Prédiction phonétique de la coarticulation labiale

Vincent Robert, Anne Bonneau, Brigitte Wrobel-Dautcourt, Yves Laprie

Prédiction phonétique de la coarticulation labiale

Vincent ROBERT, Anne BONNEAU , Brigitte WROBEL-DAUTCOURT, Yves LAPRIE

Equipe Parole, LORIA UMR 7503 BP239 –54506 Vandœuvre-lès-Nancy

URL: http://parole.loria.fr email: vrobert@loria.fr

RÉSUMÉ

En vue de l'animation d'une "tête parlante virtuelle", nous avons réalisé une étude destinée à prédire la coarticulation labiale en fonction des caractéristiques des phonèmes. Pour cela, nous avons classé les phonèmes selon différents critères. En ce qui concerne les lèvres, les critères retenus sont leur degré d'ouverture, d'étirement et de protrusion. Pour tous les phonèmes de la langue française, nous avons donc estimé dans une première phase les critères précités ; pour certains d'entre eux, notamment les voyelles, tous les critères peuvent être évalués, alors que pour les consonnes seuls certains d'entre eux le sont. Par exemple, pour le son /p/, la fermeture des lèvres est indispensable, mais il n'y a pas de contrainte particulière sur leur degré d'étirement ou de protrusion. Dans une seconde phase, nous avons établi des règles de dépendance entre les critères précédents entrant en compte dans la coarticulation. Nous avons également tenu compte du fait qu'en français (langue de notre étude), la coarticulation anticipatrice est prédominante. Ceci nous a donc permis de prédire la position des lèvres en l'absence de contraintes (critères indéterminés) et donc d'en déduire l'évolution temporelle des différents paramètres labiaux. En parallèle, nous avons effectué des mesures tridimensionnelles des paramètres labiaux sur un locuteur, notre but étant de vérifier et d'affiner nos règles de coarticulation. Les premiers résultats ont permis de mettre en évidence les écarts entre les réalisations et la prédiction de notre modèle. En conséquence, nous devons réajuster nos règles de coarticulation, qui pour l'instant sont très formelles et indépendantes de tout locuteur.

1 ETUDES EXISTANTES SUR LA COARTICULATION

La parole naturelle ne correspond pas à une simple juxtaposition de phonèmes isolés. Au contraire, les phonèmes s'influencent les uns les autres, créant une perturbation appelée phénomène de coarticulation. Trois principaux modèles de coarticulation anticipatoire ont été proposés suite à de nombreuses études. Ces modèles prédisent à la fois le début et la dynamique de la coarticulation labiale, en particulier l'anticipation du geste d'arrondissement, dans des séquences V_1CV_2 et V_1CCV_2 dans lesquelles la voyelle V_1 n'est pas arrondie, la voyelle V_2 est arrondie et la (les) consonnes(s) C sont neutres vis à vis de la coarticulation labiale. La Figure 1 présente les trois principaux modèles pour une séquence VCV puis VCCV.

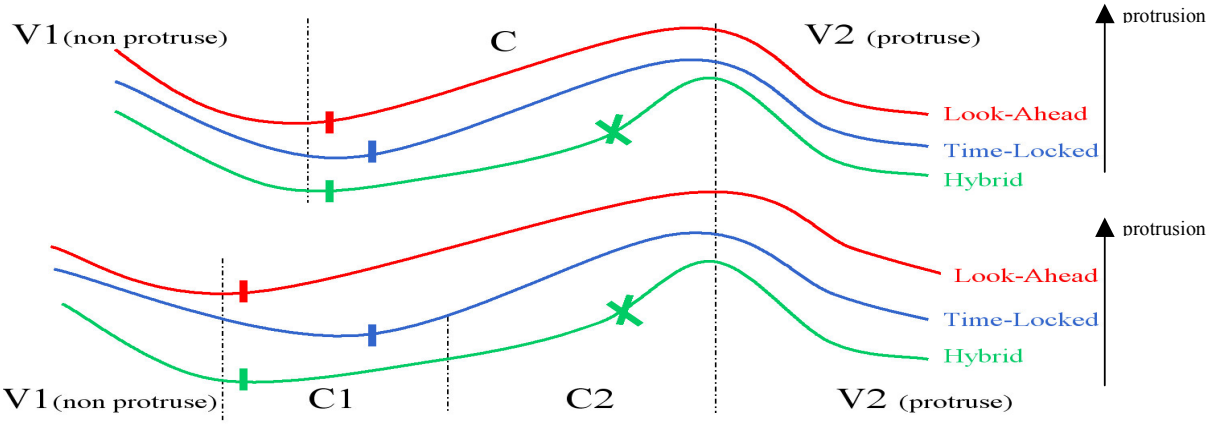

Figure 1 : Les trois principaux modèles de coarticulation
Début de la protrusion
Début de la 2ème phase du modèle hybride

Dans le modèle Look-ahead model (Henke[4], 1966), le mouvement de protrusion commence aussitôt que possible (tant qu'il n'y a pas de trait antagoniste) après la voyelle non protruse V_1.

Dans le modèle time-locked (Bell-Berti and Harris [2], 1981), le début de la protrusion commence à un instant fixe avant le début de la voyelle protruse V_2, quel que soit le nombre de consonnes intercalées. Ce modèle, défendu par l'équipe de Haskins, est repris par Boyce[3] pour une mise en évidence de gestes combinés dans un phénomène de coproduction.

Un modèle mixte aux deux précédents, le modèle hybride (Perkel and Chiang[5], 1986) a été défini pour mettre en évidence deux phases dans le mouvement. La première commence graduellement aussitôt que possible comme dans le "look-ahead model". Une seconde phase commence à un

instant fixe avant V2 comme dans le "time-locked model". Durant cette 2ème phase, des mouvements plus rapides apparaissent.

Abry et Lallouache[1] (1995) ont mis en évidence un 4ème modèle de type expansionniste. Selon eux la durée du mouvement est fortement expansible et relativement peu compressible d'où la tendance à anticiper quand le temps le permet. Ils mettent en évidence expérimentalement des coefficients d'anticipation dépendants du locuteur

Une raison importante expliquant l'existence de ces différentes théories provient du caractère limité des expérimentations. Il est indispensable, pour valider ou non un modèle de disposer d'un corpus important et d'un nombre suffisant de locuteurs afin de pouvoir étudier à la fois la variabilité inter et intra locuteurs. En outre, il est nécessaire de prendre en compte la limite des mots et de la prosodie.

2 ETUDE THEORIQUE

Les modèles précédents reposent sur l'observation d'un nombre très réduit de locuteurs. Notre but est de développer un modèle plus générique basé essentiellement sur une approche phonétique qui tente de comprendre comment les locuteurs pilotent les paramètres labiaux lors de la prononciation d'une séquence.

Nous sommes partis d'une première classification des sons en termes d'ouverture, de protrusion et d'étirement pour initialiser un modèle générique indépendant du locuteur. Cette classification s'inspire de travaux phonétiques. En parallèle, nous avons recueilli des données sur dix locuteurs qui nous ont permis de vérifier et d'affiner nos premières classifications. En ce qui concerne l'anticipation, notre modèle générique de type « look ahead » inclut des contraintes entre les différents gestes labiaux. Ce modèle doit être validé par nos données (phrases).

2.1 Paramètres labiaux des phonèmes

Le Tableau 1 montre les paramètres pour les principaux phonèmes de notre étude utilisant les lèvres. Nous avons choisi une échelle de 0 à 4 pour l'ouverture et de 1 à 4 pour l'étirement et la protrusion. Il s'agit d'un choix arbitraire. Les chiffres indiquent que le degré d'ouverture, de protrusion, et d'étirement est plus ou moins fort d'un son à un autre au sein d'une classe phonétique donnée. Par exemple, on peut dire que l'étirement croît de /a/ à /i/ en passant par /e/ et /ɛ/ pour les

voyelles antérieures non arrondies. Le niveau 0 supplémentaire pour l'ouverture est justifié par le fait que la fermeture des lèves est totale pour les bilabiales /p, b, m/ alors que le minimum de l'étirement et de la protrusion n'est pas si catégorique. Il s'agit bien à chaque fois de définir des échelles relatives et non des critères absolus (ceux-ci seront fournis par nos données).

Tous les paramètres des voyelles sont quantifiés, mais seules certaines consonnes imposent des contraintes labiales. C'est le cas des bilabiales /p, b , m/ qui nécessitent une fermeture complète des lèvres (O_0) ou de la consonne /ʃ/ qui privilégie une protrusion importante.. En ce qui concerne les semi-voyelles, la protrusion est un critère distinctif important (protrusion, importante pour /w/ et /ɥ/ et faible pour /j/).

Notons que les valeurs de protrusion, d'ouverture ou d'étirement indiquées dans le Tableau 1 correspondent à des estimations de ces paramètres pour des voyelles isolées ou des consonnes en contexte neutre. Bien sûr, ces valeurs varient en fonction du contexte et du locuteur. Il sera ainsi nécessaire lors de l'application de notre algorithme de définir un degré de résistance à la coarticulation.

L'analyse des résultats que nous faisons dans la suite (Figure 4) montre que ce tableau est globalement vérifié.

Phonème	Ouverture	Etirement	Protrusion
i	O_1	E_4	P_1
a	O_4	E_2	P_1
y	O_1	E_1	P_4
o	O_2	E_1	P_3
p	O_0		
t			
k			
f	$O_{0.5}$		
s			
ʃ			P_3

Tableau 1 : Extrait de notre classification phonétique

2.2 Interdépendance des paramètres labiaux

L'ouverture des lèvres, leur étirement et la protrusion sont trois paramètres étroitement liés. En particulier, un lien très fort existe entre protrusion et étirement qui varient en sens opposé (une anticorrélation de –0.98 a été trouvée sur nos données). La même relation existe, notamment pour les voyelles, entre ouverture et protrusion ainsi qu'entre ouverture et étirement pour les voyelles non protruses. Les règles de base sont décrites dans le Tableau 2.

Pour l'ensemble des phonèmes	**P↗ ⇔ E ↘**	**P↘ ⇔ E ↗**
Pour les voyelles et consonnes labiales	P↗ ⇔ O↘	P↘ ⇔ O↗
Pour les voyelles non arrondies	E↗ ⇔ O↘	E↘ ⇔ O↗

Tableau 2 : Liens entre protrusion, étirement et ouverture

2.3 Algorithme de coarticulation

À partir du tableau présentant les paramètres labiaux des différents phonèmes et de la relation entre ces paramètres, nous avons établi une première version d'un algorithme destiné à prédire la coarticulation labiale. L'algorithme consiste à satisfaire les règles d'interdépendance entre les paramètres énoncés ci-dessus en commençant par la fin des séquences pour tenir compte du fait que la coarticulation anticipatrice est prédominante en français. Notre modèle se rapproche du modèle Look-Ahead énoncé précédemment car nous considérons que le mouvement peut commencer dès que possible s'il n'y a pas de contraintes labiales. Mais, il ajoute les liens entre les paramètres labiaux.

Notation :

PR(Ph) : Protrusion du phonème Ph
OU(Ph) : Ouverture du phonème Ph
ET(Ph) : Etirement du phonème Ph
Ph_1 to Ph_n : Phonèmes de la séquence

Initialisation

Pour chaque phonème, recopier les paramètres du Tableau 1(s'ils existent).

```
Pour Phi ← Ph1  to Phn répéter
    |       si un des  parameters de Phi est inconnu alors
    |       |       Indiquer l'évolution de ce parameter (↗, → or ↘)
    |       finsi
FinPour
```

Corps de l'algorithme

```
Pour Phi ← Phn  à Ph1 répéter
    |       si PR(Phi) > PR(Phi-1)  alors
    |       |       si ET(Phi) est connu et ET(Phi-1) est inconnu alors
    |       |       |       si ET(Phi-1) = ↗  alors                  // incompatible avec nos règles
    |       |       |       |       ET(Phi-1) ← ET(Phi)              // stagnation
    |       |       |       finsi
    |       |       finsi
    |       |
    |       |       si Phi = voyelle ou consonne labiale alors
    |       |       |       si OU(Phi) est connu et  OU(Phi- 1) est inconnu alors
    |       |       |       |       si OU((Phi-1) = ↗ alors          // incompatible
    |       |       |       |       |       OU(Phi-1) ← OU(Phi)     // stagnation
    |       |       |       |       finsi
    |       |       |       finsi
    |       |       finsi
    |       finsi
    |
    |       //… même approche quand   PR(Phi) < PR(Phi-1)  ou
    |       // ET(Phi) < ET(Phi-1)  ou ET(Phi) > ET(Phi-1)  ou
    |       // OU(Phi) < OU(Phi-1)  ou OU(Phi) > OU(Phi-1)
finpour
```

2.4 Un exemple

Prenons l'exemple de la sequence /ipSy/. La phase d'initialisation (Tableau 3) consiste tout d'abord à reporter sur la séquence considérée la classification phonétique présentée au Tableau 1. Ensuite, une détermination du sens de variation des différents paramètres est réalisée.

Initialisation

Phoneme	/i/	/p/	/ʃ/	/y/
O	1	0		1
E	4			1
P	1		3	4

⇒

Phoneme	/i/	/p/	/ʃ/	/y/
O	1	0	↗	1
E	4	↘	↘	1
P	1	↗	3	4

Tableau 3. Report de la classification phonétique et détermination de l'évolution des paramètres

L'algorithme se déroule ensuite de la façon suivante (Tableau 3) : en partant de la fin de la séquence, une vérification des règles de dépendance entre étirement, protrusion et ouverture a lieu. Dans le cas où une incohérence dans le sens de variation de ces paramètres est détectée, une stagnation du paramètre concerné est forcée.

Corps de l'algorithme

Phoneme	/i/	/p/	/ʃ/	/y/
O	1	0	↗	1
E	4	↘	↘	1
P	1	↗	3	4

⇒

Phoneme	/i/	/p/	/ʃ/	/y/
O	1	0	1	1
E	4	↘	↘	1
P	1	↗	3	4

Impossible car la protrusion augmente

Tableau 4. Passage de /ʃ/ à /y/

Phoneme	/i/	/p/	/ʃ/	/y/
O	1	0	1	1
E	4	↘	↘	1
P	1	↗	3	4

⇒

Phoneme	/i/	/p/	/ʃ/	/y/
O	1	0	1	1
E	4	↘	↘	1
P	1	3	3	4

Impossible car l'ouverture augmente

Tableau 5. Passage de /p/ à /ʃ/

A la fin du deuxième passage (Tableau 5), les règles ont été appliquées pour l'ensemble de la séquence. Les cibles obtenues donnent une idée de l'évolution des trois paramètres mais ne permettent pas de fixer leurs valeurs avec précision. Afin d'obtenir des courbes de simulation, nous avons appliqué une approximation entre les cibles à l'aide de splines (Fonctions de lissage qui

permettent de contrôler localement les déformations). Nous avons choisi d'appliquer une approximation et non une interpolation entre les cibles, car ces dernières sont simplement des indications sur les valeurs des paramètres. Il existe un degré de liberté dans la réalisation des cibles dont on peut mieux rendre compte en utilisant des courbes de simulation. Par la suite, nous appliquerons des splines régularisantes afin de tenir compte du degré de résistance à la coarticulation des différents phonèmes. Par exemple, le phonème /p/ impose une fermeture complète des lèvres ce qui oblige la courbe d'approximation de l'ouverture à s'approcher davantage de la cible pour ce phonème.

L'approximation du premier et du dernier phonème de la séquence dépend de la position des lèvres avant et après la séquence. Nous allons considérer pour l'exemple qui suit (Figure 2) que les lèvres sont au repos en début et fin de séquence. Une analyse en composantes principales sur l'ensemble des données du corpus a mis en évidence la position de repos comme un état où les lèvres sont légèrement entrouvertes, non étirées et non protruses (O=1, E=1 ,P=1).

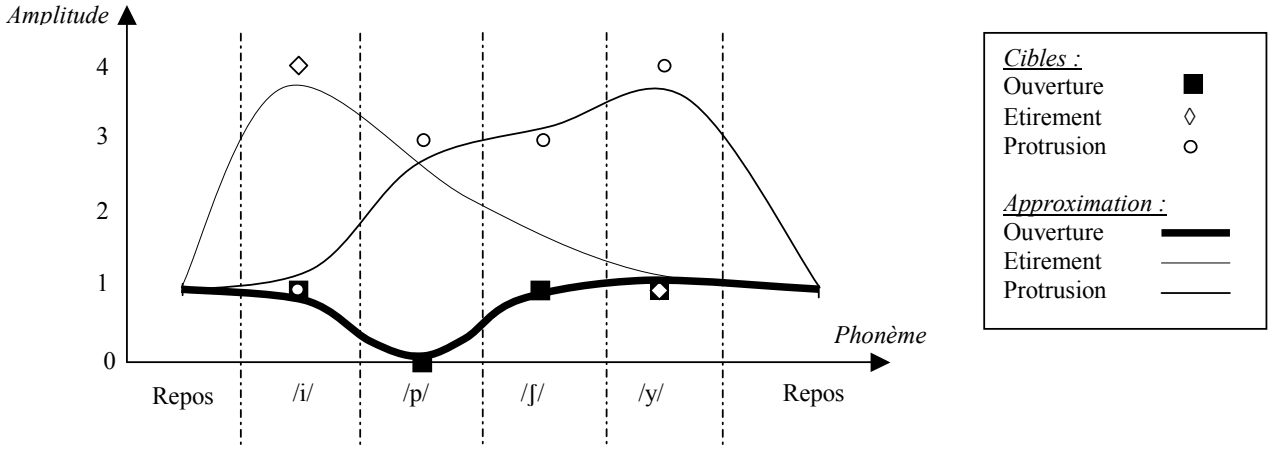

Figure 2 : Application de l'algorithme de coarticulation sur la séquence /ipʃy/

3 EXPERIMENTATION

3.1 Protocole expérimental

Nous avons développé une méthode robuste, précise et économique destinée à suivre les déformations d'un visage. Notre système, décrit en détail dans le papier de Brigitte Wrobel-Dautcourt [7], permet à partir de deux images stéréo d'un locuteur (dont le visage est peint d'une série de marqueurs) d'en déduire une reconstruction 3D (Figure 3).

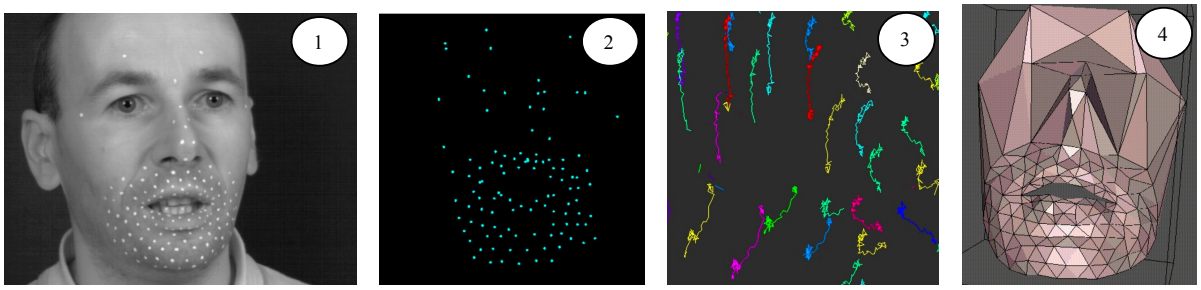

Figure 3 : Extraction des données 3D

1. Positionnement des marqueurs sur le visage
2. Identification des marqueurs
3. Suivi des déformations des marqueurs
4. Génération du maillage 3D

Dix locuteurs français (5 femmes et 5 hommes) ont été enregistrés. Notre corpus est constitué de 5 voyelles isolées (/i,y,a,o/), 6 consonnes (/p, t, d ,s, S , f /) suivies de schwa, 8 CV, 20 VCV, 18 VCCV et 2 phrases phonétiquement équilibrées. Contrairement à beaucoup d'études précédentes, nous avons aussi inclus des consonnes labiales ou ayant une influence sur les paramètres labiaux (/p , ʃ/) car notre but étant de construire une tête parlante virtuelle, nous devons nous intéresser au phénomène de coarticulation labiale dans son intégralité. Afin de limiter l'influence de la position de repos du début et de fin de séquence, les séries CV, VCV et VCCV ont été prononcées au sein d'une phrase porteuse "Trois…… lavent". Ces deux mots ont été choisis pour leur "relative neutralité" vis à vis des séquences à prononcer.

De l'ensemble des mesures, nous avons extrait nos 3 paramètres labiaux : l'ouverture des lèvres, leur étirement et la protrusion (Figure 3). L'étirement correspond à la distance entre les deux commissures tandis que l'ouverture a été choisie comme étant la distance entre un point de référence sur la lèvre supérieure et un autre sur la lèvre inférieure. En ce qui concerne la mesure de la protrusion, nous avons choisi de mesurer l'avancée de 4 points de référence (les 2 commissures ainsi qu'un point sur la lèvre supérieure et un point sur la lèvre inférieure). Contrairement à des études précédentes (Perkell et Matthies[6], Abry et Lallouache[1]) qui considèrent la protrusion comme l'avancée de la lèvre supérieure seulement, nous avons délibérément choisi cette méthode de calcul après avoir réalisé une analyse en composantes principales qui montre que les commissures et la lèvre inférieure contribuent fortement au mouvement de protrusion. La protrusion est mesurée en calculant la distance entre PF et le centre de gravité des 4 points de référence des lèvres. PF est la projection d'un point fixe du visage (par exemple le sommet du nez) sur la normale au plan formé par AB et CD.

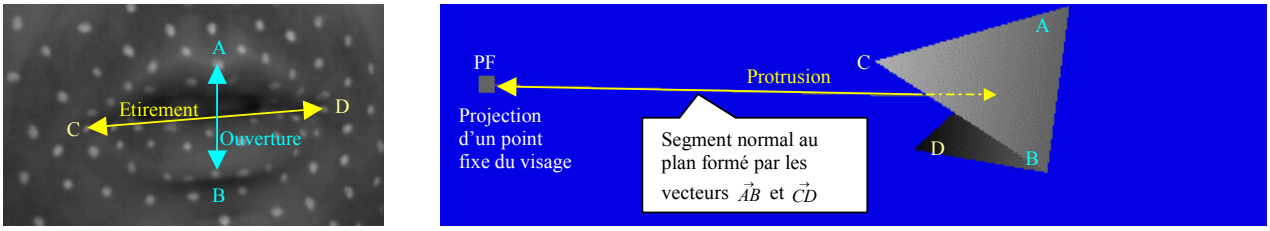

Figure 3 : Mesures de l'ouverture, de l'étirement et de la protrusion

En dépit du niveau de bruit relativement faible, nous avons appliqué un lissage à base de splines régularisantes afin de permettre une mesure plus aisée de l'établissement de la protrusion.

Les mesures obtenues en millimètres ont ensuite été centrées-réduites afin de faciliter la comparaison entre locuteurs.

3.2 Analyse des résultats obtenus

Nos résultats confirment globalement notre étude théorique sur la variation de l'ouverture, protrusion et étirement lorsqu'on passe d'un phonème à un autre (Tableau 1). On peut par exemple vérifier que l'ouverture de /a/ est plus forte que celle de /i/ et que l'étirement du /i/ est plus fort que celui du /a/.

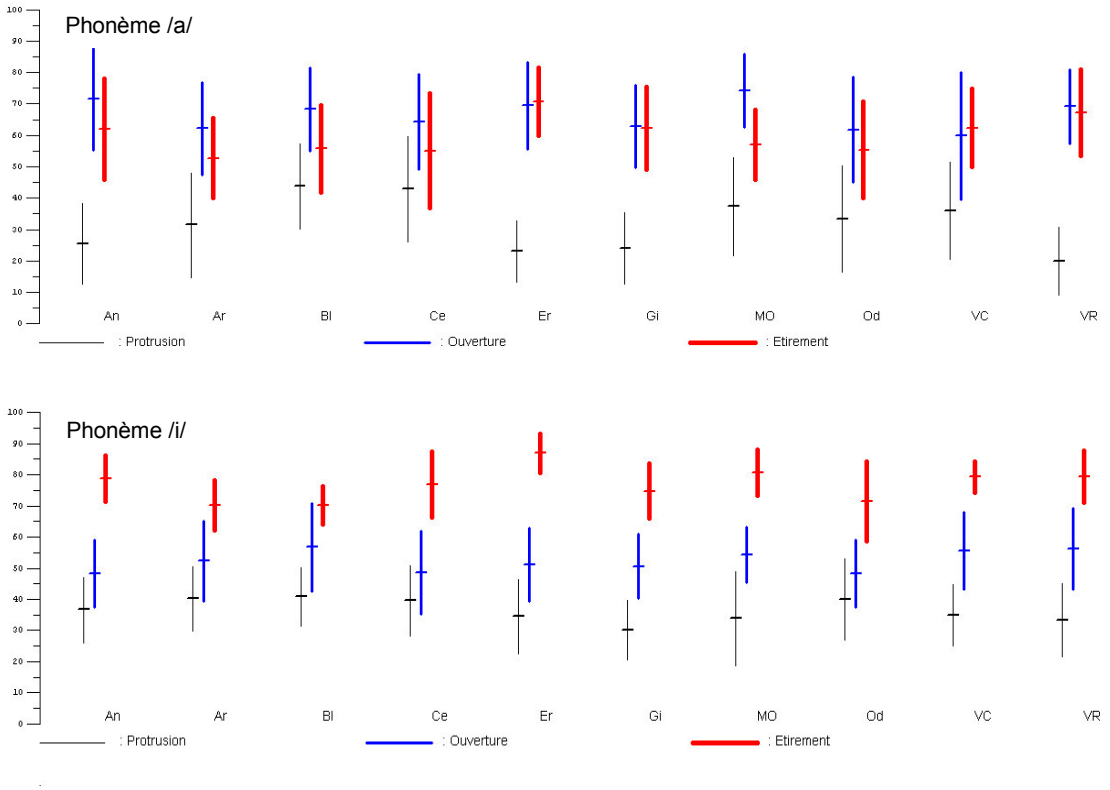

Figure 4 : Moyennes et écarts type de l'ouverture, protrusion et étirement pour les dix locuteurs lors de la prononciation des voyelles /i/ ou /a/. L'axe des ordonnées correspond à la plage de variation globale sur l'ensemble du corpus des paramètres de protrusion, d'ouverture et d'étirement. L'échelle varie de 0 (Valeur minimale) à 100 (valeur maximale).

Les premiers résultats expérimentaux montrent néanmoins une grande variabilité intra et inter-locuteurs. Pour une même séquence, les résultats diffèrent notablement entre locuteurs. Pour le démontrer, comparons la réalisation par différents locuteurs de la séquence /ipSy/ (Figure 5) à nos courbes de simulations déterminées précédemment.

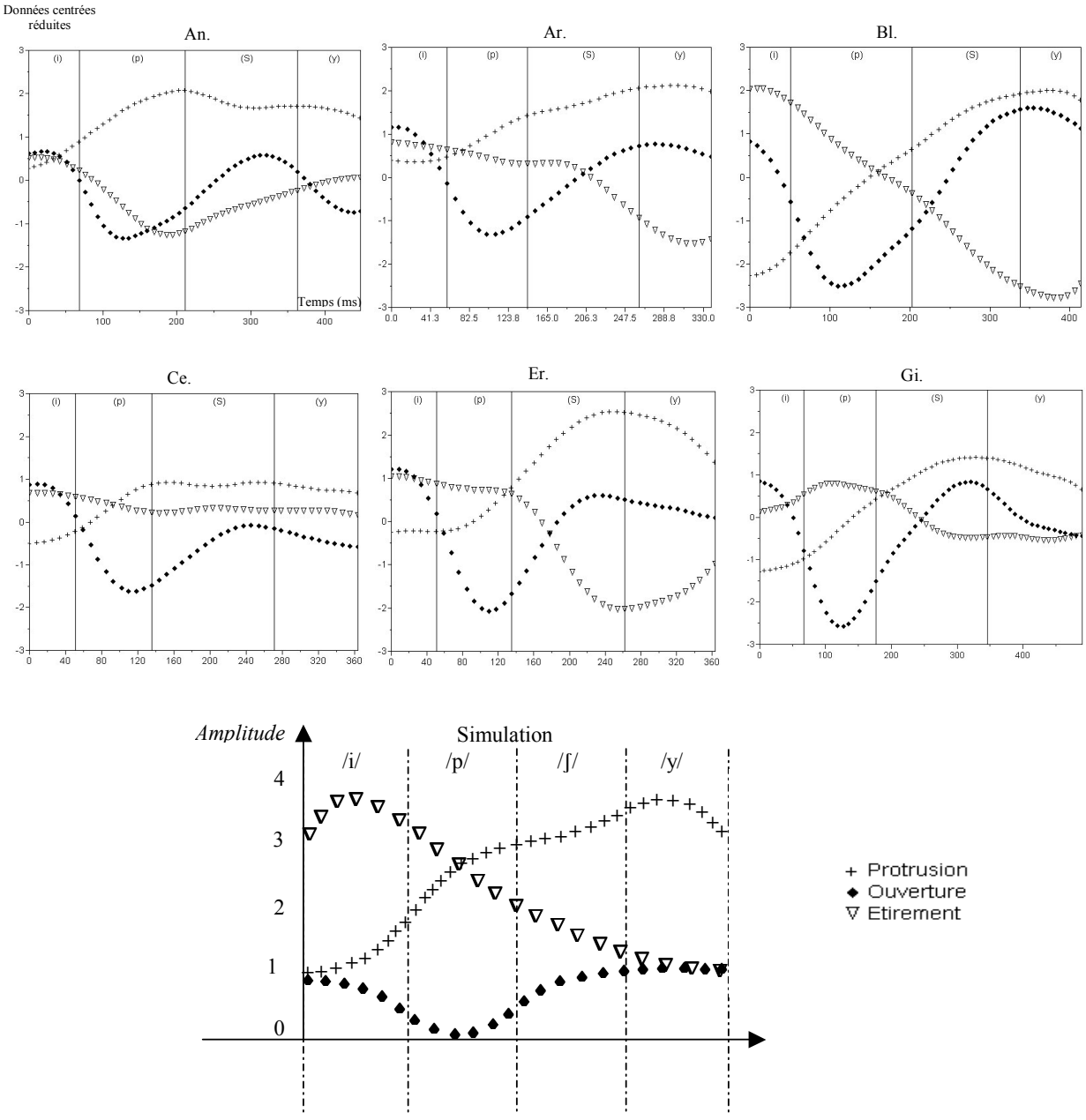

Figure 5 : Variation des paramètres d'ouverture, de protrusion et d'étirement pour la séquence /ipʃy/ .

- En haut : Séquence /ipSy/ prononcée par 6 locuteurs (abscisse = temps en millisecondes , ordonnées = Valeurs centrées réduites de la protrusion, de l'ouverture et de l'étirement.
- En bas : Courbes de simulation obtenues au paragraphe 2.4

Globalement, un lien très fort existe entre la protrusion et l'étirement qui varient en sens opposé. Une étude plus poussée a montré que sur l'ensemble du corpus, ces deux paramètres présentaient un coefficient de corrélation de –0,98.

On constate aussi une grande différence entre les courbes d'ouverture, d'étirement et de protrusion pour tous les locuteurs. En ce qui concerne la comparaison avec la courbe de simulation, il faut davantage s'intéresser à l'allure des courbes qu'à leur amplitude, car pour les courbes réelles, l'axe des ordonnées correspond à des valeurs centrées réduites alors que notre simulation correspond à une quantification de 0 à 4.

On constate en toute logique une forte baisse de l'ouverture pendant la prononciation du phonème /p/. En revanche, chez tous les locuteurs, l'ouverture est assez forte au niveau du phonème /ʃ/ ce qui n'était pas prévu par notre simulation. Ceci peut s'expliquer par notre mode de calcul de l'ouverture. Comme nous mesurons l'ouverture en calculant la distance entre un point de référence sur la lèvre supérieure et un autre sur la lèvre inférieure, cette distance augmente lors du dépliement des lèvres qui a lieu pendant la protrusion alors que l'ouverture réelle des lèvres n'a pas forcément changée. L'ouverture réelle n'étant pas mesurable directement avec notre système de stéréovision, une correction logicielle de la valeur de l'ouverture devra être effectuée. Néanmoins, il est possible de comparer l'ouverture de phonèmes ayant le même degré de protrusion. Ce problème limite la précision sur la mesure de l'ouverture mais n'invalide pas notre approche.

Quant à la décroissance de l'ouverture sur /y/, celle-ci est peut-être due au phonème suivant qui ne correspond pas à un état de repos, mais au phonème /l/, début du mot /lav/.

En ce qui concerne l'étirement, les données mesurées sont très diverses et Bl. est le locuteur qui se rapproche le plus de notre simulation. Entre An. et Gi., on constate même des variations opposées au niveau du phonème /p/. Ce paramètre semble donc avoir un grand degré de liberté lors de la réalisation de cette séquence.

La protrusion augmente comme prévu au début de la séquence. Le maximum d'accélération correspondant à l'instant de début d'établissement de la protrusion correspond assez bien avec la simulation sauf pour Er. qui commence l'arrondissement un peu plus tard que prévu. En revanche, des différences importantes apparaissent en fin de séquence. Certains locuteurs protruent davantage au niveau du /ʃ/ que du /y/. Deux explications sont possibles : soit la protrusion du /ʃ/ est plus

marquée que celle du /y/ pour le locuteur considéré, soit l'arrondissement du /y/ est réduit à cause du phonème qui le suit, en l'occurrence le phonème /l/ dans notre corpus.

4 CONCLUSION

Malgré quelques différences entre les données réelles et les données simulées, l'évolution générale des paramètres est respectée. Nous travaillons maintenant dans plusieurs directions : d'une part, il s'agit d'adapter les valeurs du Tableau 1 à chaque locuteur en vue de prendre en compte leurs caractéristiques intrinsèques. D'autre part, nous allons devoir affiner les courbes de simulation afin de prendre en compte l'importance de la résistance à la coarticulation, laquelle pourra être évaluée à l'aide des données réelles.

La validation de notre algorithme de coarticulation devra ensuite être réalisée à l'aide d'un ensemble de tests de perception proposés à des lecteurs labiaux.

5 BIBLIOGRAPHIE

[1]. Abry, C., and Lallouache, T. "Le MEM: un modèle d'anticipation paramétrable par locuteur: Données sur l'arrondissement en français*", Bulletin de la communication parlée, 3, 85-89, 1995.*

[2]. Bell-Berti, F., and Harris, K.S. " A temporal model of speech production", *Phonetica, 38, 9-20, 1981*.

[3]. Boyce, S. E. "Coarticulatory organization for lip rounding in Turkish and English", *Journal of the Acoustical Society of America, 88, 2584-95, 1990*.

[4]. Henke W. Preliminaries to speech synthesis based on an articulatory model, pages 170-171, 1967.

[5]. Perkell, J.S. and Chiang, C.M., "Preliminary support for a hybrid model of anticipatory coarticulation", *Proceeding of the XIIth International Congress of Acoustic, 1986*.

[6]. Perkell, J.S. and Mathies, M., "Temporal measures of anticipatory labial coarticulation for the vowels /u/: within-and cross subject variability", *Journal of the Acoustical Society of America, 91, 1889-95, 1992*.

[7]. Wrobel-Dautcourt B., Berger, M.O., Potard, B., Laprie, and Y., Ouni, S., "A low-cost stereovision based system for acquisition of visible articulatory data", *Audio-Visual Speech Processing, Vancouver Island, BC, Canada, 2005*.