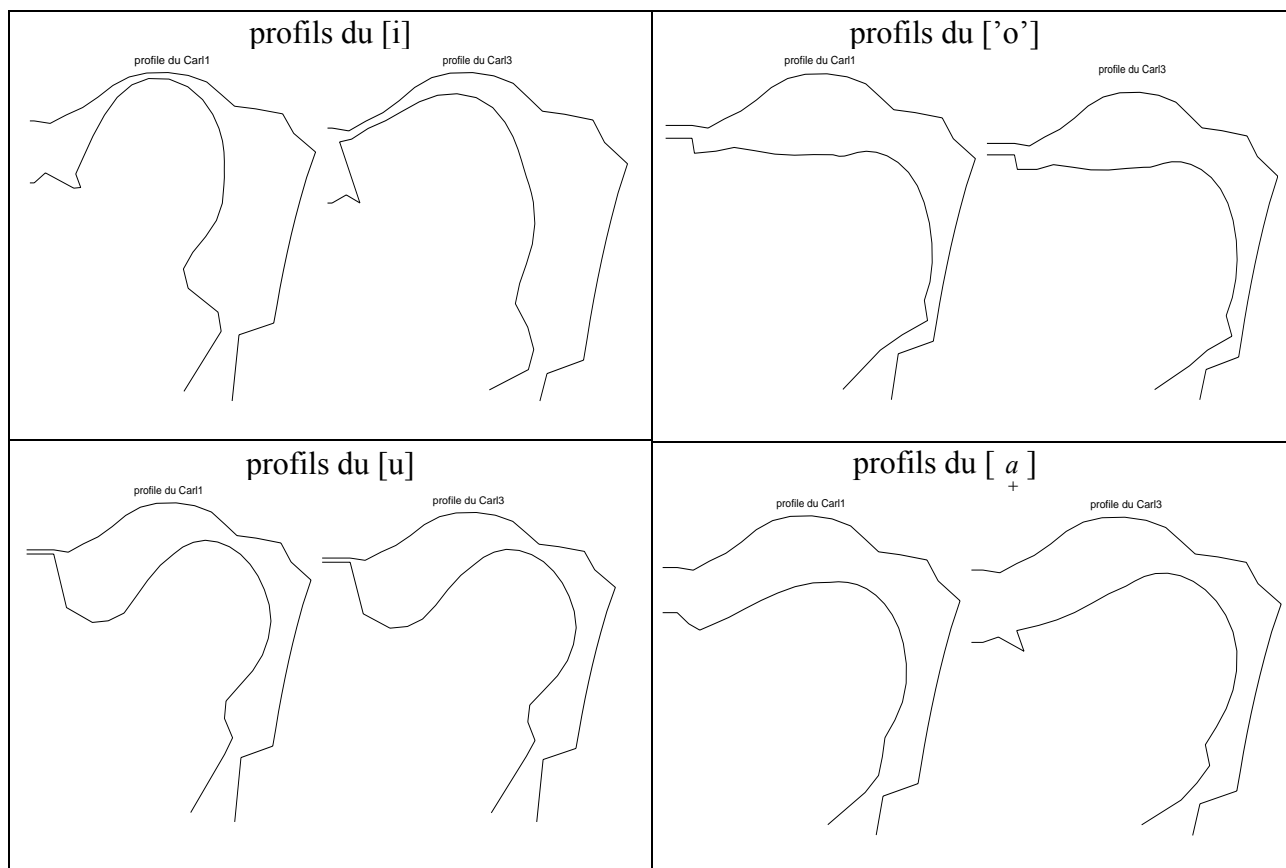


Figure 13 : coupes articutoires sagittales des *carls* 1 et 3 à la génération 1725 groupées par proximité acoustique.

6. Discussion

Nous avons montré que sans contrainte globale, uniquement par effet de coopération, des systèmes vocaliques propres à la communication parlée ont pu émerger de l'interaction entre agents sensori-moteurs. Ce sont les systèmes utilisés dans les langues du monde que l'on retrouve. Les systèmes les plus adéquats à la communication parlée émergent aux générations qui ont la fitness moyenne la plus élevée. Les produits acoustiques restent confinés dans un espace vocalique par le seul fait de la physiologie du système articuloire. De plus, chaque *carl* a élaboré sa propre stratégie articuloire pour coopérer, même si le produit acoustique est identique.

Nos résultats sont cohérents avec les statistiques sur les langues du monde (cf. Fig. 2). L'efficacité perceptive et l'économie motrice, sont les principes qui semblent régir la formation des systèmes linguistiques (Schwartz, 1995). Notre étude structure une première approche de la sorte dans le domaine de la phonétique.

Une fitness plus lisse peut être obtenue avec des populations de plus grande taille. Les fluctuations de la fitness sont dues au faible nombre de cycles d'apprentissage par rapport au nombre de voyelles, ainsi qu'à la trop petite taille de la population ce qui nuit à l'équiprobabilité des choix des voyelles lors des interactions entre les *carls*. Notre algorithme explore de la sorte l'espace des solutions, donc il n'est pas étonnant de tomber de temps en temps sur des configurations intéressantes. De récentes simulations montrent en effet que l'évolution de la fitness est similaire à celle de la simulation acoustique seule (cf. Fig. 7), lorsque l'on prend une population de 16 *carls*, pour quatre cycles d'apprentissage. Le temps des calculs prohibitifs pour l'instant des fonctions articuloires, est le frein à des simulations à plus grande échelle.

Un des résultats intéressants malgré l'effet de bord du aux valeurs des paramètres qui sont trop petits, est qu'il peut apparaître différents systèmes au cours d'une même simulation, séparés par des chutes de la fonction fitness, ce qui pourrait constituer un germe d'explication pour les phénomènes de changement phonétique.

7. Perspectives

L'éventail de recherche en phonétique qui est ouvert est vaste et novateur : l'émergence des systèmes vocaliques et syllabiques jusqu'à l'évolution phonétique des langues en recherchant des trajectoires optimales dans les phénomènes de coarticulation. L'évolution en continu dans les substrats acoustiques et articuloires offrent à nos simulations un espace de recherche très fin et inédit. Nous présentons dans la suite deux possibilités d'extension de ce travail à travers la théorie des niches écologiques, et la dynamique des changements phonétiques.

7.1. Des niches écologiques pour la formation des dialectes

Comme Lindblom et Engstrand (1990) le proposent, une quantification de facteurs socioculturels serait utile pour améliorer les prédictions : la formation et l'évolution des systèmes de sons seraient une combinaison de paramètres sociaux, perceptifs et articuloires. Nous proposons un modèle de l'émergence de dialectes dans le monde des *carls* pour donner une dimension supplémentaire à notre simulation, propre à l'étude d'apparition de dialectes, dont nous montrons la faisabilité par AGORA à travers la théorie des niches écologiques (Goldberg, 1994). Grosso (1985) a démontré que l'usage d'une géographie combinée à l'exploration génétique peut être un moyen utile d'assistance à la formation de sous populations variées, chaque sous population appartenant à sa niche écologique. Considérons qu'à un instant donné différents systèmes vocaliques sont en compétition, phénomène que nous avons observé dans nos simulations. Les niches écologiques permettent alors le développement des deux systèmes, deux dialectes, qui tendent vers leur propre optimum, grâce à des probabilités d'échanges qui sont fonction des positions géographiques des deux *carls* communiquant.

7.2. Vers un modèle de la théorie Hypo/Hyper : dynamique et étude des changements phonétiques

La théorie motrice (Lieberman et al., 1985) et la théorie Quantique (Stevens, 1989), postulent l'existence des invariants physiques en parole soit articulatoires pour l'une, soit acoustico-articulatoires pour l'autre. Cette invariance n'est pas de nature phonétique pure, mais peut être définie au niveau de la compréhension de l'auditeur : « *The invariance problem is not a phonetic issue at all for ultimately invariance can be defined only at a level of listener comprehension* » (Lindblom, 1987). Le locuteur adopterait une stratégie de Variabilité Adaptative résolvant un compromis entre la demande de compréhension et l'effort articulatoire pour réaliser la tâche. Cette adaptation du locuteur permet d'expliquer pourquoi les paramètres physiques décrivant la parole peuvent être variables pour un message linguistique invariant, ce qui est à l'origine d'un grand nombre de phénomènes de variabilité en parole. La théorie Hypo/Hyper (H&H) (Lindblom, 1995), propose aussi des explications pour les changements phonétiques. L'attention de l'auditeur peut être mobilisée par la prononciation d'un mot, autant que par son sens: « [...], on peut s'imaginer qu'une sélection, ou un rejet, d'une innovation par d'autres locuteurs est un résultat d'une évaluation qui se fait, le plus souvent, d'une manière très implicite » (Lindblom 1995). Lindblom propose un schéma (Fig. 14) de source des adaptations. La structure H&H se situerait au niveau phonologique, pour la sélection et l'adaptation.

Les variations phonétiques proviendraient des productions entre « SPEAKERS » et « LISTENERS ». Nous pourrions simuler des changements phonétiques (Labov, 1994) par AGORA en concevant des séquences de voyelles et en étudiant les phénomènes H&H. Les transitions vocaliques seraient des indissociables dans les lexiques des *carls*. Elles ne seraient pas stockées en tant que telles parce qu'elles sont susceptibles de changer selon le degré H&H. Elles seraient plutôt définies par des cibles. Elles posséderaient un caractère non-segmental du fait que les cibles n'appartiennent pas à un alphabet limité (en nombre), donné préalablement. Notre algorithme, modulo quelques modifications ferait émerger des syllabes qui maximisent à l'intérieur d'elles mêmes leur contraste perceptif entre le départ et l'aboutissement du geste, et qui minimisent leur contorsion articulatoire. Le coût articulatoire intra-syllabique serait obtenu en calculant la distance entre les paramètres articulatoires de la consonne et ceux de la voyelle; il agirait au niveau de l'adaptation, et notre algorithme de coopération agirait au niveau de la source de variation phonétique.

A plus long terme une telle étude nous amènera à mieux comprendre comment sont optimisés les parcours et négociés les objectifs de la communication parlée, ce qui est un atout pour l'élaboration de systèmes de reconnaissance automatique et de synthèse de la parole.

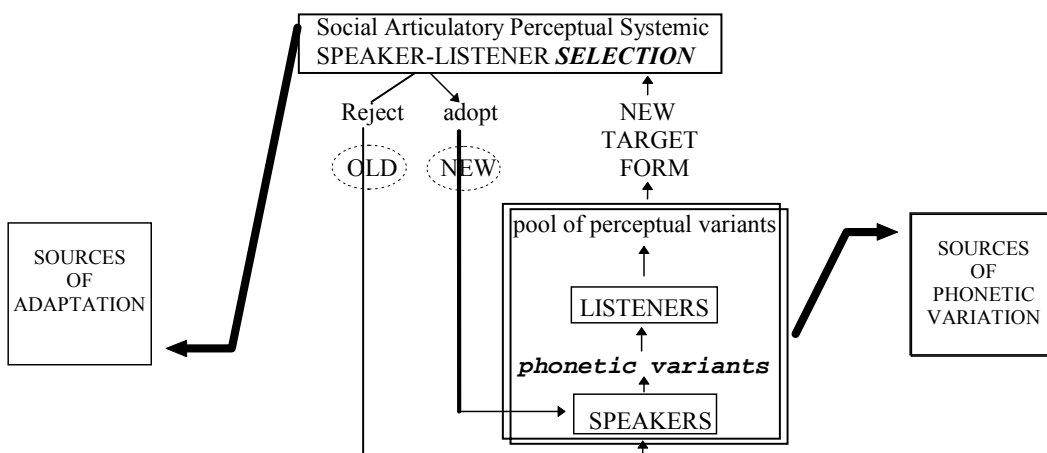


Figure 14:
d'après
« Sources of
adaptations »
Lindblom 1995

Remerciements. Nous remercions vivement Arturo Galvan pour l'implémentation des fonctions du modèle articulatoire.

Références

- Axelrod, R. (1984). *The evolution of coopération*. New York: Basic Books.
- Baker, J.E. (1985). Adaptive selection methods for genetic algorithms. *Proceedings of an international Conference on Genetic Algorithms and their applications*, pp 101-111.
- Boë, L.-J., Schwartz, J.L. & Vallée, N. (1994). The prediction of vowels systems: perceptual contrast and stability. In E. Keller (Ed.) *Fundamentals of speech synthesis, and speech recognition*, pp 185-212.
- Bosh, L.T. (1986). Architecture of vowels systems. *Proceedings from the institute of phonetic sciences of the University of Amsterdam*. The University of Amsterdam. 55-72
- Butcher, A. (1994) On the phonetics of small vowels systems: evidence from australian languages. *Proceedings of the fifth Australian International Conference on Speech Science and Technology*. Volume 1. The University of W. Australia.
- Caraty, M. (1994). *Introduction à la parole*. Laforia, Université Pierre et Marie Curie, Paris VI.
- Goldberg, D. (1994). *Algorithmes génétiques, exploration, optimisation et apprentissage automatique*. Addison-Wesley.
- Grosso, P.B. (1985). *Computer simulation of genetic adaptation: Parallel subcomponent interaction in a multilocus model*. Doctoral dissertation, University of michigan.
- Heudin, J.C. (1994). *La vie artificielle*. Paris. Hermès.
- Kugler, P. & Turvey, M.T. (1988) Self organization, Flow fields and information. *Human Movement Science*, 7,97-129.
- Laboissière, R. (1992). *Préliminaires pour une robotique de la communication parlée: inversion et contrôle d'un modèle articulatoire du conduit vocal*. Thèse de docteur de l'Institut National Polytechnique de Grenoble.
- Labov, W. (1994). *Principles of linguistic change*. Blackwell. Cambridge.
- Liberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revisited. *Cognition*, 21:1-36.
- Liljencrants & Lindblom, B. (1972). Numerical Simulation of Vowel quality systems: The role of perceptual contrast. *Language* 48, 839-862.
- Lindblom, B. (1987). Adaptive variability and absolute constancy in speech signals: two themes in the quest for phonetic invariance. *Proceedings of the Xith International Congress of Phonetic Sciences*, Tallin, Estonia, Volume 3, 9-18.
- Lindblom, B. (1990). Explaining Phonetic Variation. A Sketch of the H&H Theory. *Speech Production and Speech Modelling*, Hardcastle W.J. & Marchal A. (Eds.), Academic Publishers, pp 55-68.
- Lindblom, B. (1995). Approche intégrée de la production et de la perception. In Méloni H. (Ed.) *Fondements et perspectives en traitement automatique de la parole*. Actes de l'école thématique organisée par le GDR-PRC Communication Homme_Machine. Université d'Avignon et Pays de Vaucluse. pp 1-8.
- Lindblom, B. & Engstrand, Q. (1990). In What sense is speech Quantal. *Phon. Exp. Res.Inst. Ling. Univ. Stockholm* 8, 1-20.
- Lindblom, B., Guion, S., Hura, S., Moon, S.J. & Willerman, R. (1995, à paraître). Is sound change adaptive? Department of Linguistics, University of Texas at Austin.
- Lindblom, B., MacNeilage, P. & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. In *Explanation of Language Universals*. Mouton, pp 182-203.
- MacLennan, B.J. & Burghardt, G.M. (1994). Synthetic Ethology and the Evolution of Cooperative Communication. In *Adaptive Behavior*. Cambridge, MA: MIT Press. Vol 2, N°2, 161-188
- Maeda, S. (1988). Improved articulatory model. *Journal of the acoustical society of america*. Vol 2, pp 146.
- Meyer, J.-A. & Wilson, S.W. (1991). *Form animals to animats*. Cambridge, MA: MIT Press.
- Payan, Y. (1992) *Normalisation de l'espace de réalisation acoustique des locuteurs: intégration de contraintes de production*. D.E.A. Signal Image Parole, Institut National Polytechnique de Grenoble.
- Robert-Ribes, J. (1995). *Modèles d'intégration audiovisuelle de signaux linguistiques*. Thèse de docteur de l'INPG
- Schroeder, M.R., Atal, B.S. & Hall, J.L. (1979). Objective measure of certain speech signal degradation based on masking properties of human auditory perception. In B. Lindblom and S. Ohman (Eds.), *Frontiers of Speech Communication Research* (pp.217-229). London: Academic Press.
- Schwartz, J.L., Boë, L.-J., Vallée, N. & Abry, C. (1995). The dispersion - Focalisation Theory of vowel systems, submitted to journal of Phonetics.
- Schwartz, J.L. (1995). Perception de la parole : des représentations sensori-motrices à l'émergence de systèmes linguistiques. In *Fondements et perspectives en traitement automatique de la parole*, Méloni H.(Ed.),pp9-21
- Stevens, K.N. (1989). On the quantal nature of speech. *Journal of phonetics*, 17(1/2):3-45.
- UPSID (1981). UCLA Phonological Segment Inventory Database. UCLA WPP 53, University of California, LA.
- Vallée, N, L.-J., Boë, Chaize, M., Dequier, P., Dumay, A., Grimont, E., Perrier, P., Payan, Y. (1991). *Codebook and Vowel Prototypes of the Vocal Tract Plant*. In Speech Maps (Esprit/BR N°6975), Deliverable 17, pp 12-29.
- Vallée, N. (1994). *Systèmes vocaliques: de la typologie aux prédictions*. Thèse de docteur de l'INPG
- Varela, F.J. & Bourgine, P. (1992). *Toward a practice of autonomous systems: Proceedings of the first European Conference on Artificial Life*. Cambridge, MA: MIT Press.