MODÉLISATION RESONANTIQUE DE LA COARTICULATION DES CONSONNES OCCLUSIVES SONORES

Mhania Guerti
Ecole Nationale Supérieure Polytechnique
mhania.guerti@enp.edu.dz ou mhaniag@yahoo.fr

Résumé

Dans le message parlé, les sons sont enchaînés les uns aux autres, leurs limites sont floues : la parole résulte du mouvement continuel de l'appareil vocal, et non d'une suite de positions. Parmi les problèmes qui rendent le Traitement Automatique de la Parole délicat, il y a les difficultés liées à la variabilité omniprésente en parole continue. Celles-ci résultent du fait que l'articulation est en constante évolution et dépend même du contexte dans lequel le son est prononcé. La notion qu'un même son peut être caractérisé par une même série de formants différents est importante.

La mise à l'épreuve puis l'exploitation d'un modèle va consister à tester, interpréter, comparer les variations des caractéristiques de sortie dues aux variations des paramètres de commande. Dans ce but, un corpus $[V_1CV_2]$ ($[V_1] \equiv$ [a, i, u], $[V_2] \equiv 10$ voyelles orales et $[C] \equiv$ [b, d, g]) a été enregistré afin d'étudier la modélisation de la coarticulation. Nous inférerons des cibles consonantiques en termes de résonances en partant des cibles vocaliques adjacentes, comme points d'ancrage d'un suivi d'affiliation.

Notre étude est basée sur la représentation en résonances. Des règles de modélisation seront données. Ces dernières sont implémentées et testées par un compilateur de règles.

Mots clés

Variabilité d'un signal vocal - représentation en résonances - modélisation de la coarticulation - consonnes occlusives - compilateur de règles.

الملخص

تتوالى الأصوات في التبليغ المنطوق وتكون غير واضحة فيصعب تحديدها وتمييزها، لأن الكلام يصدر نتيجة حركة متواصلة للجهاز الصوتي وليس عن طريق مجموعة متتالية من المواضع. من بين المشاكل التي تجعل العلاج الآلي للكلام عملية دقيقة تلك الصعوبات المتعلقة بالتغيرات المتواجدة بصفة دائمة في الكلام المتواصل. وتنتج هذه التغيرات بسب التلفظ الذي هو في تطور مستمر وحتى أن موضع صدور الصوت قد يؤثر فيه. ويمكن لصوت واحد أن يتميز بنفس السلسلة من التشكلات المختلفة. إن التجربة والاستفادة من أي نموذج يتطلب اختبار وتفسير ومقارنة تغيّرات خصائص المخارج الناتجة عن مقاييس التحكم. ولهذا الغرض قمنا بدراسة بواسطة مدونة مكونة من: $[V_1C\ V_2]$ حيث ($[V_1] \equiv$ [a, i, u] ، $[V_2]$ تكافئ العشر مصوتات الشفوية، و $[C] \equiv$ [b,d,g]).

وقد سجلنا مصوتات شفوية لدراسة نموذج المخرج المزدوج انطلاقا من المصوتات المتجاورة، للوصول إلى تصويب الصوامت بالرنين. تعتمد هذه الدراسة على تمثيل الرنين باستخدام قواعد نموذجية وتّم استنتاجها بالاستفادة من النظام الحاسوبي ثم اختبرناها بواسطة معالج آلي للقواعد.

الكلمات المفاتيح

تغيير إشارة الصوت - تمثيل الرنين - نموذج الحركة المشتركة - صوامت مغلقة - المعالج الآلي للقواعد.

Abstract

In speech, sounds are linked to each other and their limits are fuzzy : the speech results of the continuous movement of the vocal apparatus and not from discrete positions. Among the problems found in the domain of Automatic Speech Processing, there is the difficulties related to the omnipresent variability in continuous speech. This latter result from the fact that the articulation is in constant evolution and depends even on the context in which the sound is pronounced. The notion that the same sound can be caracterized by a set of different formants is important.

The exploitation of a model consists to test, compare and interpret variations of features due to variations of command parameters. For this objective, a corpus of units $[V_1CV_2]$ ($[V_1] \equiv$ [a, i, u], $[V_2] \equiv$ 10 oral vowels and $[C] \equiv$ [b, d, g]) has been recorded in order to study the coarticulation modelling. We will infer consonantic targets in terms of resonances on the basis of the adjacent vocalic targets as anchorage points of a follow-up of affiliation.

Our study is based on the resonances representation. Modelling rules will be given. These are implemented and tested by a rule compiler.

Key words

Speech signal variability - resonances representation - coarticulation modelling - stop consonants - rule compiler.

Introduction

Dans la théorie quantique de la parole, il y a une antinomie entre le discret et le continu et entre le simple et le complexe [1]. La production de la parole représente un processus phonatoire finalisé par la cible auditive et les autres cibles auditives sont choisies de façon à maximiser la stabilité acoustique par rapport à la variation articulatoire [2]. La phonation et l'audition se trouvent liées par une boucle de réaction (feed-back).

L'interaction entre les sons dépend de leur nature. Par exemple, les sons alvéolaires ont une influence énorme sur les voyelles nasales qui suivent. On peut l'expliquer par le mouvement rapide de la langue. Par contre, un son labial n'affecte que peu le son qui suit, car la langue bouge à peine. Il n'impose aucune contrainte sur les postures linguales.

La coarticulation consonantique est le résultat d'une négociation entre les termes de la tâche phonologique (qui impose soit une constriction dans un certain intervalle spatial ou la réalisation de cette constriction par un articulateur donné), et un lissage articulatoire qui colore la consonne selon l'articulation vocalique adjacente. Une variation du lieu d'articulation et l'aperture labiale des consonnes, par exemple, se traduit par des loci dépendant du contexte.

Les indications que nous donnons pour le locus et les transitions formantiques s'appliquent bien à la parole synthétique et la stricte observance de ces lois suffit à obtenir une parole hautement intelligible. Des données spectrographiques suggèrent qu'il est possible d'estimer la fréquence de résonance de la cavité avant, à partir d'une information sur le signal de la parole [3, 4]. Il semble que des contributions de F_2, F_3 et des bursts des consonnes occlusives (figure 1) au lieu de la perception dynamique sont de 10 % par locus et 90 % par transition [5].

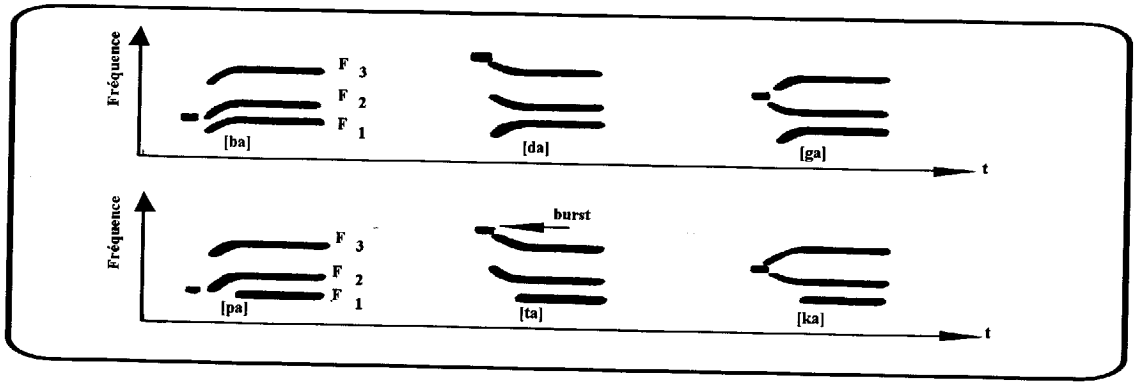

***Figure 1** : Transitions formantiques et bursts pour les syllabes [CV] ([C] ≡ occlusives ; [V] ≡ [a])*

1. Rappel sur la théorie du locus

Les recherches entreprises sur le premier synthétiseur de parole : le pattern playback, ont été innombrables. Ce synthétiseur qui a permis de transformer les spectrogrammes en sons, soit dans leur forme originale, soit dans une forme volontairement modifiée [6, 7], a été considéré à l'origine de la théorie du locus [8]. Cette dernière a connu un grand succès en phonétique et en théorie motrice de la perception (tentative d'inférence du geste vocalique et du geste consonantique) selon laquelle le message parlé est perçu par

reconnaissance des mouvements articulatoires.

Les chercheurs des Laboratoires Haskins ont révélé que les indices essentiels utilisés dans la perception de la plupart des consonnes sont les modifications que subissent les fréquences des formants lors des transitions [CV] ou [VC].

On donne le nom de locus au point de convergence de toutes les transitions de formants qui sont reliées à la perception d'un même lieu d'articulation consonantique ou de tout autre trait pertinent, quelle que soit la voyelle qui précède ou qui suit (figure 1). Ce lieu est donc, d'après P. Delattre et al., un point virtuel de la trajectoire formantique (puisque masqué par l'occlusion). Le locus était déterminé par extrapolation d'environ 50 ms à partir de l'observation des transitions formantiques pour une consonne donnée avant les différentes voyelles.

P. Delattre et al. [8] attestent *qu'il y a pour chaque consonne, des positions fréquentielles caractéristiques, ou des loci dans lesquels les transitions formantiques peuvent être vues simplement comme des mouvements de formants à partir de leurs loci respectifs aux niveaux fréquentiels appropriés pour le phonème suivant, partout où ces niveaux pouvaient être.*

L'hypothèse de départ est la suivante : étant donné que le lieu d'articulation de chaque consonne est la plupart du temps fixe, nous devrions nous attendre à trouver qu'il existe en rapport avec ceci, une position de fréquence fixe ou - locus - pour son F_2. Nous pourrions alors décrire de façon relativement simple les diverses transitions du F_2 comme des mouvements depuis ce locus acoustique jusqu'à son niveau stable dans la voyelle, et ceci quelle que soit cette voyelle.

La connaissance des loci pour les trois premiers formants de toutes les consonnes va permettre de construire des règles au niveau phonémique. Les consonnes ayant le même point d'articulation présentent la même transition de F_2, et ce sont alors les informations sur le spectre basse fréquence, qui distingueront les divers modes d'articulation.

On voit alors qu'à l'aide de 6 règles, une pour chaque point d'articulation et une pour chaque mode, on peut définir 9 phonèmes (figure 2).

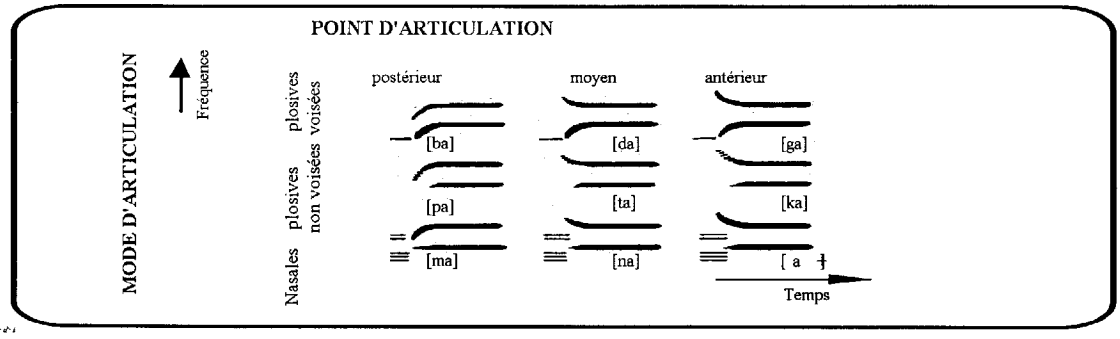

***Figure 2** : Relations entre l'articulation et les transitions formantiques des consonnes occlusives orales et nasales*

Des résultats bien connus ont montré que le F_2 contribue hautement à la perception des lieux d'articulation des consonnes occlusives dans les syllabes [CV] [8]. Néanmoins dans une parole réelle voisée, pour la voyelle [i] la région de F_3 peut être approximativement associée à la cavité avant plutôt qu'avec le F_2. Pour [a, u] dans une articulation hautement

constrictive, la région de F_2 est prédite pour être approximativement associée à la cavité avant plutôt qu'avec celle de F_3.

2. Variabilité du signal de la parole

Dans le système de la communication parlée, le processus de l'encodage-décodage passe par une négociation production - audition ou une divergence-convergence

- divergence due à la variabilité des comportements articulatoires intra-individuels et aux différentes stratégies contextuelles du locuteur (les phénomènes de la coarticulation) ;
- convergence bien obtenue, puisque la variabilité est finalement réduite pour que la communication soit assurée par l'auditeur.

Nous pouvons distinguer trois sources de variabilité liées à des différences physiologiques entre les locuteurs, aux latitudes variables de réalisation au plan linguistique et aux effets de la coarticulation.

2.1. Variabilité d'origine physiologique

Le conduit vocal d'une femme est en moyenne de 15 % plus court que celui d'un homme. En première approximation, la théorie acoustique indique que l'augmentation des fréquences de résonance est proportionnelle à la diminution de la longueur. Pour une forme de conduit vocal ayant subi une déformation uniforme, les formants sont par conséquent d'environ 15% plus élevés. Cette hypothèse de déformation uniforme est confirmée par l'observation sur des sujets réels : un adulte et un enfant (figure 3). Nous remarquons sur cette figure que les cavités antérieures ont presque la même taille et la même forme pour chaque voyelle prononcée par les deux sujets. Par contre, les cavités postérieures diffèrent complètement, car elles subissent des distorsions qui sont dues aux diminutions de la longueur disponible.

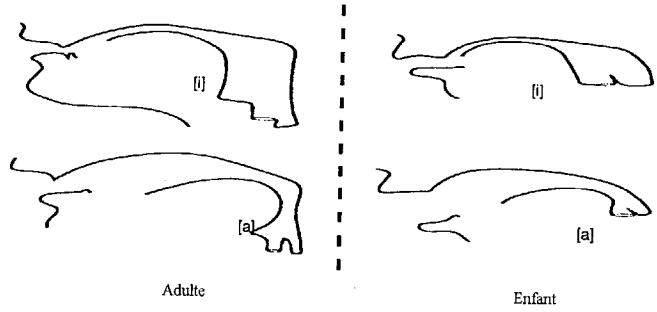

***Figure 3** : Coupes sagittales de conduits vocaux, des voyelles [i] et [a] réalisées par un adulte et un enfant*

2.2. Variabilité d'origine linguistique

Pour un même locuteur prononçant plusieurs fois un même son, la dispersion des formants est déjà importante, car l'articulation n'est jamais répétée de façon identique. Pour plusieurs locuteurs provenant d'une même région linguistique, la dispersion augmente. Plusieurs auditeurs peuvent percevoir différemment un même son à l'état isolé

(statique), prononcé par un même locuteur. Un même auditeur peut percevoir comme étant différents deux sons physiquement identiques (ceci dépend du système phonologique). Par conséquent le sexe, l'âge, le contexte, les caractéristiques du conduit vocal, les origines du locuteur et de l'auditeur, jouent un rôle très important.

2.3. Les effets de la coarticulation

La coarticulation peut être définie dans le sens large comme l'influence du contexte phonétique sur un segment déterminé. Par conséquent, elle est plus marquée dans le discours, car le son prononcé isolément n'a pas de contexte qui peut l'influencer. On a généralement estimé que les causes de la coarticulation sont à chercher dans l'inertie des muscles et des organes articulateurs [9]. Autrefois, la coarticulation était vue comme la conséquence d'une limitation d'ordre mécanique de l'appareil phonatoire. Elle est à présent considérée comme le reflet de l'organisation en unités de programmation dans la production du parlé. Elle reste un problème fondamental de la recherche phonétique. En se groupant, les sons s'influencent mutuellement et se modifient de diverses façons. Les consonnes sont soumises à l'influence acoustique des voyelles et les spectres vocaliques sont modifiés au contact des consonnes.

En prononçant les sons du langage, l'homme a tendance à obtenir le maximum d'effet avec un minimum d'effort. Une description de trajectoires acoustiques peut être simplifiée par le choix d'un espace qui interface une transformation articulatori-acoustique avec des non-linéarités minimales.

2.4. Quelques modèles de coarticulation sur le plan articulatoire

Sur le plan articulatoire, la coarticulation peut se définir comme l'influence qu'exerce un son sur un autre son contigu. Cette modification contextuelle de l'articulation résulte sans doute d'un effet d'inertie mécanique mais également d'une réorganisation du geste articulatoire, liée au concept de minimisation de l'effort articulatoire. Par exemple, pour articuler le mot [dut], le locuteur part d'une articulation antérieure apico-dentale, recule rapidement la langue vers la zone vélaire, puis doit ramener la langue en position antérieure. En débit rapide, le locuteur dispose de peu de temps pour accomplir ce déplacement important. La vitesse de déplacement de la langue doit être grande. Si le locuteur ne veut (ou ne peut) accroître l'énergie articulatoire, donc la vitesse, il lui reste la possibilité de réduire la distance à parcourir, en articulant un [u] moins extrême, plus antérieur (figure 4.a et 4.b). Ce phénomène est appelé en anglais un undershooting [10]. Sa limite est sans doute d'origine perceptive [11].

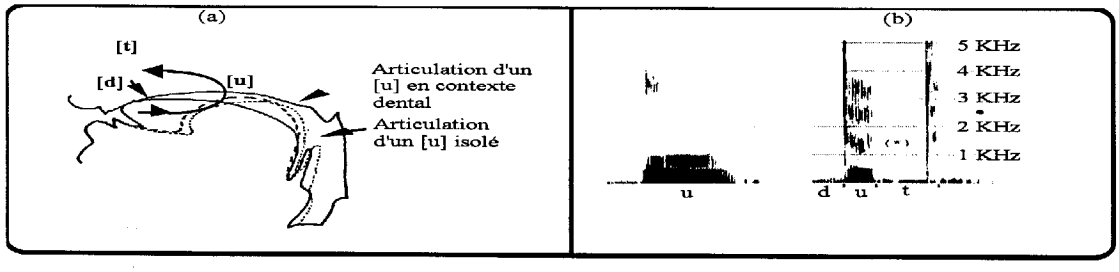

***Figure 4 a et b** : Articulation d'un [u] hors contexte et en contexte dental, dans le mot [dut] et spectrogrammes de la voyelle [u] hors contexte et en contexte dental, avec un F_2 élevé à cause de la coarticulation vers l'avant*

En étudiant la coarticulation en suédois et en anglais américain, S.E.G. Öhman élabore un modèle mathématique pour rendre compte des mouvements articulatoires en coarticulation [12]. Il propose une équation décrivant les traits essentiels de la coarticulation dans les occurrences [V_1CV_2]. Dans son modèle basé sur des relevés de films cinéradiologiques, les voyelles et les consonnes sont produites de façon différente. L'auteur suggère de décrire les mouvements de la langue par trois systèmes mécaniques contrôlables indépendamment et correspondant à des ensembles de muscles pouvant être partiellement identiques ou partiellement différents. Il attribue les noms suivants à ces trois systèmes articulateurs : apical, dorsal et le corps de la langue. Les deux premiers articulateurs permettent la production des phénomènes consonantiques, alors que le dernier est réservé aux phénomènes vocaliques. Pour modéliser les voyelles, il utilise les positions extrêmes du conduit vocal [i], [a], [u] ; par la suite toutes les voyelles seront représentées par des combinaisons linéaires de ces trois positions. L'équation proposée est la suivante :

$$S(x, t) = V(x, t) + k(t) [C(x) - V(x, t)] W C(x) \quad (1)$$

- x : distance longitudinale entre les lèvres et la glotte ;
- S(x, t) : définit la forme du conduit vocal à l'instant t durant l'occurrence [VCV] ;
- V(x, t) : composante vocalique qui est une combinaison linéaire des 3 positions extrêmes des voyelles cardinales, avec les coefficients variables dans le temps ;
- C(x) : représente la cible idéale de la consonne ;
- W C(x) : fonction de coarticulation ;
- k(t) : facteur qui varie dans le temps. Il représente le degré d'écart entre la cible idéale et la réalisation concrète sur le plan articulatoire.

Dans une analyse spectrographique des séquences [V_1CV_2] ([C] est une occlusive sonore), S.E.G. Öhman a pu montrer que [V_1] subit l'effet de la coarticulation de [V_2] à travers la consonne intervocalique. La deuxième voyelle était en quelque sorte programmée pendant la réalisation de la première voyelle. Il suggère que les mouvements linguaux sont la base pour produire les voyelles et que les mouvements nécessaires pour produire les consonnes sont superposés au substrat vocalique avec lequel ils coarticulent d'une manière plus ou moins claire. Pour produire les voyelles, le mouvement est constant et interrompu et le mouvement consonantique est simplement superposé au mouvement vocalique qui ne s'arrête jamais. Autrement dit, le geste vocalique est lent et le geste consonantique est rapide. Ceci correspond à deux canaux différents : les voyelles, canaux en raideur, et les consonnes, canaux en force.

Les conclusions tirées semblent, pour l'essentiel, confirmées par l'étude de Mac Neilage et de Declerck, qui constatent que le système articulatoire ne travaille pas à partir d'instructions données pour chaque phonème, mais plutôt à partir d'instructions pour la syllabe entière [13]. Leurs résultats confirment le point de vue que la syllabe [CV] est le seul type universellement attesté et que les effets de la coarticulation fournissent à l'auditeur des indices qui facilitent l'identification des segments.

En français, la syllabe [CV] est le cadre de la coarticulation. Cette dernière est fréquente à travers la consonne alvéolaire dans les syllabes [VtV], et beaucoup moins fréquente à travers une vélaire dans les [VkV]. Aussi, la voyelle [i] est fréquemment

source d'une coarticulation dans les deux sens (de gauche à droite et inversement). La conclusion la plus intéressante est que le degré de la coarticulation semble varier selon le lieu d'articulation. Etant donné que jusqu'à présent on n'est pas arrivé à cerner avec précision les limites de la coarticulation, on admet tout de même que son rôle est celui de garantir le caractère naturel du langage et que les transitions articulatoires d'une cible à l'autre se réalisent avec la plus grande facilité possible, sans heurts ni sauts abrupts.

La zone du conduit vocal où se situe la constriction reste relativement invariante, tandis que les autres zones sont susceptibles de varier largement en fonction du contexte [14]. Ces dernières assimilent les formes des voyelles adjacentes. Les phénomènes de la coarticulation sont donc en général fonction des différentes aires constitutives du conduit vocal. Certaines de ces aires sont très affectées tandis que d'autres sont relativement insensibles.

S.E.G. Öhman propose des améliorations à son modèle numérique de coarticulation [12]. Il confirme qu'il serait nécessaire d'établir les fonctions d'aires du conduit vocal pour que l'on puisse obtenir les équations acoustiques correspondant aux mouvements articulatoires. Il ajoute qu'il faudrait pouvoir tenir compte de la nasalité et incorporer une méthode pour contrôler les types de consonnes (sourdes ou sonores) et localiser la source. Les inconvénients majeurs de ce modèle demeurent dans son impuissance à générer des séquences Consonne - Consonne et surtout à générer des occlusives faisant intervenir l'articulateur dorsal. En effet, pour celles-ci, il paraît difficile de localiser le lieu d'articulation. Selon les entourages vocaliques de ces consonnes, les lieux d'occlusion sont très différents et en tout cas trop éloignés des écarts prévus par le modèle.

Après une période d'enthousiasme, on s'est donc rendu compte que les trois loci étaient insuffisants pour caractériser correctement les consonnes [k] et [g]. Delattre a défini un certain nombre d'indices acoustiques des consonnes françaises, qu'il a classés en indices de lieu[1] (loci de F_2 et de F_3, fréquence de bruit pour les fricatives) et de mode d'articulation (forme et rapidité des transitions formantiques, locus de F_1, présence ou absence de bruit de friction, etc.) [6]. Depuis, on a découvert d'autres indices comme le VOT (Voice Onset Time : délai d'établissement de voisement). Ainsi une consonne est caractérisée par l'ensemble des valeurs que prennent ces indices au cours d'une transition, de même qu'une voyelle est caractérisée par les fréquences formantiques F_1, F_2 et F_3. Comme il y a une grande variabilité des transitions, S.E.G. Öhman fut amené à mettre en doute la théorie du locus. Ses remarques critiques donnent l'occasion à Delattre de souligner que le concept de locus était un concept perceptuel et ne pouvait être affecté par la coarticulation [15].

En réalité, la théorie du locus souffre de très nombreuses exceptions. Nous pouvons dire, en première approximation, que les variations dans une transition [CV] ou [VC] dépendent de la consonne et de la voyelle. Toutefois, en parole spontanée, le contexte phonémique qui conditionne la transition est plus étendu [16]. Cette théorie suppose que la consonne est issue d'une position articulatoire idéale, bien fixe, quel que soit l'entourage vocalique. Cette conception s'est avérée très riche mais trop simplificatrice. On a dû admettre que le locus de F_2 pour les vélaires est largement variable et se situe à

[1] En français, l'articulation de la consonne est toujours plus en avant que celle de la voyelle.

environ 3 kHz pour les voyelles antérieures et écartées, et à environ 1000 - 1500 Hz pour les voyelles postérieures et arrondies. Ceci signifie que le point d'articulation peut être différent pour une même consonne et en particulier lorsque l'articulation est palatale. Il est décalé vers l'avant pour les voyelles antérieures et vers l'arrière pour les postérieures (il dépend de l'articulateur mis en jeu.).

Dans l'étude radiocinématographique, la coarticulation observée sur le plan articulatoire ne se reflète pas nécessairement dans les spectrogrammes. Les influences exercées par les phonèmes les uns sur les autres sont certainement explicables, dans une large mesure, par les propriétés dynamiques du système phonatoire.

La théorie du locus a permis de simplifier et de systématiser la recherche d'éléments acoustiques pertinents dans les premiers essais de parole synthétique. Mais, dès lors qu'elle était présentée comme une véritable théorie d'une perception consonantique idéale, elle se heurtait à des controverses parfaitement justifiées, en particulier par D. Klatt, pour la théorie du locus modifié [7]. D. Klatt a trouvé que cette conception du locus est très riche mais trop simplificatrice, car elle s'applique au mieux à F_2. Il avait des difficultés pour synthétiser les occlusives anglaises. Il a aussi essayé de déterminer si un concept de locus modifié pouvait être créé ou si une liste est nécessaire pour tabuler les débuts des fréquences pour les trois premiers formants avant chaque voyelle. Quand tous les points des données restent sur une ligne droite, on peut prédire le commencement de la fréquence F_2 voyelle par une équation de la forme :

$$F_2 \text{ onset} = F_2 \text{ locus} + k\,(F_2 \text{ voyelle} - F_2 \text{ locus}) \qquad (2)$$

Avec k : degré de coarticulation.

Plusieurs facteurs complexes causent l'échec de la théorie du locus. Une transition peut avoir à la fois une composante rapide et lente, due au relâchement rapide de l'obstruction suivie par des mouvements graduels du corps de la langue ; une voyelle qui suit peut influencer le F_2onset d'une transition [CV]. Le F_2 peut être relativement insensible aux constrictions orales quand il est affecté à une résonance de la cavité avant comme le cas de [i]. D. Klatt a fait une hypothèse sur les influences avant et arrière du corps de la langue et de l'arrondissement des lèvres. Il a divisé l'ensemble des voyelles en [+avant, +arrondies] et [-avant, -arrondies] et il a trouvé à l'intérieur de chaque ensemble que les données sont suffisamment régulières pour être approximées par des lignes droites (figure 5).

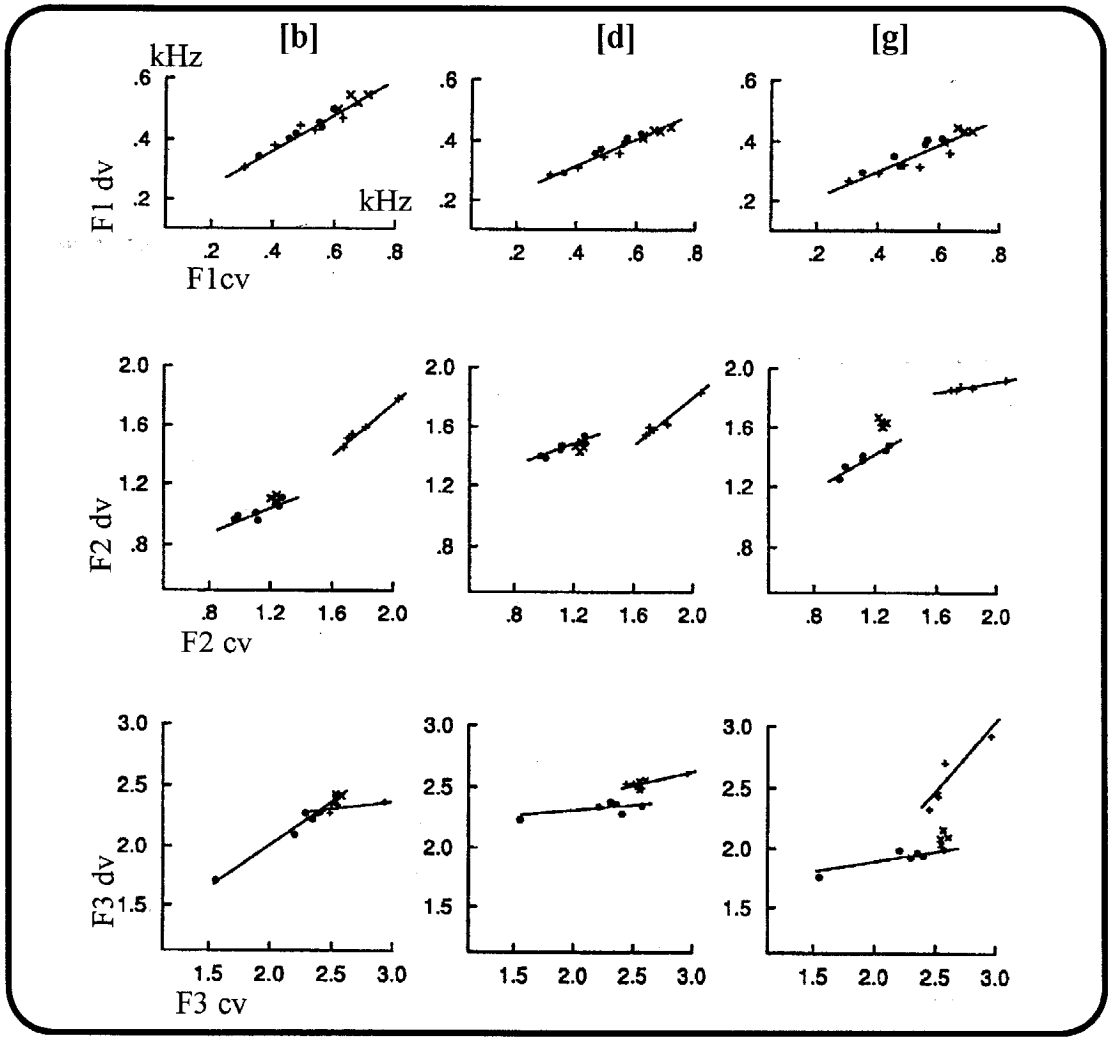

***Figure 5** : Fréquences du début de la transition de F_1, F_2, F_3 en fonction de celles de la voyelle cible, pour les 16 voyelles américaines après les consonnes [b, d, g]*

Nous avons superposé les tracés des fréquences F_2, F_3 concernant les [gV] en les remettant à la même échelle (figure 5). Nous remarquons que les portions de droites correspondant aux voyelles intermédiaires se superposent parfaitement (figure 6). Pour les voyelles basses, nous observons que $R_2 \equiv F_2$ alors que pour les voyelles hautes $R_2 \equiv F_3$. La courbe de coarticulation R_2([g]) en fonction de R_2 de la voyelle adjacente ne donne pas de saturation. Nous pouvons en faire de même pour [d].

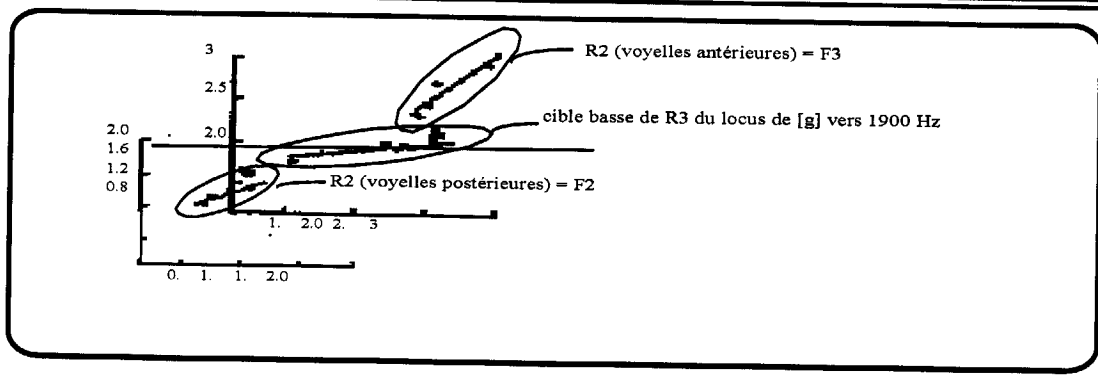

Figure 6 : Réinterprétation en résonances des schémas de coarticulation : F_2([g]) et F_3([g])

3. Modélisation de la coarticulation des occlusives sonores

Dans l'objectif de modéliser la coarticulation, nous avons constitué un corpus de logatomes du type $[V_1CV_2]$ avec [C] ≡ [b, d, g], $[V_1]$ ≡ les trois voyelles cardinales et $[V_2]$ ≡ 10 voyelles orales. L'environnement vocalique est choisi parce qu'il donne à la fois l'information sur la fermeture et la relaxation autour de la consonne. La structure théorique de résonances va être confirmée avec des données expérimentales de ces stimuli. Le nombre total des différentes combinaisons est de 90 réalisations. Nous avons utilisé deux techniques : la technique cepstrale et la LPC à Pitch Synchrone à glotte fermée. Les fréquences de R_2 sont mesurées dans les parties stables de chaque voyelle et au milieu des consonnes occlusives, de manière à assurer une bonne continuité des trajectoires.

3.1. Mesures en résonances

La première étape de ce travail a consisté à analyser un ensemble d'enregistrements effectués par deux locuteurs afin d'en extraire les paramètres spectraux pertinents, à savoir les valeurs des cinq premières résonances. Nous avons procédé à des regroupements de manière à nous trouver en présence de classes limitées, à partir desquelles une interprétation sur la nature des consonnes peut être tentée. Les fréquences de ces résonances ont été mesurées en un point médian de l'état stable : la fréquence cible de la voyelle. Pour les transitions, nous avons relevé la différence de fréquence par rapport à la partie stable de la voyelle.

3.2. Etude des trois occlusives sonores

Les consonnes occlusives fournissent un type de constriction extrême que peut produire une cavité avant bien définie. Pour étudier les consonnes, Stevens considère un modèle idéalisé constitué de deux résonateurs séparés par une constriction : l'un postérieur (fermé) et l'autre antérieur (ouvert) [1]. Il suppose qu'il n'y a pas de couplage entre eux. Si la constriction est très étroite, la fréquence de R_1 est très basse et ce sont les fréquences des résonances supérieures, et en particulier celles de R_2 et de R_3, qui sont déterminantes. Celles-ci correspondent aux fréquences propres de ces deux résonateurs. Or, lorsque l'on fait varier le rapport de l_p /l_a (respectivement la longueur de la cavité arrière ou postérieure et celle de la cavité avant), on constate que les résonances

changent de cavités en des points singuliers correspondant à l'intersection des courbes de fréquences propres (figure 7). Sur cette figure, la longueur totale du tube est de 16 cm et celle de la constriction est de 3 cm. Les lignes en tirets représentent les deux plus basses résonances de la cavité avant., Les lignes continues représentent les quatre plus basses résonances de la cavité arrière. Les lignes en pointillés près des points de convergence des deux résonances représentent les fréquences de résonances pour le cas où il y aurait une faible quantité de couplage entre les cavités avant et arrière.

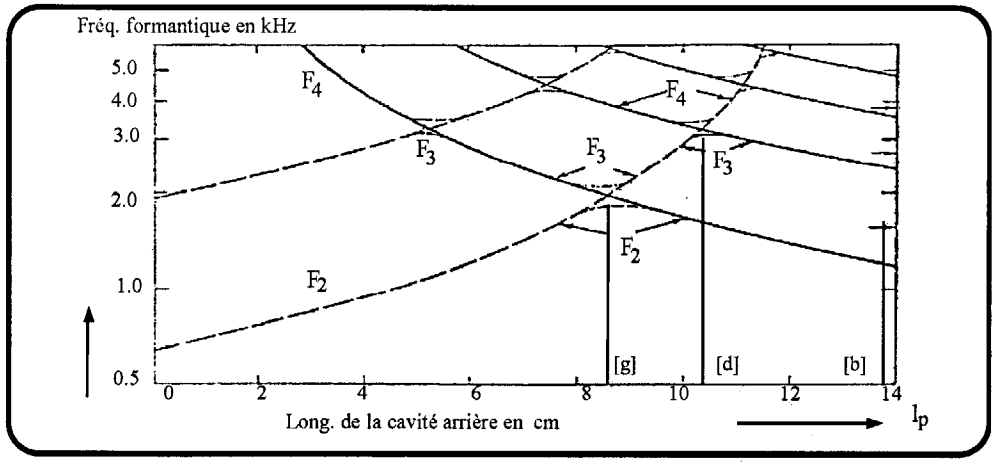

***Figure** 7 : Relations entre les fréquences naturelles et la position de la constriction*

3.2.1. L'occlusive labiale [b]

La réalisation d'une labiale consiste en une fermeture des lèvres. L'ouverture de sortie du résonateur buccal devient plus petite ou peut être supprimée totalement ; il est évident que sa fréquence de résonance est plus basse que quand il est ouvert. Mais, pour prononcer n'importe quelle labiale suivie d'une voyelle, on n'adopte aucune autre articulation que celle de la voyelle, par conséquent les transitions R -R sont toujours montantes à des degrés divers[2]. Si nous avons une [VC] (C ≡ labiale), c'est l'inverse qui se produit, les transitions des labiales étant réversibles. Cette montée universelle de R est faible pour les voyelles graves à médianes et devient nettement plus accentuée avec les voyelles aiguës. Dans les expériences de Delattre et al. [8], le locus de [b] a une fréquence plus basse que le R_2 le plus bas de toutes les qualités vocaliques utilisées (soit 700 Hz). Quant aux transitions résonantiques de R_1, elles dépendent surtout de la nature sourde ou sonore de ces consonnes. Compte tenu de ces détails, elles sont dans un [CV] toujours montantes et dans un [VC] descendantes, quelle que soit la voyelle qui précède ou qui suit (du fait de l'occlusion partielle ou complète de tout le conduit vocal).

[2] La proportionnalité de cette montée dépend en grande partie de l'anticipation vocalique (mise en position des organes comme pour prononcer la voyelle qui suit).

3.2.2. L'occlusive apico-dentale [d]

Lors de l'explosion de [d] dans les séquences [dV], le résonateur buccal est hors service du fait de l'articulation dentale. Cette dernière est très largement indépendante de celle des voyelles qui suivent ; le résonateur buccal doit donc nécessairement, avant d'être à même de produire ces voyelles, passer momentanément par des configurations géométriques différentes. Ce qui explique aisément cet aspect de divergence des transitions de R à partir d'un point unique : locus, qui est une notion bien réelle, du fait qu'il possède une véritable corrélation articulatoire et acoustique. Dans ces cas, les transitions de :

- R sont montantes à degrés divers ;
- R sont descendantes en contexte postérieur et montantes en contexte antérieur ;
- R sont toutes descendantes sauf pour le [i].

Par conséquent, avec toutes les voyelles, même si les transitions résonantiques ne correspondent pas absolument avec les loci de R - R, les transitions de ces résonances semblent toujours provenir de ces valeurs de base.

La lecture de sonagrammes peut faire naître des confusions avec des [dV], [V] ≡ voyelles aiguës, la différenciation de lecture se faisant alors surtout pour la transition de R à montée très abrupte après une labiale. L'anticipation prononcée avec les voyelles médianes limite les pentes des transitions de R_2. Elle devient nettement faible avec les voyelles aiguës, ce qui amplifie les pentes de R_2, les rendant généralement plus abruptes pour [bi] que pour [di], le tout résultant en un semblant de divergences des R de toutes les voyelles depuis un locus commun, au moins dans des [bV] artificiels soumis au choix forcé où une pente R plus accentuée suffit à différencier auditivement [di] de [bi]. Cependant, dans des [CV] naturels présentant une assez forte anticipation vocalique, il y a souvent si peu de différence entre les transitions de R_2 de [di] et de [bi] que la différenciation acoustique ne se fait plus pratiquement que par la nature du bruit d'explosion. La transition de F est importante pour le [d] dans [di] mais non pas pour le [d] dans [du]. Ce résultat peut être dû au fait que la résonance fondamentale de la cavité avant est fortement associée à F_3 de [i] et à F_2 du [u]. Par conséquent, la résonance de la cavité avant peut jouer un rôle dans la perception de [d].

4.2.3. L'occlusive vélaire [g]

Pendant l'élocution d'une vélaire suivie d'une voyelle, l'ouverture de sortie de la cavité buccale ne joue aucun rôle puisqu'elle reste absolument inchangée. Cependant, pour l'élocution de la consonne, le dos de la langue est accolé au palais, et ce dans des positions très diverses suivant la voyelle qui suit. Mais, quelle que soit l'articulation, quand on passe de la consonne à la voyelle, le dos de la langue se décolle du palais, ce qui produit nécessairement un agrandissement du volume du résonateur buccal. Naturellement, il s'ensuit un abaissement de F en cours de transition, et ce dans tous les cas, quelle que soit la voyelle attenante (sauf pour [i, e] et [y]). Le point d'articulation (point de contact du dos de la langue avec le palais) change selon la voyelle qui suit à cause du phénomène de la coarticulation. Par contre, le volume de cette cavité avant est au moment de la réalisation de la consonne, du fait de l'occlusion palatale, forcément plus

petit que lors de l'élocution de la voyelle suivante et vice-versa. Dans ce cas, la transition de R est toujours descendante à des degrés divers (exception faite pour les voyelles antérieures). La pente des transitions de R suit grossièrement le degré d'ouverture des voyelles adjacentes. Elle est forte avec les voyelles ouvertes (grande augmentation de volume du résonateur buccal en passant de la consonne à la voyelle) et devient progressivement plus faible au fur et à mesure qu'on a affaire à des voyelles plus fermées, aiguës ou graves. Autrement dit, cette pente est peu prononcée pour [gi] et [gu], et nettement plus prononcée pour [ga] par exemple. Avec toutes les qualités intermédiaires entre [a] et [o], le [g] est toujours nettement perçu ce qui confirme la validité d'un locus double pour les vélaires. Les pentes des transitions de R suivent le degré d'ouverture des voyelles attenantes. Nous savons que, lors de la prononciation d'une vélaire, la taille de la cavité avant varie beaucoup en fonction de la forte coarticulation linguale. Ce phonème devient palato-vélaire quand il est associé à des voyelles ou des consonnes à articulation antérieure ou bien vélaire dans les autres contextes (figure 8). Ceci provoque d'importantes variations dans les fréquences de résonances de R et de R . Par exemple, dans les séquences [aga] et [agØ], le R passe d'environ 1250 Hz pour le [a] à 1600 Hz pour le [Ø]. Les deux voyelles ne sont pas antérieures, donc il n'y a pas de changement d'affiliation entre les cavités lors de ce passage. Par contre, pour [iga], le R de [i] passe d'environ 3000 Hz (voire 3500 Hz) à 1250 Hz et pour [igØ] à 1600 Hz. Ceci explique la forte descente de R et le changement d'affiliation que nous remarquons : le F de [i] devient égal à R (F) de [a] ou bien de [Ø].

La figure 7 illustre bien ce changement d'affiliation. Pour $l_p < 8.5$ cm, le F_2 est produit par le résonateur antérieur et le F_3 par le résonateur postérieur. A la traversée de $l_p = 8.5$ cm, il y a échange : F_2 devient produit par le résonateur postérieur et F_3 par le résonateur antérieur. Si l'on élargit alors la constriction (relâchement de l'occlusion), F_2 et F_3 se séparent. Leurs transitions résonantiques T_2 et T_3 sont convergentes et divergent ensuite ($T_2 \equiv T_3$ après $T_2 \neq T_3$). Ce fait caractérise bien sûr les vélaires. Notons que si les transitions des résonances R -R ont une grande valeur perceptive de reconnaissance avec les voyelles médianes (où ces transitions sont accentuées), cette valeur devient beaucoup plus faible avec les voyelles aiguës et surtout graves, où la concentration spectrale de l'explosion joue un rôle acoustique prépondérant. Le fait de supprimer F dans des [gV] oblige à relever la transition de F en lui donnant une pente très supérieure à sa pente réelle. Ceci confirme la valeur perceptuelle des directions des transitions telles que celles pilotées par des loci théoriques. Delattre et al. [8] cherchent à appliquer aux vélaires la théorie perceptuelle des loci. Cependant, ils emploient une grande restriction, en imaginant deux loci virtuels distincts, l'un, dont semblent diverger les F des voyelles graves, placé vers 1 kHz, et l'autre, dont semblent diverger les F des voyelles aiguës à médianes, placé à 3 kHz. La seule catégorisation possible de [g] est obtenue avec un F droit à 3 kHz (une sorte de [i] très fermé, confinant à une qualité semi-consonantique de [j]).

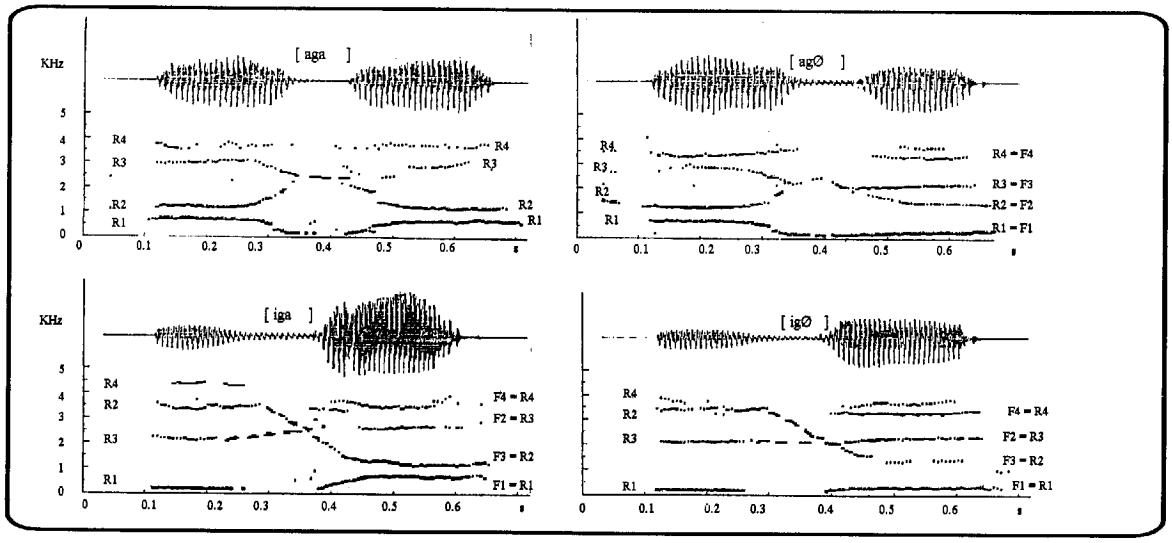

***Figure 8** : Suivi de résonances dans les séquences [agV] et [igV] [V] = [a, Ø]*

3.3. Equations des loci dans les séquences [V_1CV_2]

Conformément aux résultats de P. Delattre et al. [8], nous avons trouvé dans le cas du [g] deux loci : l'un à une fréquence basse d'environ 1200 Hz pour des contextes impliquant des voyelles postérieures et un deuxième locus haut aux alentours de 3500 Hz pour les voyelles antérieures [17, 18, 19]. Ces résultats indiquent clairement qu'en français, la voyelle subséquente [V_2] a une influence sur les transitions de la voyelle précédente [V_1]. La voyelle [a], a presque toujours des transitions plus longues que celles de [i]. Sa variation de fréquence est plus importante au contact des consonnes vélaires qu'au contact des labiales ou alvéodentales. Des trajectoires de résonance utilisent un critère de continuité et un son - une trajectoire. Des résonances sont définies pour chaque phonème par un ensemble de trajectoires connectées par des fonctions d'interpolation : X_1, X_2D, X_2G, X_3. Basées sur cette représentation, des règles sont données pour décrire une coarticulation, des trois occlusives sonores dans des contextes [V_1CV_2].

En utilisant les données acoustiques comme un point de départ, nous présentons les règles de coarticulation de [ibV_2], [idV_2], [V_1gV_2] :

$$R_2(C) = f(R_2(V2))$$

$$R_2[b] = 70\sqrt{(R_2(V2) - 1841)} \quad (3)$$

$$R_2[d] = -8\sqrt{(R_2(V2) + 2500)} \quad (4)$$

$$R_2[g] = 2300 + 1500 * Sig[0.004 * R_2(V_2) - 6] \quad (5)$$

Avec $Sig = \dfrac{(ex-1)}{(ex+1)}$

Ces règles ont été implémentées et testées à l'aide d'un compilateur de règles. Autrement dit, la fréquence d'une résonance au milieu d'une voyelle tend, suivant une fonction exponentielle de sa durée, vers une valeur fixe (cible).

4. Discussion

L'examen des spectrogrammes et le suivi de résonances des séquences [V_1CV_2] prononcées par deux sujets, suggère un type de coarticulation sensiblement différent en français, de celui constaté par S.E.G. Öhman [12] : *The articulatory system prepares for the medial consonant during all of the initial vowel.*

Ce qui est confirmé pour le français, c'est que l'influence de la voyelle subséquente ne se fait pas sentir avant la consonne. Certaines composantes de cette voyelle se réalisent déjà pendant la voyelle précédente, d'où des transitions variables en fonction de la voyelle transconsonantique. Lorsqu'on examine les séquences où les deux voyelles sont identiques [aga, igi, ...], nous constatons que les transitions de départ et d'arrivée de la même voyelle varient légèrement en étendue. La tendance est à ce que les transitions de départ soient plus longues que celles d'arrivée, ce qui semble pouvoir s'expliquer, si nous supposons que les instructions sont à chaque fois renouvelées pour le segment le plus proche. Nous pouvons également supposer que l'arrondissement labial coarticule indépendamment de la position linguale. Tout ceci reste à vérifier et constitue autant de directions de recherche pour l'avenir.

Conclusions

Cette étude montre l'importance du phénomène de la coarticulation dans les systèmes de synthèse par règles durant les perturbations consonantiques (variabilité du locus). La théorie du locus qui tente de définir chaque consonne par le point de convergence des formants vocaliques n'est pas vérifiée dans les faits. D'ailleurs, nos données confirment la proposition de D. Klatt, pour une théorie de locus variable selon la voyelle attenante.

A notre connaissance, les déformations dues au contexte n'ont pas fait l'objet d'études systématiques importantes. La segmentation de la parole continue en dépend. Il est d'ailleurs probable qu'elles soient un obstacle à la reconnaissance automatique de la parole. Il existe d'importantes variations individuelles dans l'étendue et le détail des faits de la coarticulation. Mais, il est clair que l'effet est d'autant plus marqué que le débit est plus grand. Nous pensons que, sans être indispensable à l'intelligibilité de la parole synthétique, la prise en compte de ces inflexions en contexte doit en améliorer la compréhension et surtout le caractère naturel. Quelle que soit l'explication que nous donnons à la coarticulation, il faut admettre qu'elle constitue une caractéristique importante du langage humain dans tous ses aspects : acoustique, physiologique, articulatoire et perceptuel.

Les découvertes de S.E.G. Öhman ne semblent pas être universelles. Il est probable que chaque langue ou famille de langues possède ses propres règles de coarticulation.

L'interprétation des spectrogrammes comme des structures de résonances permet une inversion acoustico-articulatoire plus claire et permet ainsi une voie facile de la modélisation des trajectoires formantiques. Des loci consonantiques ont été révisés pour tenir compte de cette nouvelle architecture. Nous pensons que cette représentation acoustique (en résonances) du signal de la parole peut suggérer une nouvelle perspective pour un nombre de problèmes-clés tel que le calcul du formant effectif F_2'.

REFERENCES

[01] K.N. Stevens, Bases for universals in the properties of the speech production and perception systems. Proc. of the IXth ICPhS. Vol 2, Copenhague, pp. 53-59, 1972.

[02] B.E.F. Lindblöm, Some phonetic null hypotheses for biological theory of language, IXth ICPhS, 2, Copenhague, pp. 33-40, 1972.

[03] G.M. Kuhn, Stop consonant place perception with single formant stimuli: evidence for the role of the front cavity resonance. J. Acoust. Soc. Am. 65 (3), pp. 774-788, 1979.

[04] H. Hermansky & D. Broad, The effective second formant F2' and the vocal tract front cavity. ICASSP, pp. 480-483, 1989.

[05] K.N. Stevens & S.E. Blumstein, The search for invariant acoustic correlates of phonetic features. In Eimas P. & Miller J. (Eds.). Perspectives on the study of speech. Hilsdale, New Jersey : Lawrence Erlbaum Associates, 1981.

[06] P. Delattre, Les attributs acoustiques de la nasalité vocalique et consonantique, les indices acoustiques de la parole. Studies in French and Comparative Phonetics, Mouton, the Hague, 1966.

[07] D.H. Klatt, Review of Text-To-Speech conversion for English. J. Acoust. Soc. Am. 82 (3), 737-1221, 1987.

[08] P., Delattre, A.M. Liberman & F.S. Cooper, Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Am. 27 (4), pp. 769-774, 1955.

[09] R. Hammarberg, On redefining coarticulation. Journal of Phonetics, N°10, 123-137, 1982.

[10] M. Guerti & G. Bailly, Anticipation et rétention dans les mouvements vocaliques du Français. XVIIIème JEP. 28-31 Mai, Montréal-Canada, pp. 292-295, 1990.

[11] Calliope, La parole et son traitement. Edition Masson, 718 pages, 1989.

[12] S.E.G. Öhman, Numerical model of coarticulation. J. Acoust. Soc. Am. 41 (2), 310-320, 1967.

[13] W. Strange, Dynamic specification of coarticulated vowels spoken in sentence context. J. Acoust. Soc. Am. 85 (5), 2135-2153, 1989.

[14] L.J. Boë, P. Perrier & G. Bailly, The geometric vocal tract variables controlled for vowel production : proposals for constraining acoustic-to-articulatory inversion. Journal of Phonetics, Vol. 20, N°1, 27-38, 1992.

[15] P. Delattre, Coarticulation and the locus theory. Studia Linguistica 23, pp. 1-26, 1969.

[16] D. Recasens, Vowel-to-vowel coarticulation in Catalan VCV sequences. J. Acoust. Soc. Am. 76 (6), 1624-1635, 1984.

[17] M. Guerti, Speech synthesis by rule, 8th International Conference on Computer Theory and Applications ICCTA'98, IEEE Alexiandria Chapter, Alexandria - EGYPT, III.12-III.15, 15-17 September 1998.

[18] M. Guerti, Speech synthesis by diphones, 9th International Conference on Computer Theory and Applications, ICCTA'99, IEEE Alexiandria Chapter, Alexandria - EGYPT, Vol.2, 278-281, 27-30 August 1999.

[19] M. Guerti, Résonance du triangle vocalique, Séminaire National sur l'Automatique et les Signaux, SNAS'99, Annaba Algérie, pp. 110-116, 9 et 10 Novembre 1999.