Extraction automatique de contour de lèvre à partir du modèle CLNF

Li Liu, Gang Feng, Denis Beautemps

Extraction automatique de contour de lèvre à partir du modèle CLNF

Li Liu[1,2] Gang Feng[1,2] Denis Beautemps[1, 2]

(1) Univ. Grenoble Alpes, GIPSA-lab, F-38040 Grenoble
(2) CNRS, GIPSA-lab, F-38040 Grenoble

li.liu@gipsa-lab.grenoble-inp.fr, gang.feng@gipsa-lab.grenoble-inp.fr,
denis.beautemps@gipsa-lab.grenoble-inp.fr

RESUME

Dans cet article nous proposons une nouvelle solution pour extraire le contour interne des lèvres d'un locuteur sans utiliser d'artifices. La méthode s'appuie sur un algorithme récent d'extraction du contour de visage développé en vision par ordinateur, CLNF pour *Constrained Local Neural Field*. Cet algorithme fournit en particulier 8 points caractéristiques délimitant le contour interne des lèvres. Appliqué directement à nos données audio-visuelles du locuteur, le CLNF donne de très bons résultats dans environ 70% des cas. Des erreurs subsistent cependant pour le reste des cas. Nous proposons des solutions pour estimer un contour raisonnable des lèvres à partir des points fournis par CLNF utilisant l'interpolation par spline permettant de corriger ses erreurs et d'extraire correctement les paramètres labiaux classiques. Les évaluations sur une base de données de 179 images confirment les performances de notre algorithme.

ABSTRACT

Automatic lip contour extraction using CLNF model.
In this paper a new approach to extract the inner contour of the lips of a speaker without using artifices is proposed. The method is based on a recent face contour extraction algorithm developed in computer vision. This algorithm, which is called Constrained Local Neural Field (CLNF), provides 8 characteristic points (landmarks) defining the inner contour of the lips. Applied directly to our audio-visual data of the speaker, CLNF gives very satisfactory results in about 70% of cases. However, errors exist for the remaining cases. We offer solutions for estimating a reasonable inner lip contour from the landmarks provided by CLNF based on spline to correct its bad behaviors and to extract the suitable labial parameters A, B and S. The evaluations on a 179 image database confirm performance of our algorithm.

MOTS-CLES : modèle CLNF, spline, contour des lèvres, paramètres labiaux, parole visuelle.

KEYWORDS: CLNF model, spline, lip contour, lip parameters, visual speech.

1 Introduction

Cet article traite de l'extraction du contour interne des lèvres à partir d'enregistrements vidéo du visage « naturel » (c.à.d. sans utilisation d'artifices) vu de face dans le contexte du traitement automatique de la parole. Ce contour constitue en effet une étape indispensable pour obtenir les

paramètres portant l'information visuelle de la parole en suivant une approche forme par opposition à une approche d'apparence (voir par exemple Potamianos et al., 2012). Les bénéfices de l'information visuelle pour la perception de la parole (lecture labiale) sont bien connus. Depuis les travaux de Sumby et Pollack (1954), à ceux de Benoit et collègues pour la langue française (1992) en passant par Summerfield et collègues (Summerfield, 1979 ; Summerfield et al., 1989), il est bien établi que l'information fournie par le mouvement du visage (principalement celui des lèvres), est utilisée pour améliorer la perception de la parole dans des situations de bruit ambiant. Les expériences en shadowing (répétition de la parole de l'autre) ont montré le bénéfice de l'apport de la collaboration audiovisuelle en situation de parole « audio claire » (Reisberg et al., 1987 ; Scarbel et al., 2014). L'effet McGurk manifeste dans le cas où les informations audio et vidéo sont incohérentes la capacité d'intégrer ces informations par l'identification d'un percept différent de celui porté par chacune des deux modalités seules (McGurk and MacDonald, 1976; MacDonald and McGurk, 1978). Le contour des lèvres et les paramètres labiaux (étirement A, aperture B et aire S) qui en sont extraits sont très utiles en traitement automatique et plus particulièrement en décodage visuo-phonétique par la reconnaissance de l'articulation labiale, domaine qui connait un regain d'intérêt pour les enjeux en surveillance ou en communication avec les sourds, réel enjeu de santé publique. Historiquement pour extraire ces contours, les lèvres étaient maquillées en bleu avant l'enregistrement vidéo. Le contour interne des lèvres était alors obtenu par application d'un simple seuil dans le plan « bleu » de l'image codée RGB (Lallouache, 1990 ; Lallouache, 1991 ; Aboutabit et al., 2007).

Plusieurs travaux ont eu pour objectif de s'affranchir de l'utilisation de maquillage des lèvres. Ainsi dans le domaine de la parole, Ming et al. (2010) ont proposé d'estimer directement les paramètres labiaux par les coefficients d'une décomposition en Cosinus Discrets à 2 dimensions de la région d'intérêt des lèvres non maquillées. Dans le domaine du traitement des images les approches s'appuyant sur des modèles de contour actif multi-paramétrés ont permis pour les plus récentes méthodes de segmenter les lèvres en ajustant les contours par plusieurs polynômes d'ordre trois et l'application de seuils multiples sur le paramètre de luminance pour ce qui concerne le contour interne (Stillitano et al. 2012). Dans le domaine de la vision par ordinateur, les méthodes s'appuient sur des modèles de formes et d'apparence et le modèle CLNF se situe dans ce contexte.

L'objectif de notre travail est d'appliquer cette dernière méthode issue du domaine de la vision par ordinateur au traitement automatique de la parole dans le domaine visuel (en phonétique, analyse/synthèse, reconnaissance audio-visuelle) afin d'extraire le contour interne des lèvres sans utilisation de quelque artifice expérimental que ce soit. Nous étudions ses performances en tant que méthode générique et proposons des améliorations pour tenir compte des spécificités rencontrées en parole. Nous montrons que cette approche permet d'obtenir de manière efficace les paramètres labiaux des lèvres sans passer par des modèles de lèvres complexes.

2 La base de données visuelles

Les données sont composées d'images vidéo vues de faces de voyelles extraites d'un corpus de 50 mots isolés du Français prononcés par un sujet et précédemment enregistré dans le contexte d'un travail sur la Langue Parlée Complétée (Ming et al., 2010). On obtient des images toutes les 20 ms. L'enregistrement audio synchrone a permis de segmenter les voyelles. Dans ces intervalles on sélectionne 2 à 3 images successives correspondant à la partie stationnaire de chaque voyelle. Enfin, afin d'équilibrer la base de données, nous avons fait un tri de ces voyelles dont la répartition est donnée dans la Table 1.

L'ensemble a permis de constituer une base de 179 images. Le contour interne des lèvres a été extrait manuellement par un expert, un des auteurs. L'expert a placé une quarantaine de points décrivant fidèlement le contour interne des lèvres tout en s'assurant qu'il s'agit du contour au sens

« articulatori-acoustique ». Les paramètres (A, B, S) sont extraits à partir du contour selon la méthode classique en parole visuelle (Lallouache, 1990, 1991) et exprimés en cm ou cm^2. En traçant dans le plan (A, B), nous avons bien observé la répartition en trois groupes classiques (groupe I : voyelles ouvertes et étirées aux lèvres, groupe II : voyelles ouvertes et arrondies, groupe III : voyelles fermées et arrondies).

	Groupe I					Groupe II			Groupe III				
Voyelle	[a]	[ɛ]	[ɛ̃]	[e]	[i]	[ã]	[ɔ]	[œ]	[o]	[ø]	[õ]	[y]	[u]
Effectif	26	18	15	21	21	24	12	12	6	6	6	3	9

TABLE 1 : Répartition des voyelles dans la base de données et leur répartition en trois groupes

3 Le modèle CLNF

En vision par ordinateur, l'algorithme AAM *Active Appearance Models* (Cootes et al., 1998) introduit un modèle statistique conjoint de forme (ensemble de points placés sur le visage) et d'apparence en niveaux de gris du visage vu de face ainsi qu'un algorithme d'ajustement linéaire du modèle sur les visages. Le modèle CLM *Constrained Local Model* (Cristinacce and Cootes, 2006) applique le modèle conjoint pour générer des *templates* (imagettes rectangulaires d'une dizaine de pixels centrées sur les 68 points du modèle) qui estiment les points à partir d'une relation non linéaire. En effet, les *templates* sont utilisés pour trouver les bords des segments du visage en optimisant une fonction de réponse de surface sous une contrainte de forme. Enfin le modèle CLNF *Constrained Local Neural Field* (Baltusaitis et al., 2013) est une amélioration du modèle CLM avec l'estimation des *templates* par la méthode LNF (*Local Neural Fields*) et l'utilisation d'une fonction d'optimisation s'appuyant sur la méthode *Non-Uniform RLMS*. Dans la méthode LNF, la probabilité d'une position étant donné le *template* est calculée à partir d'une distribution sigmoïde dont les paramètres sont estimés par un réseau de neurones artificiels à noyaux convolués. *Non-Uniform RLMS* (Saragih et al., 2011) est une méthode pour minimiser une fonction de coût composée du terme RLMS dont le terme en jacobien est pondéré par la matrice de covariance des *templates*. Enfin, le modèle CLNF a été construit à partir de 4000 visages vus de face extraits des bases de données indépendantes du locuteur HELEN, LFPW et Multi-PIE (communication personnelle de Tadas Baltrusaitis). Le CLNF améliore l'estimation des *templates* du module LNF et le module d'optimisation « non-uniform » RLMS.

4 Problèmes rencontrés et solutions proposées

Dans notre expérience, pour chaque voyelle retenue dans la base de données, la méthode CLNF a été appliquée aux images de chacun des mots contenant la voyelle considérée. Nous avons retenu les points correspondants aux lèvres en particulier les 8 points du contour interne pour chacune des images de la voyelle considérée. Lors de l'application de la méthode sur notre base de données, nous avons constaté qu'elle donne d'excellents résultats pour le contour interne des lèvres dans environ 70% des cas, et ce malgré un nombre relativement faible de points (8 points seulement). Nous montrons dans la figure 1.a un exemple.

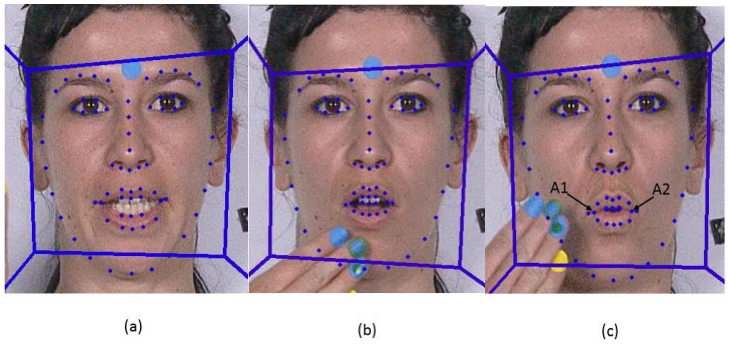

FIGURE 1: Illustration du résultat de l'application directe du modèle CLNF sur des visages de la base de donnée. On observe les 68 points distribués sur l'ensemble du visage et en particulier ceux de la région des lèvres. De gauche à droite, (a) montre des points bien placés, (b) les contours interne (et aussi externe) de la lèvre inférieure mal déterminés, (c) les points extrêmes du contour interne largement au-delà de l'ouverture des lèvres.

Cependant, une partie des images de la base de données présente des défauts manifestes. En particulier, on peut constater que le contour de la lèvre inférieure est parfois très mal déterminé (Figure 1.b). On n'a remarqué aucun problème pour la lèvre supérieure. Ce phénomène peut être expliqué par le fait que le modèle CLNF s'appuie sur un dictionnaire d'imagettes (*templates*). Si pendant la phase d'apprentissage, la région des lèvres n'a pas été bien prise en compte, il peut manquer des imagettes lors de la phase d'optimisation. Ce phénomène affecte davantage la lèvre inférieure car la région concernée est souvent très complexe (langue et des dents pouvant être partiellement visibles, voir Figure 1). Par ailleurs, nous avons constaté que pour les lèvres arrondies de faible ouverture, comme pour les voyelles [u], [y] par exemple, les deux points A1 et A2 marquant les extrémités horizontales du contour interne peuvent être bien éloignés du véritable contour au sens de celui de la relation articulatori-acoustique (Figure 1.c). En effet, d'un point de vue « géométrique », A1 et A2 ne sont pas faux car dans ce cas le contour interne peut réellement atteindre ces deux points d'extrémité. Cependant, d'un point de vue articulatori-acoustique, ces deux points ne définissent pas le paramètre d'étirement aux lèvres.

Pour avoir une idée claire sur les performances objectives de la méthode, nous comparons les points déterminés par CLNF avec le contour déterminé par l'expert. Cette comparaison se fait en termes des paramètres A, B et S, et non par une erreur quadratique moyenne entre deux contours. Le problème est délicat : comment estimer un contour à partir de seulement 8 points fournis par la méthode CLNF? On peut naturellement effectuer une interpolation linéaire permettant de relier deux points adjacents pour former un contour. Mais cette méthode présente des erreurs assez importantes vis-à-vis du véritable contour. Ceci étant dit, pour une première comparaison CLNF - contour d'expert, nous avons adopté l'interpolation linéaire car l'objectif premier est de déceler les erreurs importantes de la méthode CLNF et de les corriger. Nous proposons par la suite une interpolation non linéaire mieux adaptée au contour des lèvres. Une fois le contour déterminé, nous calculons les paramètres A, B et S. Nous traçons ensuite ces paramètres pour le contour estimé à partir des points CLNF et pour le contour d'expert, ainsi que leurs écarts (Figure 2). Nous constatons que les erreurs de CLNF concernant la lèvre inférieure mal déterminée sont clairement traduites par le paramètre B : leurs valeurs sont en général beaucoup plus petites que celles du véritable contour (Figure 2.b). En revanche, pour les erreurs correspondant aux lèvres arrondies de petite ouverture, c'est le paramètre A qui trahit l'anomalie (Figure 2.a).

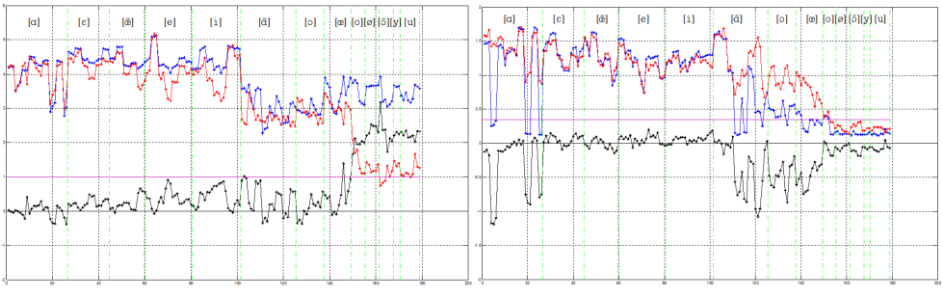

FIGURE 2 : Tracés de la valeur des paramètres A (a), B (b) issus du contour interne des lèvres déterminé par l'expert (tracé rouge) et produit par le modèle CLNF après interpolation linéaire (tracé bleu), et leur écart (tracé noir). Le trait de couleur mauve horizontal indique l'erreur quadratique moyenne. Les pointillés verticaux de couleur verte délimitent les voyelles concernées.

4.1 Solutions proposées

Etant donné la complexité de la méthode CLNF, en particulier sa phase d'apprentissage (4000 différents visages ont été utilisés), il n'est pas possible pour le moment de corriger les erreurs constatées en intervenant directement dans l'algorithme de base. Nous cherchons ainsi à corriger les défauts de l'algorithme par d'autres moyens. Ainsi nous proposons une estimation de la partie inférieure des lèvres basée sur le contraste de la luminance de l'image. En effet, sur la partie centrale des lèvres, lorsque l'on passe de la langue (ou des dents) à la lèvre inférieure, la luminance varie sensiblement. La frontière correspond à la variation la plus grande de la luminance dans la direction verticale. Certes, la recherche d'un champ de gradient dans la zone concernée permettrait de déceler les variations dans toutes les directions. Mais appliquée à des images assez bruitées cela ne donne pas de résultats convaincants. Ainsi nous décidons de chercher la position des extrema de la dérivée de la luminance dans la direction verticale, dans un intervalle délimité par les points fournis par le modèle pour la lèvre inférieure (Figure 3). Ce calcul de dérivée, très sensible au bruit, est envisageable seulement si on lisse préalablement les données. Ainsi nous proposons un lissage par spline avec un poids correctement choisi (p=0,1 valeur déterminée expérimentalement) (Feng, 1998). L'avantage de cette méthode est que la dérivée est obtenue naturellement lors du lissage. La figure 3 montre un exemple de ce lissage ainsi que la dérivée correspondante. Nous vérifions que le minimum de la dérivée correspond bien à la position de la lèvre inférieure.

Testée sur de nombreuses images contenant des erreurs de la méthode CLNF, l'estimation de la lèvre inférieure par détection de la plus grande variation de la luminance donne des résultats tout à fait satisfaisants. Il faut cependant savoir quand cette correction est nécessaire. Pour cela, nous constatons que le rapport A/B calculé à partir des points fournis par le modèle CLNF initial constitue un excellent indicateur. En effet, avec les erreurs de la CLNF, le paramètre B est anormalement petit, de telle sorte que A/B devient anormalement grand. Sachant que ce rapport A/B est relativement stable (typiquement entre 2 et 5 comme observé à travers tout le corpus), une valeur anormalement grande de A/B permet de déceler les erreurs. Ainsi nous effectuons la correction quand ce rapport est supérieur à 5. Les résultats d'évaluation par la suite confirment parfaitement cette valeur. A noter que l'utilisation d'un seuil ici n'apparait que comme indice de détection d'une anomalie de CLNF.

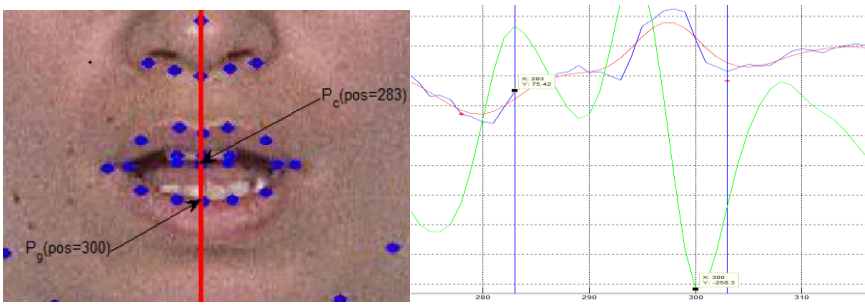

FIGURE 3 : Image illustrant une erreur de placement du contour interne de la lèvre inférieure (les points bleus prévus pour la partie inférieure se situent vers la lèvre supérieure) (à gauche). A droite, on présente la valeur de la luminance le long de la ligne verticale rouge (tracé bleu) et son lissage par spline (tracé rouge), ainsi que la dérivée du tracé lissé (tracé vert). On peut observer que le point du contour interne des lèvres correspond au minimum de la dérivée. Les 2 traits pleins verticaux de couleur bleue délimitent l'intervalle de recherche.

Nous avons évoqué que l'interpolation linéaire ne donne pas de résultats satisfaisants pour le contour estimé à partir des points CLNF, étant donné la très grande distance qui sépare les deux points d'extrémité (A1 et A2) et les autres points. Nous proposons l'utilisation d'une interpolation spline. Nous constatons que l'application de cette méthode, simple mais efficace, donne des contours excellents pour la partie inférieure des lèvres. En revanche, les trois points au centre de la lèvre supérieure forment souvent un « V », ce qui rend une interpolation spline aberrante dans le contour. Par ailleurs nous constatons que pour la partie supérieure, l'interpolation linéaire donne des résultats assez convenables, nous avons décidé de conserver l'interpolation linéaire pour la lèvre supérieure. La figure 4.a illustre le résultat d'un contour obtenu par cette méthode.

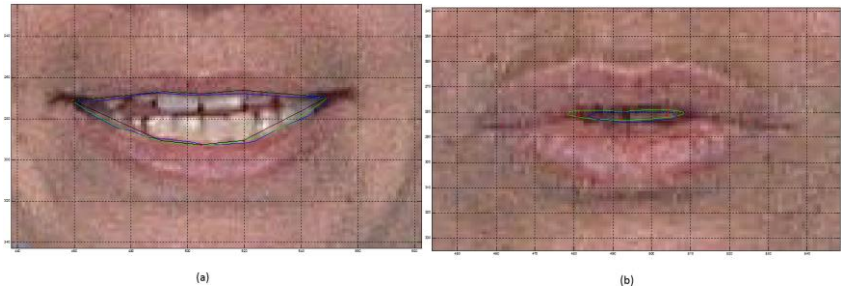

FIGURE 4 : A gauche, image illustrant l'effet de l'interpolation non linéaire. Le contour d'expert en tracé bleu, les 8 points de couleur rouge fournis par le modèle CLNF avec leur interpolation linéaire (tracé noire) et l'interpolation par spline concernant la partie inférieure (tracé vert). A droite le contour d'expert (tracé bleu), les 8 points de couleur rouge fournis par le modèle CLNF et l'interpolation par spline pour le contour interne entier mais excluant les 2 point extrêmes qui sont totalement erronés (tracé vert).

Pour les lèvres arrondies de petite ouverture, les deux points d'extrémité déterminés par CLNF sont manifestement faux. Une interpolation (linéaire ou spline) ne peut corriger cette anomalie. Remarquons que dans ce cas, les six points issus de CLNF (trois points supérieurs et trois points inférieurs) sont déterminés correctement. Nous proposons ainsi une estimation du contour uniquement

à partir de ces six points. Les deux points d'extrémité étant ignorés, on ne peut estimer un contour que si on considère la partie supérieure et la partie inférieure comme étant un ensemble et non deux parties séparées. Nous proposons une interpolation du contour entier à partir de ces six points. Voici la méthode proposée. Nous dilatons d'abord l'échelle verticale de telle sorte que les 4 coins de ces 6 points forment un carré. On convertit ensuite les coordonnées cartésiennes en coordonnées polaires pour le contour entier afin d'assurer la continuité du lissage aux deux extrémités. On effectue ensuite une interpolation avec ce système de coordonnées. Après l'interpolation, on revient à l'échelle initiale et on obtient un contour entier interpolé. Les résultats obtenus sont satisfaisants et leur évaluation est présentée dans la section suivante. Un exemple d'une telle interpolation est illustré à la figure 4.b. Ce traitement est uniquement appliqué aux lèvres arrondies de petite ouverture caractérisées par un très fort rapport A/B issu de la méthode CLNF supérieur à 8.

5 Evaluation des résultats et discussion

Nous avons évalué la méthode CLNF incluant nos propositions d'amélioration en utilisant la même base de données contenant 179 images. Nous montrons les paramètres A, B et S obtenus en intégrant toutes les propositions, comparées naturellement avec (A, B, S) du contour fourni par l'expert, ainsi que leur écart. Les résultats sont présentés à la figure 5 (ne concernant que les paramètres A et B). On constate que la correction des erreurs de la lèvre inférieure issues de CLNF donne des résultats totalement satisfaisants. En effet, dans la figure 5.b, on peut constater que toutes les erreurs présentes dans la figure 2.b ont été corrigées. Les valeurs de B suivent assez bien celles du contour de l'expert avec un écart cohérent avec les zones où on n'effectue pas cette correction. Rappelons que les résultats de la figure 5.b correspondent déjà à un contour de la lèvre inférieure interpolée par spline. Mais cette interpolation modifie très peu le paramètre B. Donc la différence entre la figure 5.b et la Figure 2.b résulte essentiellement de la correction sur la lèvre inférieure. On peut constater que l'erreur quadratique moyenne (toutes les voyelles confondues) passe de 0,3 cm pour la figure 1.b à 0,1 cm pour la figure 5.b, montrant l'efficacité de l'amélioration.

Examinons maintenant l'amélioration apportée par une interpolation du contour entier pour des lèvres arrondies de faible ouverture. Nous savons que ceci affecte essentiellement le paramètre A car CLNF donne les deux points d'extrémité beaucoup trop distants. On peut constater dans la figure 5.a que le paramètre A pour ces voyelles est beaucoup plus raisonnable, réduisant considérablement les écarts. Notons que le paramètre A n'est affecté ni par la correction sur la lèvre inférieure, ni par une interpolation non linéaire du contour inférieur, la différence entre la figure 5.a et la figure 2.a concerne uniquement les voyelles de ce groupe. L'erreur quadratique moyenne passe de 1,0 cm à 0,4 cm.

On peut remarquer que le paramètre A pour le groupe III reste supérieur à la valeur du contour expert. Ceci est essentiellement dû au fait que les six points issus de CLNF présentent en général une distance entre eux peu variable quelques soient les voyelles, ce qui constitue une limite de la méthode pour les très faibles ouvertures aux lèvres.

Nous examinons le paramètre S. La correction apportée au contour se répercute naturellement sur le paramètre S. La valeur de ce paramètre suit maintenant très bien celle issue du contour d'expert, et les écarts sont encore plus homogènes. L'erreur quadratique moyenne concernant S passe finalement de 0,89 cm^2 à 0,35 cm^2.

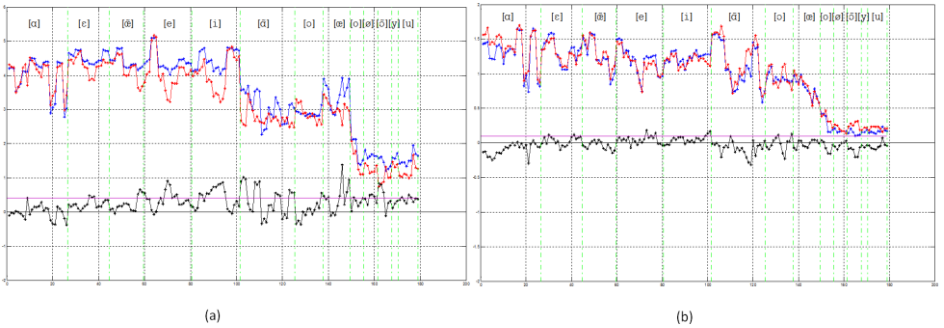

FIGURE 5 : Tracés de la valeur des paramètres A (a) et B (b) issus du contour interne des lèvres déterminé par l'expert (tracé rouge) et produit par le modèle CLNF incluant les 3 propositions de correction (tracé bleu), et leur écart (tracé noir). Le trait de couleur mauve horizontal indique l'erreur quadratique moyenne. Les pointillés verticaux de couleur verte délimitent les voyelles concernées.

Dans l'ensemble l'erreur absolue obtenue est homogène que l'on considère les grandes ou petites ouvertures aux lèvres, ce qui est caractéristique de la méthode. En conséquence l'erreur relative augmente considérablement pour les petites ouvertures ce qui diminue son intérêt. Mais pour pouvoir comparer avec la littérature nous avons reporté dans la table 2 les erreurs relatives qui se trouvent plus faibles que celles obtenues par Stillitano et al. (2013).

Erreur relative moyenne et (écart-type)	A : 13,4% (16%)	B : 9,8% (12%)	S : 13,6% (13%)

TABLE 2 : Erreur relative moyenne en % et écart-type pour les paramètres labiaux

6 Conclusion

En conclusion, nous retenons que le modèle CLNF issu du domaine de la vision par ordinateur et développé pour l'extraction de partie du visage complet reste très prometteur pour des données visuelles en production de parole. En effet, ce modèle générique permet d'extraire les paramètres du contour interne correctement dans environ 70% des cas sans aucune intervention spécifique. Nous avons montré pour le reste, qu'il a fallu développer des méthodes visant à corriger les erreurs en s'appuyant sur les points centraux fournis par le modèle CLNF. En effet, la correction a consisté à les repositionner sur le véritable contour en cherchant le maximum de contraste sur la luminance, en utilisant une interpolation spline ainsi que sa dérivée. Et dans le cas des petites ouvertures, nous avons pu proposer une interpolation spline de l'ensemble du contour interne s'appuyant sur les 6 points centraux issus du modèle CLNF. Les performances atteignent une précision de 1 mm pour le paramètre d'aperture B, de 4 mm pour le paramètre d'étirement A et de 0,35 cm^2 pour l'aire intéro-labiale S en terme d'erreur quadratique moyenne. Ces résultats sont comparables à ceux de la littérature mais sont obtenus à partir de lèvres sans maquillage. Ils indiquent que le modèle générique CLNF est tout à fait approprié. Enfin les améliorations apportées ici ne touchent pas le cœur du modèle CLNF et ses propriétés. Comme perspectives, il restera à élargir le corpus de données, en intégrant la variété des unités de parole et la variabilité liée à plusieurs locuteurs. On s'intéressera aussi à l'application de cette méthode à des situations plus complexes telles que par exemple les occlusions main-visage ou les variations dans les conditions d'enregistrement.

Références

ABOUTABIT N. (2007). Reconnaissance de la Langue Française Parlée Complétée. Manuscrit de thèse, Université de Grenoble.

AUER E.T., BERNSTEIN L.E. (2007). Enhanced Visual Speech Perception in Individuals With Early-Onset Hearing Impairment. *Journal of Speech, Language, and Hearing Research*, 50, 1157-1165.

Baltrusaitis T., Morency L.-P., and Robinson P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. In *Computer Vision Workshops (ICCV-W)*, Sydney, Australia, 2013 IEEE Conference on. IEEE, 2013.

BENOIT C., LALLOUACHE T., MOHAMADI T., ABRY C. (1992). A set of French visemes for visual speech synthesis. In: Bailly G., Benoit C. (Eds.), Talking Machines: Theories, Models and Designs. Elsevier Science Publishers, Amsterdam, pp. 485-504.

COOTES TF , EDWARDS G.J., TAYLOR C.J. (1998). Active Appearance Model. *Actes de European Conference on Computer Vision*, 484-498.

CRISTINACCE D. AND COOTES T. (2006). Feature detection and tracking with Constrained Local Models. *Actes de British Machine Vision Conference, Vol. 3,* 929-938.

FENG G. (1998). Data Smoothing by Cubic Spline Filters. *IEEE Transactions on Signal Processing*, 46, 2790-2796.

LALLOUACHE T. (1990). Un poste Visage-Parole. Acquisition et traitement des contours labiaux. *Actes des Journées d'Etudes de la Parole*, Montréal.

LALLOUACHE T. (1991). Un poste Visage-Parole couleur. Acquisition et traitement automatique des contours des lèvres. Thèse de doctorat, Institut National Polytechnique de Grenoble.

MCGURK AND MACDONALD J, 1976. "Hearing lips and seeing voices", Nature 264, 746-748.

MACDONALD J., MCGURK H., 1978. Visual influences on speech perception processes. *Perception and Psychophysics* 24, 253-257.

MING Z., BEAUTEMPS D., FENG G. AND SCHMERBER S. (2010). Estimation of Speech Lip Features From Discrete Cosine Transform. Interspeech proceedings. Tokyo, Japan.

REISBERG D., MCLEAN J., GOLDFIELD A. (1987). Easy to hear but hard to understand: a lipreading advantage with intact auditory stimuli. In: Dodd, R., Campbell, R. (Eds.), Hearing by Eye : The Psychology of Lipreading. Lawrence Erlbaum Associates Ltd, Hillside, NJ, pp. 97-113.

POTAMIANOS, G., NETI, C., LUETTIN, J., AND MATTHEWS I. (2012). Audiovisual automatic speech recognition. In G. Bailly, P. Perrier, E. Vatikiotis-Bateson (Eds), *Audiovisual Speech Processing*, pp. 193-247.

Saragih, J., Lucey, S. and Cohn, J. Deformable Model fitting by Regularized Landmark Mean-Shift. *IJCV*, 2011.

SCARBEL L., BEAUTEMPS D., SCHWARTZ J.-L. & SATO M. (2014). The shadow of a doubt? Evidence for perceptuo-motor linkage during auditory and audiovisual close-shadowing. Front. Psychol.

STILLITANO S., GIRONDEL V., CAPLIER C. (2013). Lip contour segmentation and tracking compliant with lip-reading application constraints.