République Algérienne Démocratique et Populaire
Ministère de L'enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique
Département d'Electronique
Laboratoire Signal & Communications

Thèse de Doctorat

en Electronique
Option : Signal et Communications

Présentée par :
Mr Hamidatou Mohamed Lamine
Magister en Electronique USD - BLIDA

Intitulé

Perception Visuelle de la Parole en Vue de la Lecture Labiale

Soutenue Publiquement le **17/12/2014** devant le jury composé de :

Présidente :	HAMAMI Latifa	Professeur	ENP - Alger
Rapporteur :	GUERTI Mhania	Professeur	ENP - Alger
Examinateurs :			
	CHITROUB Salim	Professeur	USTHB - Alger
	BENBLIDIA Nadjia	MCA	USD - Blida
	BENSELAMA A. Zoubir	MCA	USD - Blida

ENP 2014

Ecole Nationale Polytechnique (ENP)
10, Avenue des frères Oudek, Hassen Badi, BP, 182, 16200 El Harrach, Alger, Algérie
www.enp.edu.dz

Dédicaces

A la mémoire de mon père Ali Hamidatou qui m'a appris les bons principes de la vie ;

A ma mère Zohra Chihi qui m'a toujours entouré de son affection. Que Dieu lui accorde sa sainte miséricorde, santé et longue vie, afin que je puisse la combler à mon tour ;

A mes chers sœurs et frères, ainsi qu'à toutes leurs familles ;

A ma femme pour sa patience et son infaillible soutien ;

A mes enfants Karima, Sana, Abdelhakim et Sid Ali.

ملخـــص

من أجل تسليط الضوء على الإدراك البصري للكلمة، والمساعدة في قراءة الشفاه، عملنا هو الاستخراج التلقائي في الوقت ذاته للثوابت الخاصة بالشفاه من سلسلة فيديو بدون استعمال إضاءة أو مكياج.
دراستنا تتضمن مرحلتين هامتين: التهيئة الشبه آلية لاستخلاص المحيط الخارجي للشفاه للصورة الأولى من سلسلة فيديو، والمرحلة الثانية تتمثل في تتبع حركة الشفاه من أجل القراءة الشفاهية للكلام. لهذا اقترحنا خوارزميات لتحديد بعض النقاط المميزة للشفاه، وكذالك اقترحنا طريقة الميل الأمثل لرسم قطع تشكيلية للمحيط الأولي للشفاه.
بالنسبة لتتبع النقاط المميزة، اعتمدنا على طريقة وضع الملائمة للتقليل من أخطاء التتبع نستخدم خوارزمية إرجاع النقاط لمكانها لتتبع أفضل للمحيط الخارجي للشفاه.
وفي الأخير للتحقق من جدية خوارزمياتنا استعملنا أيضا تهيئة يدوية للنقاط المميزة. النتائج المتحصل عليها أثبتت دقة خوارزمياتنا وذلك من خلال المقارنة بين الرسوم البيانية التي تحصلنا عليها بالتهيئة اليدوية مع تلك المتحصل عليها بالتهيئة الشبه آلية.
كلمات المفاتيح : تجزئة، تتبع النقاط، المحيطات النشطة، القراءة الشفهية، سلسلة فيديو، إرجاع النقاط لمكانها، قطع.

RÉSUMÉ

Dans le but de mettre en évidence la perception visuelle de la parole, et afin de contribuer à la lecture labiale, notre travail vise particulièrement l'extraction automatique en temps réel, des contours des lèvres et des paramètres labiaux d'une séquence vidéo, sans contraintes d'éclairage ou de maquillage. Notre étude comporte deux étapes essentielles : l'initialisation semi-automatique pour l'extraction du contour externe des lèvres de la première image d'une séquence vidéo, et le suivi des mouvements des lèvres en vue de la Lecture Labiale (L.L). Pour cela, nous avons proposé des algorithmes pour la détermination de certains Points Caractéristiques (PC) des lèvres, ainsi qu'une méthode de pente optimale permettant de tracer les cubiques formant les contours initiaux des lèvres. Pour le suivi des PC, nous nous sommes basés sur la méthode de la mise en correspondance. La minimisation des erreurs de suivi se fait à l'aide d'un algorithme de recalage afin de suivre convenablement les contours des lèvres. Finalement, pour s'assurer de la performance de nos algorithmes, nous avons utilisé aussi une initialisation manuelle des PC. Les résultats obtenus ont montré la rigueur de nos algorithmes et cela en comparant les graphes obtenus par l'initialisation manuelle avec ceux de l'initialisation semi-automatique.

Mots clés : Segmentation, Suivi des points, Contours actifs, Lecture Labiale, Séquence Vidéo, Recalage des points, Cubiques.

ABSTRACT

with the aim to show the visual perception of the word, and to contribute to lip reading our work aim particularly to the automatic extraction in real time, the contours of the lips and labial parameters of a video sequence, without constraints of lighting or make-up. Our study comprises two crucial steps: semi-automatic initialization for the extraction of the external contour of the lips of the first image of a video sequence, and the follow-up of the movements of the lips for Lip Reading. For that, we proposed algorithms for the determination of certain characteristic points of the lips, as well as a method of optimal slope allowing to trace the cubic ones forming initial contours of the lips. For the follow-up of the characteristic points, we based ourselves on the method of the mapping. The minimization of the errors of follow-up is made using an algorithm of retiming in order to follow contours of the lips suitably. Finally to make sure of the performance of our algorithms, we also used a manual initialization of the characteristic points. The results obtained showed the rigor of our algorithms, and that by comparing the graphs obtained by manual initialization with those of semi-automatic initialization.

Key words: Segmentation, Tracking points, Active Contours, Lip Reading, Video Sequence, Retiming points, Cubic.

Remerciements

Je commence par remercier ALLAH, le tout puissant, qui m'a donné le courage, la force et la volonté pour bien mener ce travail.

A l'issue de la rédaction, je suis convaincu que la thèse est loin d'être un travail solitaire. En effet, je n'aurais jamais pu réaliser ce travail doctoral sans le soutien d'un grand nombre de personnes dont la générosité, la bonne humeur et l'intérêt manifesté à l'égard de ma recherche m'ont permis de progresser dans cette phase délicate.

En premier lieu, je tiens à remercier ma directrice de thèse, Madame GUERTI Mhania, professeur au Département d'Electronique, Ecole Nationale Polytechnique d'Alger pour la confiance qu'elle m'a accordée en acceptant d'encadrer ce travail de thèse, pour ses multiples conseils et pour toutes les heures qu'elle a consacrées à diriger cette recherche. C'est grâce à ses orientations, ses encouragements et sa confiance que j'ai pu mener à terme ce travail. Qu'elle trouve ici ma profonde reconnaissance pour tout ce qu'elle a fait pour moi.

J'exprime ma gratitude à Madame HAMAMI Latifa, professeur au Département d'Electronique, Ecole Nationale Polytechnique d'Alger, d'avoir accepté de présider le Jury de la soutenance de cette thèse, Je la remercie également pour son soutien moral à chaque fois que j'ai sollicité son aide, ainsi que pour ses multiples encouragements.

Mes remerciements vont également à Monsieur CHITROUB Salim, professeur au Département d'Electronique de l'USTHB, pour avoir accepté de participer à ce jury de thèse. Pour m'avoir accompagné avec tant de sympathie et conseils. Veuillez trouver ici le témoignage de ma considération et de ma gratitude.

Je tiens à remercier Madame BENBLIDIA Nadjia, Maître de conférences au Département d'Electronique, de l'Université Saâd Dahleb Blida, qu'elle me fait

l'honneur de juger mon travail. Veuillez accepter mes vifs remerciements et mon respect.

A Monsieur BENSELAMA Abdeslem Zoubir Maître de conférences au Département d'Electronique, de l'Université Saâd Dahleb Blida, pour avoir accepté de participer à ce jury, et d'avoir apporté son soutien et son expérience. Qu'il trouve ici le témoignage de mon respect.

J'adresse toute ma gratitude et mes remerciements à Monsieur AIT-AOUDIA Samy, professeur à l'Ecole Supérieure d'Informatique -ESI- et Chef d'équipe Med-Ima au laboratoire de recherche LMCS, pour son soutien constant et sa disponibilité à tout moment.

Ainsi qu'à tous mes collègues chercheurs. Je remercie toutes les personnes formidables que j'ai rencontrées par le biais des écoles ESI, ENP et ENSA. Merci pour votre soutien et vos encouragements.

Je suis particulièrement reconnaissant à Monsieur HAMADI Billel et Madame MEHDAOUI Lalia de l'intérêt qu'ils ont manifesté à mon égard.

Enfin, les mots les plus simples étant les plus forts, j'adresse toute mon affection à ma famille, et en particulier à ma Mère qui m'a fait comprendre que la vie n'est pas faite que de problèmes qu'on pourrait résoudre grâce à des formules mathématiques et des algorithmes. Son intelligence, sa confiance, sa tendresse, son amour me portent et me guident tous les jours, je la remercie pour avoir fait de moi ce que je suis aujourd'hui.

Une pensée pour terminer ces remerciements pour toi, mon Père, qui n'a pas vu l'aboutissement de mon travail, mais je sais que tu en aurais été très fier de ton fils.

Que soient remerciés toutes celles et tous ceux, qui de près ou de loin, m'ont aidé, par leur travail et leur soutien, à accomplir cette thèse de doctorat.

Table des matières

Liste des abréviations

Analyse en **C**omposante **P**rincipale	**ACP**
Contours **A**ctifs	**CA**
Contours **A**ctifs **H**ybride	**CAH**
Cubique	**Cub**
Gradient **V**ecteur **F**low	**GVF**
Langage **P**arlé **C**omplété	**LPC**
Lecture **L**abiale	**LL**
Minimum de **L**uminance	**Lmin**
Mise en **C**orrespondance	**MC**
Modèles de **D**istribution de **P**oints	**PDM**
Modèles **A**ctifs de **F**orme	**ASM**
Modèles **A**ctifs d'**A**pparence	**AAM**
Points **C**aractéristiques	**P C**
Perception **V**isuelle de la **P**arole	**PVP**
Reconnaissance **A**utomatique de la **p**arole	**RAP**
Teinte, **S**aturation, **L**uminance	**TLS**
Voyelle **C**onsonne **V**oyelle	**VCV**

Liste des figures

Liste des tableaux

Introduction Générale

Dans le domaine très large du traitement de l'image, la vision par ordinateur a pris une part de plus en plus importante au fur et à mesure que les opérations de base et techniques de segmentation d'image se perfectionnaient.

L'analyse d'image ne se limite plus à segmenter une image en régions ou à en détecter les contours, l'ordinateur doit à présent être capable d'exploiter ces informations pour donner un sens aux images, de la même façon que l'homme est capable d'interpréter son environnement grâce à sa vue. Les champs applicatifs sont très diversifiés, nous pouvons citer à titre d'exemples la télésurveillance, la télédétection, la biométrie, la médecine, et la labiométrie, etc.

Parmi toutes les images pouvant servir de sujet d'étude, celles représentant le visage humain sont devenues un centre d'intérêt particulièrement fort du domaine de la vision par ordinateur. Le visage fournit en effet une quantité d'informations que le cerveau humain peut relever et interpréter de façon naturelle ; la plus évidente est l'identité de la personne mais on peut relever également l'état émotionnel, la direction du regard et donc le centre d'attention, etc.

Si les traits du visage sont donc remplis de sens, la zone qui en contient le plus est celle de la bouche car elle contient des informations caractéristiques liées à la parole et donc à la communication entre êtres humains.

La volonté d'obtenir des interfaces établissant des rapports de plus en plus naturels entre l'Humain et la Machine, cette volonté fournit donc aussi un terrain d'application pratique privilégié pour les nombreux algorithmes de segmentation du visage et des lèvres développées au cours des vingt dernières années [1]. De nombreuses méthodes ont été proposées pour résoudre le problème de la segmentation des lèvres.

Les lèvres jouent un rôle important dans la perception visuelle de la parole. La forme et le mouvement des contours labiaux donnent des informations visuelles permettant d'améliorer la compréhension de la parole dans un environnement bruité [2]. Un nombre croissant d'études porte sur le Traitement Automatique de la Parole Visuelle, en particulier sur la lecture labiale qui est l'art permettant « d'entendre » les mouvements des lèvres, ce que chacun dit par perception visuelle des gestes articulatoires du locuteur.

Sur le plan scientifique, les travaux d'analyse labiale remontent aux années 80 à l'Institut de la Communication Parlée (ICP- INPG) dans lesquels un maquillage bleu était utilisé pour capturer le mouvement des lèvres [3]. Depuis, cette activité s'est largement

développée et plusieurs travaux ont été menés au Laboratoire des Images et des Signaux (LIS) comme par exemple ceux de P. Delmas qui a présenté un algorithme de détection des contours des lèvres intégrant une détection préalable des coins et des extrema horizontaux de la bouche [4].

N. Eveno a proposé d'utiliser des courbes paramétriques cubiques pour décrire les contours en imposant les conditions et les limites, aux dérivées des courbes aux niveaux des points saillants [5].

Dans son travail P. Gacon a séparé l'apparence globale en deux quantités : une apparence *statique*, caractéristique d'un locuteur donné, et une apparence *dynamique* qui correspond aux fluctuations induites par le mouvement (et donc plus particulièrement la parole) [6].

C. Bouvier a présenté un algorithme pour l'extraction du contour extérieur de la bouche, où il a utilisé deux types de méthodes, une déterministe et une autre statistique [7].

À l'heure actuelle, malgré le nombre important de méthodes de segmentation des lèvres qui existent, certaines applications nécessitent en effet une grande précision dans la modélisation du contour et une grande robustesse par rapport aux changements de conditions de l'environnement et de locuteurs. De tels algorithmes sont très dépendants de la précision de la segmentation. Comme tout problème de segmentation, l'extraction du contour des lèvres est un problème complexe. Les algorithmes doivent pouvoir gérer des lèvres qui subissent des déformations importantes et très rapides lors de la prononciation d'une phrase [8].

L'information sur la position des lèvres peut permettre d'améliorer la robustesse des systèmes de reconnaissances de parole. En effet, l'apport de l'image aide à la compréhension orale et contient donc des informations sur le contenu sonore. Il convient d'extraire sur une image d'un visage parlant des paramètres pertinents permettant de décrire la parole. Or, ces paramètres sont pour la plupart en rapport avec la position des lèvres.

La lecture labiale, ou lecture sur les lèvres, consiste à décrypter sur les lèvres de l'interlocuteur les mots qu'il prononce [9]. En effet, chaque voyelle ou consonne a une forme caractéristique. La lecture labiale reste la meilleure solution pour rompre l'isolement dû à la surdité, elle permet de continuer à communiquer avec les malentendants. Malgré, ses limites (on ne peut pas lire sur les lèvres dans toutes les situations), c'est un

remarquable et indispensable outil de réinsertion sociale, aussi bien au niveau professionnel que personnel.

L'un des enjeux principaux est d'améliorer les conditions d'interaction entre deux interlocuteurs notamment en termes d'intelligibilité des messages échangés. Or, il est connu que la vision de l'interlocuteur améliore la compréhension du message. Cela est évidemment vrai pour les sourds et les malentendants capables de reconnaître des mots par simple lecture labiale mais pas seulement. Enfin, même pour des messages sans bruits, la compréhension est améliorée dans le cas de messages complexes comme par exemple lors d'une discussion dans une langue étrangère. Cette application nécessite avant tout une grande rapidité d'exécution de la part des algorithmes afin de pouvoir compresser en temps-réel la vidéo. La précision doit rester tout de même convenable afin de refléter au mieux la forme des lèvres.

L'objectif visé dans notre étude est de proposer une méthode permettant de modéliser précisément la zone de la bouche avec la meilleure robustesse possible. Une méthode fiable qui ne nécessite pas de réglage de paramètres et qui réalise une segmentation fidèle des contours externes des lèvres. L'algorithme permettant de suivre le contour des lèvres met en évidence la perception visuelle de la parole pour chaque image de la séquence. Cet algorithme se base sur la segmentation statique, puis dynamique, et aussi nous avons rajouté à notre étude la détermination des paramètres labiaux qui sont des éléments nécessaires pour la lecture labiale.

Dans la segmentation statique, nous optons pour une initialisation semi- automatique afin de rechercher les Points Caractéristiques (PC) des lèvres. Par la suite, nous traçons le contour à l'aide d'une méthode que nous avons proposée, en utilisant la pente optimale [10].

Dans la segmentation dynamique, nous avons implémenté une méthode de mise en correspondance pour le suivi des PC. Un algorithme de réajustement a été ajouté pour segmenter rigoureusement les lèvres durant leurs mouvements [11- 12].

En outre, nous avons inclus à la segmentation statique une segmentation basée sur l'emplacement manuel et rigoureux des PC, afin de comparer les deux méthodes, manuelle et semi-automatique.

Nous avons aussi exploité cette segmentation par l'extraction des paramètres labiaux et le temps d'exécution, afin de pouvoir identifier la parole prononcée par un locuteur dans le cadre de la lecture labiale [13].

Le document présentant notre travail est structuré en quatre chapitres:

- Le premier expose quelques généralités sur la parole et sa perception auditive. Nous avons donné un aperçu sur les modes et lieux d'articulation et la production de la parole. Nous avons aussi présenté le système auditif et sa perception, et enfin un bref rappel sur les surdités.
- Le second est dédié à l'état de l'art où nous présentons la perception visuelle de la parole, l'effet McGurk montrant une interférence entre l'audition et la vision lors de la perception de la parole. Nous avons aussi présenté l'intérêt de la lecture labiale et ses limites. Enfin un état de l'art de l'analyse labiale a été donné présentant les différentes techniques de segmentations des lèvres.
- Le troisième présente les notions fondamentales utilisées dans le traitement d'images, ainsi que les techniques de prétraitement et de segmentation déjà existantes. Nous nous intéressons par la suite à la méthode de segmentation par contours actifs ; nous étudions l'espace couleur, le gradient hybride, et la pseudo-teinte. Nous présentons aussi les différentes formes de snacks utilisées pour la segmentation labiale. Enfin nous exposons les différents résultats de segmentations des lèvres par différentes techniques.

Le dernier est consacré à la segmentation et suivi des mouvements des lèvres, où nous avons proposé des techniques pour la détermination des PC, afin de créer le modèle des lèvres, et par la suite nous appliquons les méthodes de mise en correspondance pour pouvoir faire le suivi des lèvres. Un réajustement des points a été établi pour maintenir le modèle conforme pendant le mouvement. On expose aussi les résultats expérimentaux obtenus durant cette étude, avec des interprétations et des discussions sur les résultats obtenus.

Une conclusion générale clôture ce travail de thèse, et quelques perspectives pour la continuité de ce travail de recherche.

CHAPITRE 1 :
Généralités sur la Parole et sa Perception Auditive

1.1 Introduction

La parole est multi-sensorielle de par sa nature [14]; elle résulte d'un ensemble de gestes articulatoires rendus audibles mais aussi visibles.

La vue est le sens le plus étudié parmi les cinq sens intervenant dans la perception humaine. De même, l'ouïe possède une grande importance dans la perception du monde, prenons simplement comme exemples les rôles primordiaux que ce sens a dans la communication et l'art. Effectivement, il peut paraître moins évident et plus difficile à étudier que la vue, mais ces deux sens sont finalement comparables. Ce ne sont pas des sens de proximité, ils sont tous les deux largement véhiculés par les médias et ils sont complémentaires dans bien des situations. La vue peut par exemple compléter l'ouïe dans un environnement bruyant, où on est amené à lire sur les lèvres pour mieux comprendre le message que l'on nous communique.

La perception visuelle est la fonction mentale impliquée dans la discrimination de la forme, de la taille, de la couleur et d'autre stimulus oculaires.

Ce premier chapitre a pour but d'une part, de présenter les connaissances essentielles qui décrivent les natures physiologiques et phonétiques de la parole, et d'autre part la perception auditive de la parole.

1.2. Production de la parole

La parole est la faculté d'exprimer et de communiquer la pensée au moyen du système des sons du langage articulé émis par les organes phonateurs. Le processus de production de la parole est un mécanisme très complexe qui repose sur une interaction entre le système neurologique et physiologique. Il y a une grande quantité d'organes et de muscles qui entrent dans la production de sons des langues naturelles.

1.2.1. Articulateurs de la parole

Le fonctionnement de l'appareil phonatoire humain repose sur l'interaction entre trois grandes classes d'organes : les poumons, le larynx, et les cavités supra-glottiques.

1.2.1.1. Système sous-glottique

Ce système est composé de l'abdomen, du diaphragme, du thorax, des poumons et de la trachée artère (Figure 1.1). Lors de l'inspiration, la contraction du diaphragme permet

d'élargir la cavité pulmonaire pour faire rentrer de l'air, et l'élévation du diaphragme ainsi que la contraction des muscles intercostaux permet d'expulser cet air dans la trachée, avec une pression quasi constante [15].

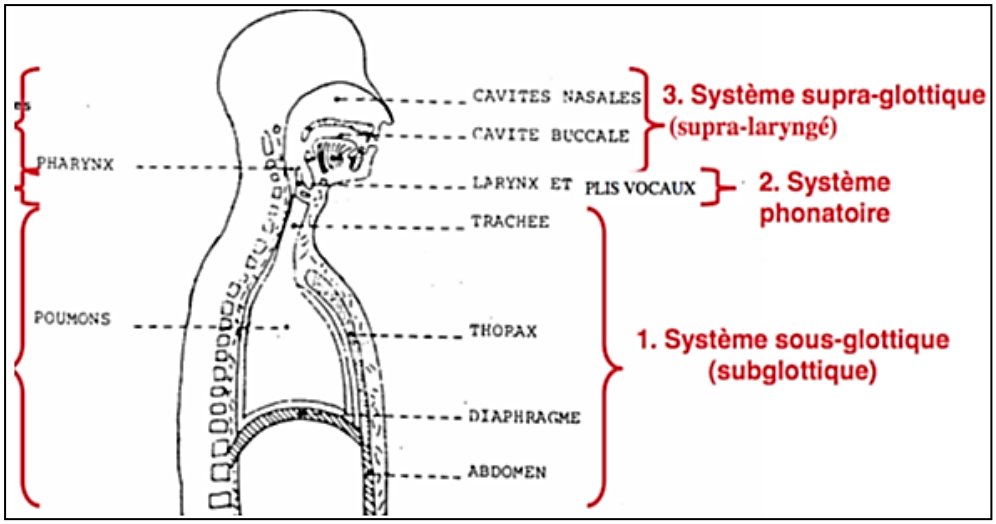


Figure 1.1 : Présentation schématique des principaux organes de la phonation [15]

1.2.1.2. Système phonatoire

Le larynx fait l'intermédiaire entre la trachée et le pharynx. Il abrite les plis vocaux. L'air qui arrive de la trachée traverse le larynx, et donc les plis vocaux (figure 1.2). Le passage d'air situé entre les plis vocaux s'appelle la glotte. Grâce à la rotation de deux cartilages du larynx, les aryténoïdes, la glotte peut être ouverte pour permettre la respiration, et fermée pour permettre la phonation (figure 1.3). Si la pression sous-glottique est différente de la pression supra-glottique (l'air envoyé par la soufflerie augmente la pression sous-glottique), les plis vocaux vont alors entrer en vibration, selon l'effet Bernoulli (effet de rétro-aspiration de la muqueuse cordale) [16]. C'est cette vibration qui crée la voix et ses différentes harmoniques. Selon les propriétés intrinsèques aux plis vocaux et les différences de pression sous et supra-glottique, la fréquence de vibration des plis vocaux sera différente, ce qui caractérisera la voix propre à une personne.

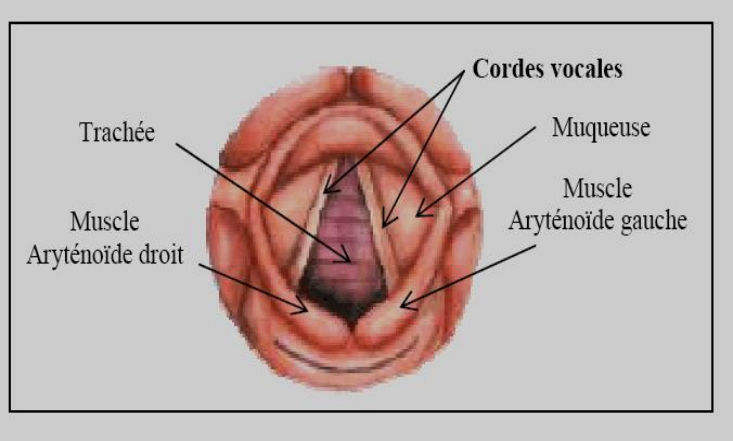


Figure 1.2 : Schéma du larynx

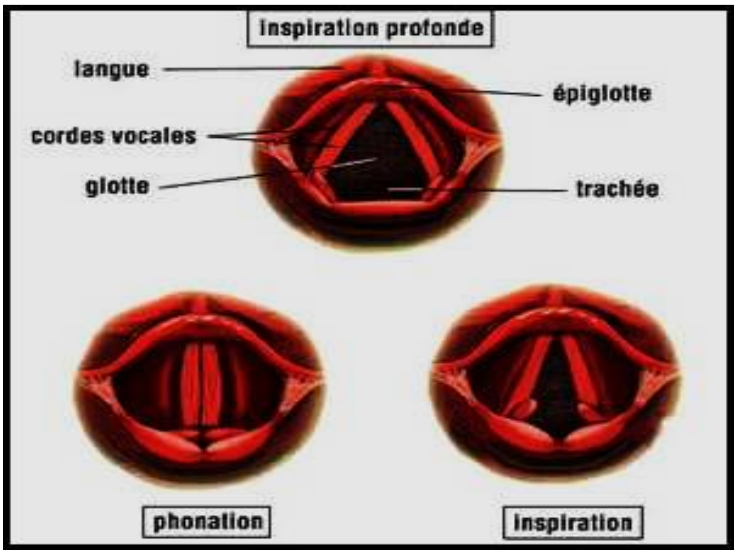


Figure 1.3 : Position des cordes vocales

1.2.1.3. Système supra-glottique

L'air, après son passage dans le système phonatoire, va traverser le conduit vocal. Celui-ci pourra prendre diverses formes d'après la variété de ses constituants. Le système supra-glottique est composé de différents résonateurs qui sont les cavités : labiale, buccale, nasale et pharyngale, ainsi que de différents articulateurs : les lèvres, les dents, les alvéoles, le palais dur, le velum, l'uvule, la langue. Suivant l'endroit où l'air résonne (par exemple cavité orale ou cavité nasale), et suivant la résistance imposée au passage de l'air par les

résonateurs, le son qui arrive du système phonatoire (qui possède déjà quelques propriétés acoustiques), finira d'être transformé en un son particulier, appelé phonème, qui appartiendra au système phonologique de la langue voulue.

1.3. Système phonologique

Chaque langue possède un ensemble de règles de grammaire, et d'autres liées à la formation des mots. Mais ce n'est pas tout; il existe également des règles concernant les sons. Ceux-ci sont organisés dans un ensemble qu'on appelle le *système phonologique* d'une langue.

En linguistique, on distingue les *phones* (aussi appelés plus simplement *sons*) des *phonèmes*. Les phonèmes sont des représentations abstraites des sons établies les unes par rapport aux autres selon leur fonction au sein d'un ensemble organisé (le système phonologique).

La phonétique traditionnelle classe les phonèmes en *voyelles*, *consonnes* et *semi-voyelles* (ou *semi-consonnes*). La distinction entre voyelles et consonnes s'effectue de la manière suivante :

- si le *passage de l'air* se fait *librement* à partir de la glotte, on a affaire à une voyelle ;
- si le *passage de l'air* à partir de la glotte est *obstrué*, complètement ou partiellement, en un ou plusieurs endroits, on a affaire à une consonne ;
- les *semi-voyelles* présentent la même articulation que les voyelles, mais se comportent dans la syllabe comme les consonnes : plus précisément, les consonnes et les semi-voyelles ne peuvent constituer à elles seules une syllabe, les voyelles si : par exemple, le mot *abbaye* [a /be / i] comporte des voyelles alors que le mot *abeille* [a / bɛj] comporte ausi une semi-voyelle notée [j].

Le système phonologique du Français possède 36 phonèmes : 17 consonnes, 16 voyelles et 3 semi-voyelles. Cependant, selon le contexte et différents facteurs comme l'origine, le sexe et l'âge du locuteur, chacun de ces phonèmes peut être prononcé de différentes manières. Ces variantes sont les phones, c'est-à-dire les réalisations concrètes de ces phonèmes.

1.3.1. Articulation des consonnes

Les *consonnes* se différencient des voyelles par la présence d'un *obstacle* qui empêche le libre écoulement de l'air. La qualité de cet obstacle, ou *mode d'articulation* (occlusif ou constrictif), est le critère principal qui permet de les distinguer entre elles. Le second critère de classification est la position de cet obstacle, ou *lieu d'articulation* (des lèvres au palais mou) et l'adjonction éventuelle d'antirésonances (Tableau 1.1).

Les sons produits sont dits voisés ou sonores. On y trouve les voyelles, les semi-consonnes et certaines consonnes [b, d, g, v, z, r, l, m, n,..]. Si la glotte est resserrée mais non fermée, l'air peut circuler sans mettre en action les cordes vocales, les sons produits alors sont dits non voisés ou sourds [p, t, k, f, s,..].

Tableau 1.1 : Tableau des consonnes du Français [17].

LIEU D'ARTICULATION / MODE D'ARTICULATION			Bilabiales	Labiodentales	LabioPalatales	Labio vélaires	Alvéodentales	Alvéolaires	Post-alvéolaires	Palatales	Vélaires	Glottal		
Occlusives	Orales	Sourdes	(p)				t/t^h		ʈ	c	k	ʔ	Explosives	Momentanées
		Sonores	b				d							
	Nasales	Sonores	m				n				ɳ	ŋ		
Constric-tives	Orales	Sourdes		f				s	ʂ		x	h	Fricatives	Continues
	sonores	Médianes		v				z	ʐ		ɣ			
		Vibrante						r					Liquides	
		Latérale						l						

1.3.1.1 Mode d'articulation

Le mode d'articulation fait référence à d'éventuelles entraves au passage de l'air dues aux organes articulatoires. Les consonnes qui, à un moment de leur articulation, donnent lieu à un blocage du passage de l'air sont nommées les occlusives, elles sont des sons bruités de courte durée, caractérisés par un silence provenant de la fermeture complète du conduit vocal en un point précis. Dans le cas contraire, on parlera de consonnes constrictives ou fricatives. Ces consonnes sont des sons bruités produits par l'écoulement turbulent de l'air, lorsque cet écoulement rencontre un rétrécissement, un lieu de constriction, il se produit un bruit de friction.

On dénombre 10 consonnes occlusives [p, t, k, b, d, g, m, n, …] et 8 consonnes constrictives (6 médianes) [f, s, v, z,..] ; une latérale [l] et une vibrante [R].

1.3.1.2 Lieu d'articulation

Le lieu d'articulation d'un son correspond à l'endroit le plus étroit de la cavité pharyngo-buccale lors de la production de ce son. Pour les consonnes occlusives, il s'agira du point d'occlusion du canal vocal. Le lieu d'articulation est illustré dans le tableau 1.2 :

Tableau 1.2 classement articulatoire des consonnes

	consonnes	**Lieu d'articulation**
Bilabiale	[p], [b], [m]	les lèvres forment un contact entre elles
Apico-dentales	[t], [d], [n]	la pointe de la langue se déplace vers les alvéoles, région située directement après les dents, en touchant les dents
Dorso-vélaires	[k], [g],[ŋ]	le dos de la langue se déplace vers ou contre le voile du palais
Labio-dentales	[f], [v]	la lèvre inférieure forme un contact avec les dents
Prédorso-alvéolaire	[s],[z]	le dos de la langue se déplace vers les alvéoles
Prédorso-prépalatales-labiales	[ʃ], [ʒ]	la pointe de la langue se place à l'arrière de la région des alvéoles,
Apico-alvéolaire	[l]	lorsque la pointe de la langue se déplace vers les alvéoles
Dorso-uvulaire	[R]	lorsque la luette vibre ou se déplace
labio-palatal	[j]	la langue se déplace vers le palais, les lèvres étant rapprochées
labio-vélaire	[w]	la langue se déplace vers le voile du palais, les lèvres étant arrondies

1.3.1.3 Antirésonance

L'air peut passer ou non par la cavité nasale. Lorsque le voile du palais se trouve en position abaissée, une partie de l'air peut passer par le nez. Ces consonnes sont appelées

nasales [m, n, ŋ]. Lorsque le voile du palais est relevé, l'air ne peut passer et les consonnes sont dites orales.

1.3.1. Articulation des voyelles

La classification des voyelles s'organise autour de deux critères : le mode d'articulation et le lieu d'articulation. Contrairement aux consonnes qui se différencient selon le degré d'ouverture du canal buccal (les consonnes occlusives sont fermées et les consonnes constrictives sont partiellement fermées) ; les voyelles se distinguent néanmoins par le volume du résonateur buccal qui est fonction du degré d'aperture dû à l'écartement des mâchoires. On note habituellement quatre degrés d'aperture : fermé, mi- fermé, mi- ouvert et ouvert. La configuration des lèvres est également une variable pertinente. Les lèvres peuvent être arrondies ou plus ou moins écartées. Pour les voyelles arrondies [y, u], les lèvres sont projetées vers l'avant. Les voyelles écartées ou étirées [i, e, a] sont quant à elles caractérisées par un accolement des lèvres contre les dents. Ensuite, comme les consonnes, les voyelles peuvent être nasales ou orales selon que la position du voile du palais fasse ou non intervenir l'anti résonateur. Enfin, le lieu d'articulation permet de différencier les voyelles antérieures [i, y, e], articulées en avant de la cavité buccale, des voyelles postérieures [u, o].

A partir de la figure 1.4, on peut identifier 7 caractéristiques fondamentales : les voyelles aigues, graves, neutres, ouvertes, fermées, orales ou bucco-nasales.

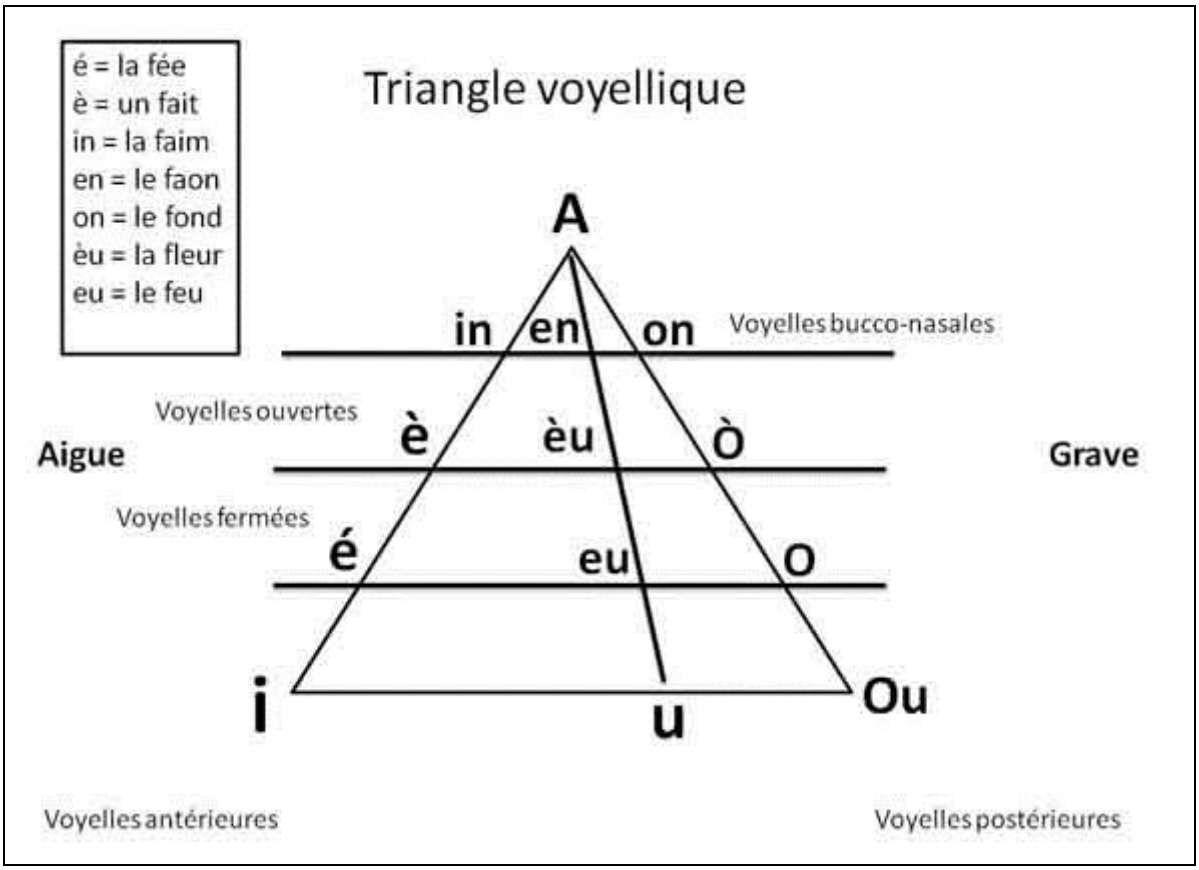


Figure 1.4 : Triangle vocalique pour les voyelles du Français [18].

A l'instar des caractères pour le langage, les phonèmes permettent de décomposer et d'écrire toute parole. Dans le domaine visuel, ces éléments simples sont appelés visèmes. Ces visèmes sont définis à partir de groupe de phonèmes (figure 1.5).

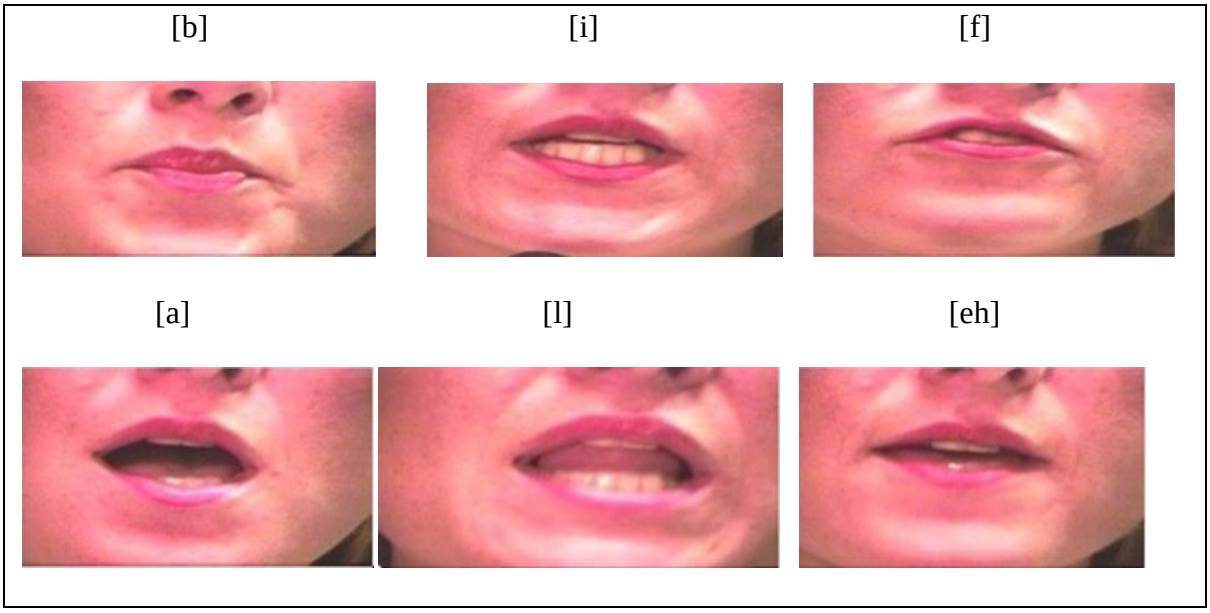


Figure 1.5 **:** Exemples de visèmes [19]

1.4. Perception auditive

Le décodage de la parole nécessite sa réception sensorielle, sa transduction et son intégration sous forme d'informations neuronales par le système auditif puis leurs transferts vers les centres nerveux et corticaux par les voies auditives. La première étape consiste en une transmission et adaptation de l'énergie sonore à l'oreille interne au travers des structures suivantes : pavillon auriculaire, conduit auditif externe, chaîne tympano-ossiculaire. Sur le plan fonctionnel, nous ne retiendrons que la notion d'une zone de résonance autour de fréquence de 2 kHz induite par les caractéristiques physiques de ces structures et jouant un rôle important dans l'intelligibilité de la parole [20].

1.4.1. Système auditif

Le système auditif est composé de deux parties (figure 1.6) : le système périphérique (oreille externe, oreille moyenne, oreille interne et le nerf auditif) et le système central (voies auditives au niveau du tronc cérébral et du cortex auditif).

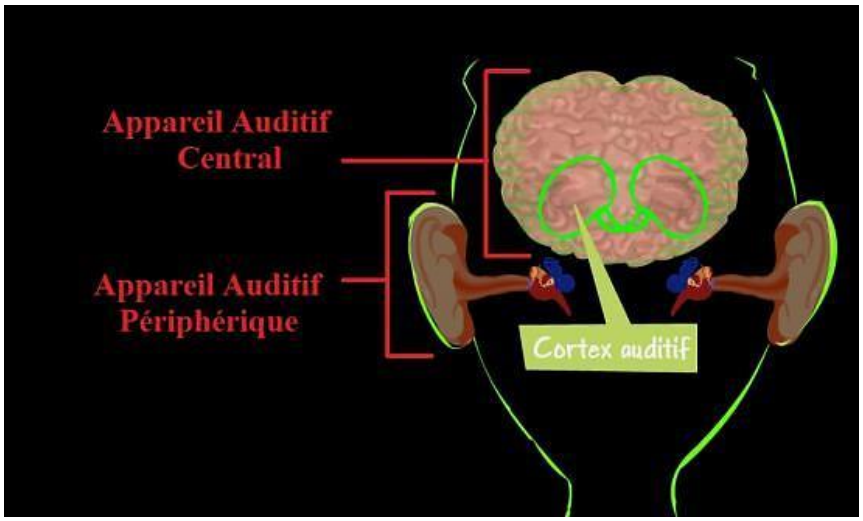


Figure1.6: Système auditif humain

1.4.1.1. Système auditif périphérique

Le système auditif périphérique permet la transmission du signal sonore du pavillon de l'oreille externe jusqu'aux premiers neurones du nerf auditif (Figure 1.7). Chaque partie qui le compose participe à cette transmission. L'oreille comprend trois parties [21] :

- **l'oreille externe** comprend le pavillon et le conduit auditif. Le pavillon collecte les sons et le conduit auditif les guide jusqu'au tympan qui est mis en vibration. Le conduit auditif sécrète une substance cireuse, le cérumen, dont l'excès peut former un bouchon produisant une baisse temporaire de l'audition.
- **l'oreille moyenne** comprend une chaîne de 3 osselets minuscules : le marteau, l'enclume et l'étrier. Cette chaîne transmet les mouvements du tympan à l'oreille interne.
- **l'oreille interne** se compose du vestibule et de la cochlée. Le vestibule est responsable de l'équilibre et la cochlée est responsable de l'audition. La cochlée est une structure en spirale à l'intérieur de laquelle on retrouve l'organe de Corti. Celui-ci contient des cellules ciliées, qui se baignent dans un liquide appelé périlymphe. Lorsque les osselets de l'oreille moyenne transmettent les vibrations à ce liquide, les fibres ciliées se mettent en mouvement. Ceci active un influx nerveux qui sera ensuite transmis au cerveau par le nerf auditif

L'oreille est en permanence ouverte sur l'environnement, et ne se focalise sur un son que lorsque l'évènement présente un intérêt et constitue, au sens large, une «alerte». La vigilance constante du système auditif fait que l'audition est un dispositif d'alerte très efficace à courte distance.

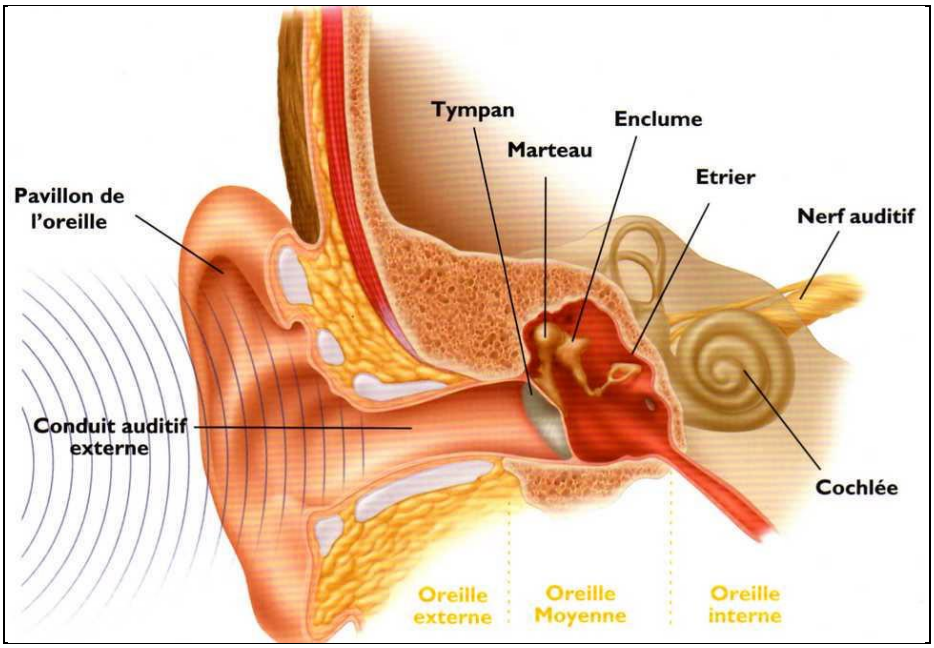


Figure 1.7: système auditif périphérique

1.4.1.2. Système auditif central

Le système auditif central permet la transmission de l'information sonore des premiers neurones du nerf auditif jusqu'au cerveau. C'est ce système qui est responsable de l'interprétation de l'information sonore.

Le système auditif périphérique communique avec le système auditif central à l'aide de fibres nerveuses afférentes, qui partent de l'organe de Corti vers le cortex auditif, et de fibres nerveuses efférentes, qui font le chemin contraire.

Les cellules dans l'organe de Corti se séparent en deux catégories, les cellules ciliées externes et les cellules ciliées internes. Ces deux types de cellules sont reliés à des fibres nerveuses et c'est ce qui forme les deux nerfs auditifs (8ème paire crânienne). L'information au niveau du nerf auditif est ensuite envoyée au cerveau en passant par plusieurs relais au niveau du tronc cérébral (noyaux cochléaires, complexe de l'olive supérieure, lemniscus latéral, colliculus inférieur et corps genouillé médian). Ces différents relais permettent un traitement additionnel du message auditif (figure 1.8).

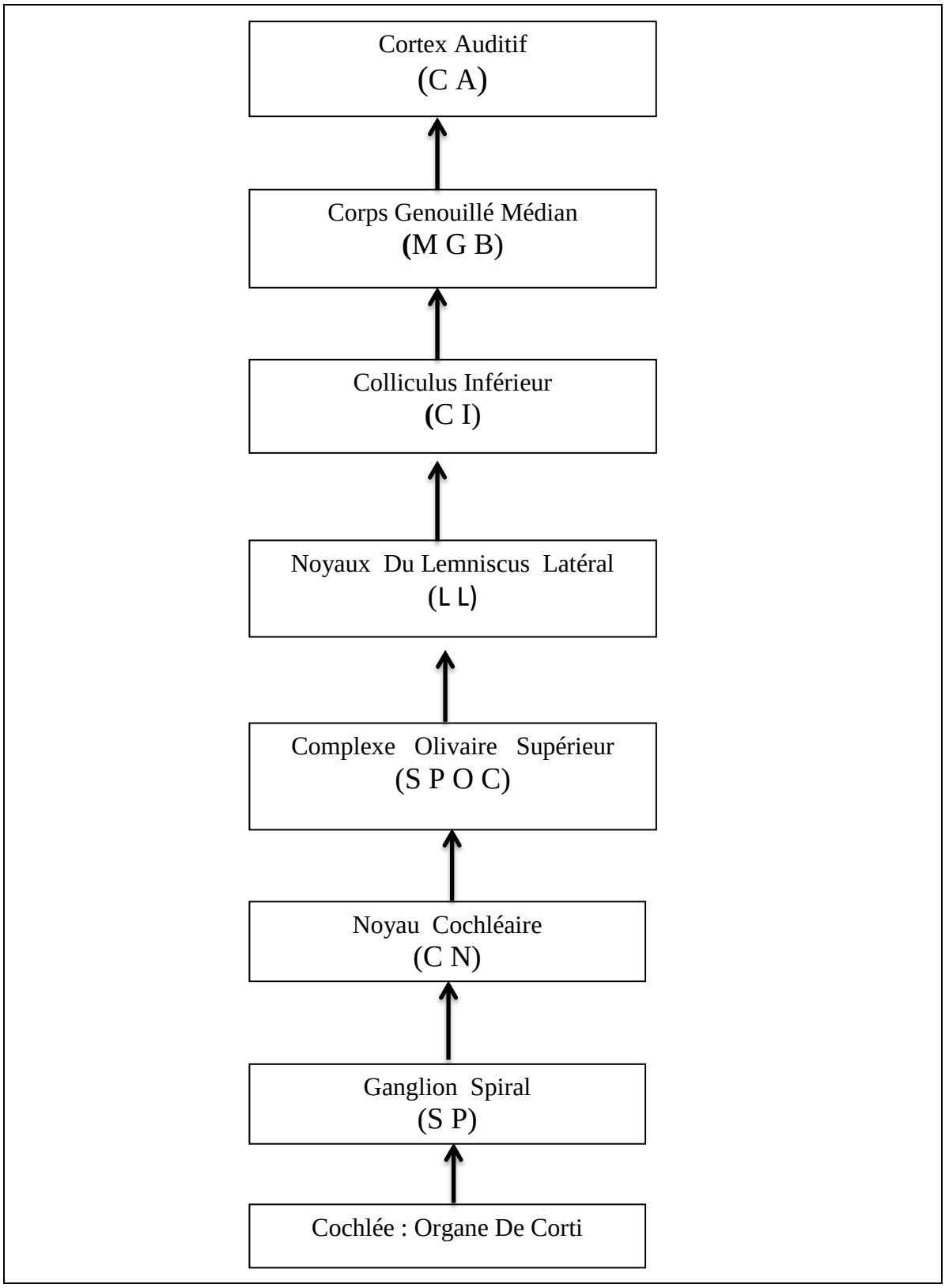


Figure 1.8 : Hiérarchie ascendante des principaux niveaux du système nerveux auditif [22].

Par analogie à l'être humain nous présentons une vue unilatérale schématique du système auditif du chat, donnant une idée grossière des positions anatomiques relatives des différentes étapes du système nerveux auditif d'après Ehret [23]. Dans la figure 1.9, le cervelet et la partie caudale du cortex gauche ont été retirés pour montrer le tronc cérébral. L'abréviations utilisée dans la figure ci-dessous est : AC cortex auditif, MGB corps genouillé médian, IC colliculus inférieur, SC colliculus supérieur, LL lemniscus latéral, SPOC complexe olivaire supérieur, DCN noyau cochléaire dorsal, AVCN noyau cochléaire ventro-antérieur, PVCN noyau cochléaire postéro ventral.

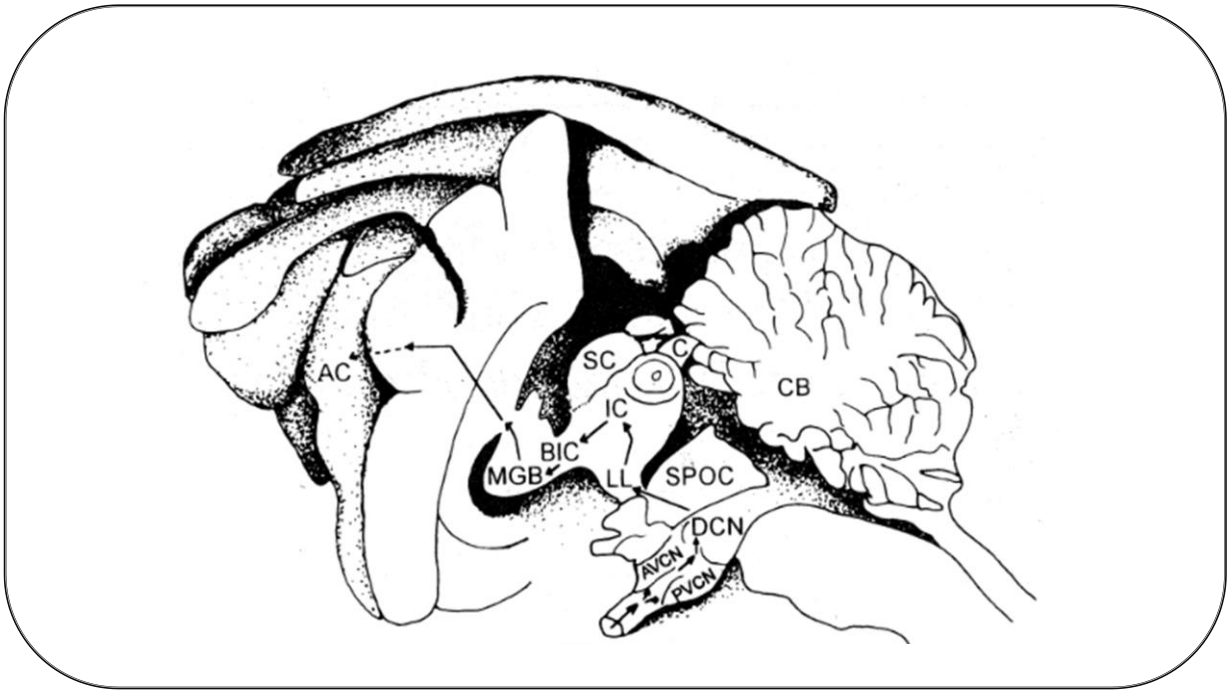


Figure 1.9 : Vue unilatérale schématique du système auditif du chat [22],

En plus de transmettre l'information sonore, les différentes voies auditives fournissent l'information relative à la fréquence du son, à l'intensité et à la position de la source sonore dans l'espace.

1.4.2. Déficience auditive

La surdité est la conséquence d'une atteinte pathologique de la fonction auditive. L'étude de la déficience sensorielle et de ses conséquences cognitives n'épuise pas la question de la surdité, qui se prolonge dans de nombreux domaines n'ayant plus qu'un rapport éloigné avec la physiologie auditive et la réhabilitation médicale [20]. Cependant, la connaissance approfondie des processus auditifs est un nécessaire préalable pour le psychologue afin de comprendre pourquoi les enfants sourds ont tant de difficulté à acquérir le langage verbal.

On distingue 3 grands types de surdité : la surdité de perception, de transmission ou mixte.

1.4.2.1. Surdité de perception

C'est le type de surdité le plus fréquemment dépisté. Elle est due à un problème au niveau de l'oreille interne qui n'assure pas ou plus correctement la transmission des impulsions neuro-électriques jusqu'au cerveau. C'est pourquoi elle est aussi connue sous le nom de surdité neurosensorielle. Elle a pour conséquences, une tendance à parler fort; des difficultés à entendre une voix forte ou au contraire un chuchotement, ou la présence d'acouphènes. Ce type de surdité peut apparaître dès la naissance ou intervenir plus tardivement et toucher les deux oreilles ou une seule.

La lésion de l'oreille interne peut être le résultat du vieillissement, d'une infection virale ou bactérienne comme la varicelle et la grippe, de l'hérédité, d'une médication, des tumeurs, de la rétention de liquides et des conséquences d'un traumatisme crânien, une exposition répétée aux bruits.

La presbyacousie est une des causes fréquentes de surdité de perception. Elle est due au vieillissement naturel du système auditif qui entraîne une perte d'audition progressive. Elle touche en majorité les personnes âgées de plus de 50 ans. Ce type de surdité peut donc engendrer une gêne sociale voire l'isolement de la personne atteinte par cette baisse de l'audition.

La surdité de perception ne peut pour l'instant pas être guérie. L'audition peut cependant être améliorée dans une grande partie des cas, particulièrement en cas de presbyacousie,

des prothèses auditives peuvent être ainsi une solution efficace :

- l'intra-auriculaire : oreillette qui se loge dans le conduit auditif (figure 1. 10. a);
- le contour d'oreille : placé autour de l'oreille avec un embout qui conduit le son dans l'oreille (figure 1. 10. b);
- l'implant cochléaire : sorte d'oreille interne artificielle réservée aux surdités de perception sévère (figure 1. 10. c).

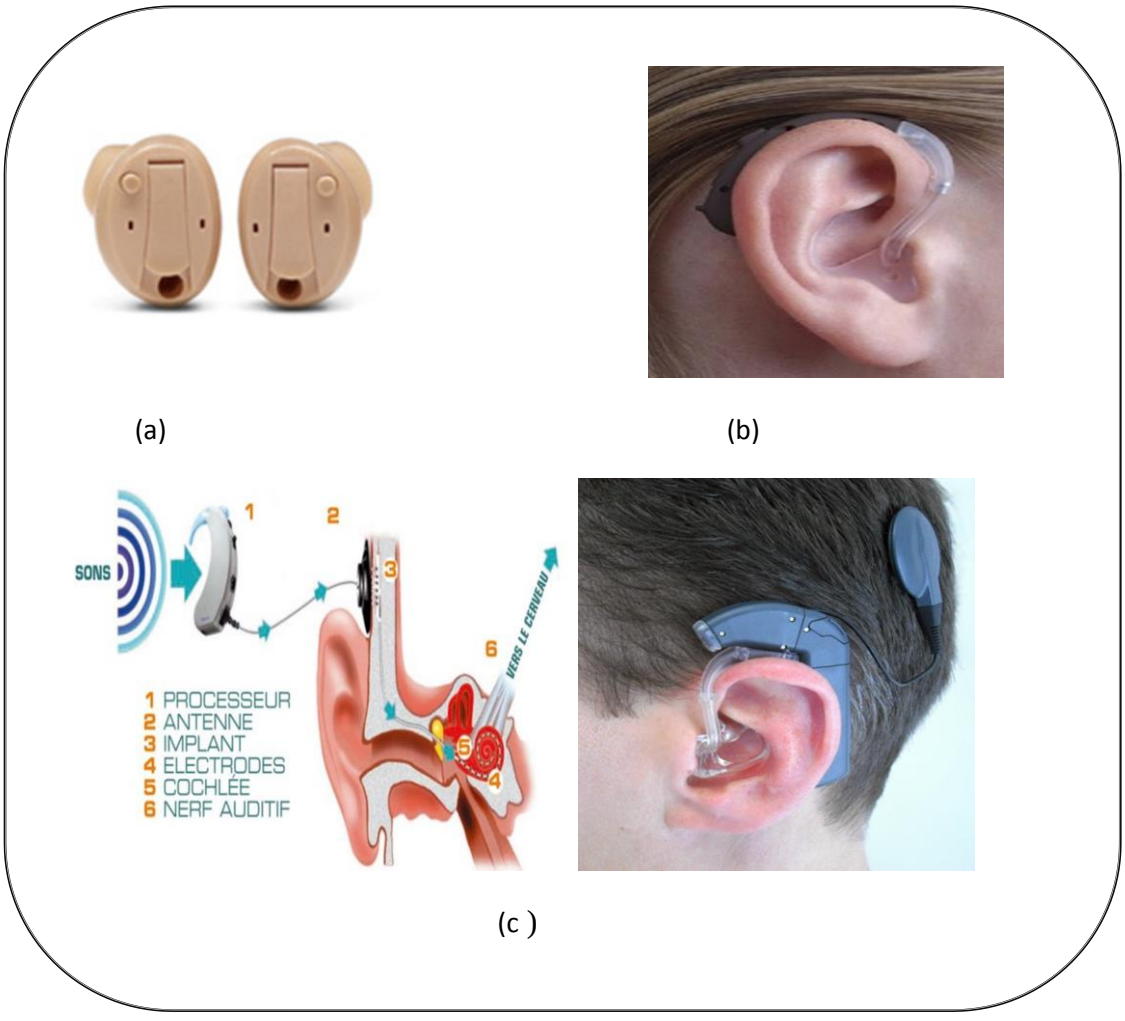


Figure1.10 : Prothèses auditives pour la surdité de perception

a- L'appareil auditif intra-auriculaire, b- contour d'oreille et c- L'implant cochléaire

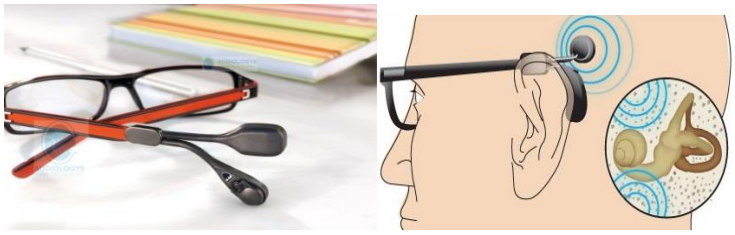
Quel que soit l'appareil adéquat, il doit être accompagné d'une rééducation auditive afin de garantir au patient une amélioration notable et confortable de l'audition.

1.4.2.2. Surdité de transmission

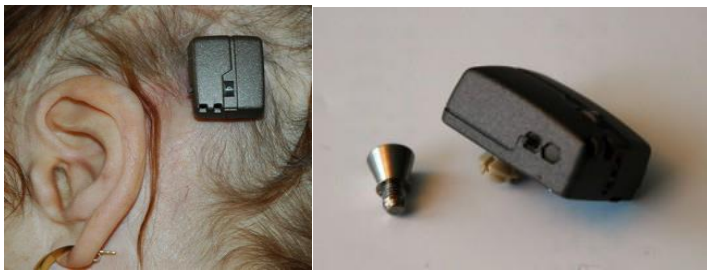
La surdité de transmission est due à un problème de fonctionnement de l'oreille externe ou moyenne qui n'achemine plus les sons jusqu'à l'oreille interne. Elle peut être causée par un bouchon de cire (cérumen), un tympan perforé, l'hérédité, une anomalie congénitale ou une infection d'oreille. Les conséquences de cette surdité sont souvent une tendance au chuchotement, car la personne a l'impression de parler fort, ou des difficultés à entendre les fréquences graves. La surdité de transmission peut généralement être traitée par médicaments, par chirurgie ou par un appareil auditif (figure 1.11), tel :

- qu'un appareil auditif classique (contour d'oreille par exemple) ;
- qu'une prothèse à conduction osseuse : vibrateur monté sur une branche de lunettes ou un contour d'oreille et appliqué sur l'arrière de l'oreille afin de créer un signal vibratoire qui se propage via les os du crâne directement à l'oreille interne (figure 1.11.a) ;
- qu'un implant à ancrage osseux : fonctionne sur le même principe que la prothèse à conduction osseuse mais cette fois le vibrateur est implanté directement dans l'os derrière l'oreille (figure 1.11.b).

Une combinaison de déficience auditive de transmission et perception (neurosensorielle) est appelée déficience auditive mixte. La déficience auditive mixte survient lors d'une lésion nerveuse dans le cerveau ou dans les voies cérébrales.



(a)



(b)

Figure 1.11 : Prothèses auditives pour la surdité de transmission :
(a) une prothèse à conduction osseuse, (b) Implant à ancrage osseux.

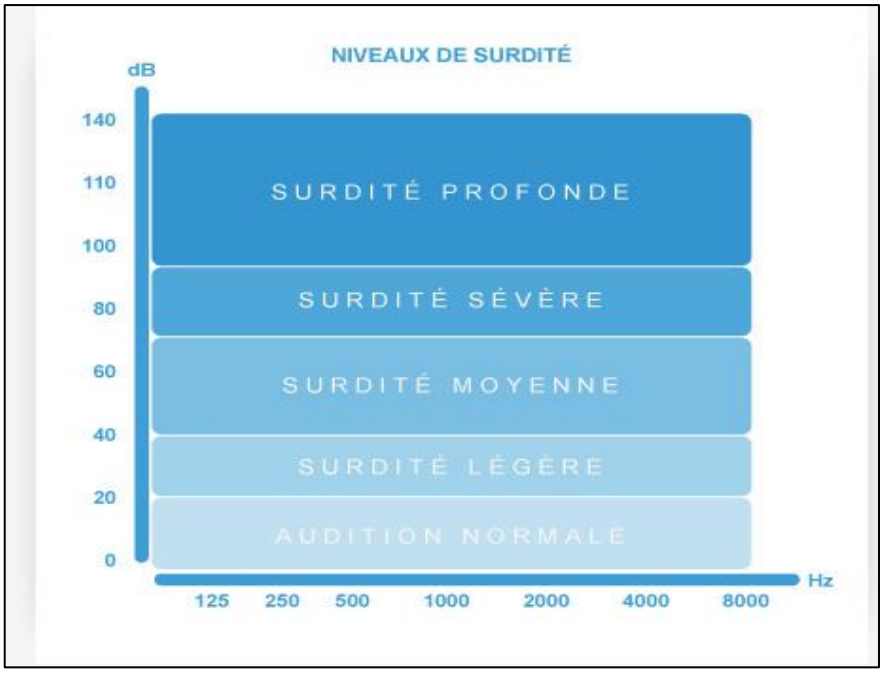
1.5. Degrés de surdité

Pour mesurer l'importance du handicap provoqué par une perte auditive, les surdités sont classées en 5 degrés [24] (Tableau 1.3) :

- légère : de 21 à 40 dB**,** la parole est perçue à voix normale, difficilement à voix basse, comme tous les sons faibles ou lointains ;
- moyenne : de 41 à 70 dB, de 1er degré (41 à 50 dB), la parole est entendue si on élève la voix, mais mal comprise. De 2ème degré (51 à 70 dB), la personne comprend mieux si elle regarde son interlocuteur **(**perception visuelle de la parole**)**. Quelques bruits familiers sont encore perçus ;
- sévère : de 70 à 90 dB, le handicap est important. Seuls les bruits forts et les voix proches sont perçus ;
- profonde : supérieur à 90 dB, à ce stade, et jusqu'à 120 dB de perte, la parole n'est plus du tout perçue et seuls les bruits très puissants sont entendus sans être nécessairement identifiés ;

- totale : pas d'audition mesurable**.**

Tableau 1.3 : Niveau de surdité [24]



1.6. Conseils pour communiquer avec une personne malentendante

- parlez clairement. Parlez sur un ton de conversation normal et articulez clairement sans exagération ;
- rythmez votre élocution et faites des pauses normalement. Il est difficile de suivre une élocution rapide ou lente ;
- ne criez pas, ne chuchotez pas, n'exagérez pas vos mouvements de la bouche. Cela empêche de bien comprendre ce qui est dit. En réalité, le fait de crier déforme le son et est dérangeant. Personne n'aime qu'on lui crie après ;
- reformulez ce qui a été dit. Cela peut donner d'avantage d'indices ou de meilleurs indices de ce qui a été dit ;
- épelez les noms et utilisez un nom pour identifier une lettre. Par exemple B comme Bernard et P comme Pierre ;
- utilisez le langage corporel, votre expression donne de nombreux indices sur la teneur du message excitation, colère, ennui, etc ;

- établissez le contact visuel ou effectuez un toucher léger pour attirer l'attention avant de parler ;
- si possible, tenez une conversation dans un environnement calme. Le bruit de fond est très difficile à filtrer.

Enfin on peut dire que la vision et l'audition seraient liées, car les chercheurs pensent que la lecture sur les lèvres intervient étroitement dans la compréhension de la conversation chez ces patients touchés par une surdité après l'acquisition du langage, ce qui expliquerait cette suractivité des aires visuelles et auditives liées au langage. Au fur et à mesure de la récupération auditive, une forte synergie entre ces deux sens s'établit et la vision permet un meilleur apprentissage des informations auditives grossières fournies par l'implant. Leurs travaux ouvrent de nouvelles pistes pour la rééducation de ces patients, qui pourraient grandement améliorer leur prise en charge [25].

1.7. Perception de la parole

La Perception de la parole est le processus par lequel les humains sont capables d'interpréter et de comprendre les sons utilisés dans le langage. L'étude de la perception de la parole est reliée aux champs de la phonétique, de psychologie cognitive et de perception en psychologie.

Plusieurs travaux ont montré que la perception de la parole, et plus particulièrement le décodage des voyelles et des consonnes par l'auditeur, tire profit non seulement des caractéristiques de l'onde sonore produite par le locuteur, mais également des informations visuelles transmises par la position de sa mâchoire et de ses lèvres. On rencontre pourtant fréquemment des situations impliquant des interactions de communication uni modale réussies entre interlocuteurs ; par exemple, lorsqu'un individu parle au téléphone à un autre individu, seule la modalité auditive est utilisée, et lorsque deux personnes échangent des paroles alors qu'une vitre les sépare et les empêche de s'entendre, seule la modalité visuelle est utilisée. Des privations sensorielles telles que la surdité et la cécité constituent des contraintes physiologiques qui suppriment une des deux modalités de la parole audio-visuelle.

1.8. Conclusion

Comme nous avons vu dans ce chapitre, la perception de la parole est caractérisée par des éléments de modalités sensorielles (et en particulier auditives et visuelles) qui peuvent donner naissance à un percept unitaire, c'est que les informations traitées par les différents systèmes sensoriels doivent interagir quelque part.

Apparemment ces systèmes indépendants coopèrent pour offrir une perception cohérente et unifiée du monde environnant. Cette coopération est concrétisée avec une influence du système visuel sur le traitement spatial auditif.

CHAPITRE 2 :
Perception Visuelle de la Parole et Lecture Labiale

2.1. Introduction

Pour percevoir la parole, le cerveau humain utilise les informations sensorielles provenant non seulement de la modalité auditive mais également de la modalité visuelle. En effet, de précédentes recherches ont mis en évidence l'importance de la lecture labiale dans la perception de la parole, en montrant sa capacité à améliorer et à modifier celle-ci. C'est ce que l'on appelle l'intégration audio-visuelle de la parole.

Dans ce chapitre, nous allons introduire un état de l'art de la perception visuelle de la parole, et la lecture labiale. Ce chapitre s'articule autour de l'analyse labiale en présentant les principales techniques de segmentation des lèvres.

2.2 Vision

La vision est un processus qui, à partir d'images d'un environnement extérieur à l'observateur, produit une description utile et dépouillée d'informations superflues [26].
La vision est considérée comme le sens le plus développé chez l'Homme et représente le système sensoriel le plus étudié. Les données électrophysiologiques et lésionnelles chez l'animal et chez l'Homme ont permis de mettre en évidence l'existence de plusieurs voies parallèles de traitement de l'information visuelle, commençant dès la rétine et mettant en jeu des structures sous-corticales (corps genouillés externes) et un grand nombre d'aires corticales organisées hiérarchiquement (cortex visuels primaires) (figure 2.1).

Des expériences très simples dans lesquelles des sujets sont placées dans un environnement visuel uniforme révèlent des comportements intéressants. En l'absence de tout point de repère. L'observateur ne se contente pas de fixer un point indéterminé, mais au contraire parcourt inlassablement le champ visuel à la recherche d'un point de référence. Le système visuel aurait donc besoin de la présence d'objets dans le champ visuel pour fonctionner. Ce qui amène à la remarque suivante, la Vision est un processus de reconnaissance, elle est :

- associative c'est - à - dire association de vues ou de propriétés avec des concepts et des représentations ;
- interprétative car elle recherche à répondre à des questions spécifiques à propos de l'environnement ;
- dirigée parce que chaque comportement oriente vers un certain type de calculs ;

- sélective car les informations inappropriées pour la tâche en cours sont rejetées [27].

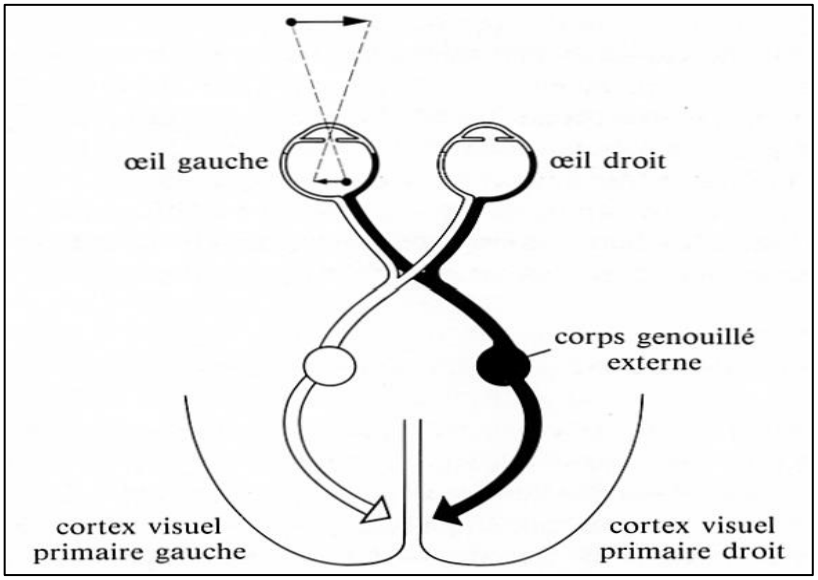


Figure 2.1 : Schéma d'un système visuel [28]

2.3. Théories de la perception visuelle

L'étude du problème de la vision a donné lieu à de nombreuses approches, séparées pour des raisons historiques et pour différentes interprétations du fonctionnement de la vision naturelle, qu'elle soit humaine ou bien animale [29]. Le domaine de la vision concerne des chercheurs aussi variés que des psychologues, des biologistes, des neurobiologistes, des ingénieurs, des informaticiens ou mathématiciens, ainsi nous pouvons dégager trois familles d'approches [30] :

- psycho-visuelle, la plus ancienne car attachée aux aspects psychologiques de la perception visuelle ;
- analytique, qui cherche à comprendre comment fonctionnent les mécanismes sensoriels et neuronaux de la vision à un niveau biologique. C'est le cas, en particulier, des théories neurophysiologiques ;
- calculatoire, qui traite des problèmes algorithmiques de l'acquisition, du traitement et de l'interprétation des informations visuelles. Il s'agit des théories engendrées par la vision par ordinateur.

Les différentes approches de la Vision par Ordinateur doivent beaucoup aux théories développées depuis plus d'un siècle par les neurophysiologistes et les psychologues. Elles subissent également les influences millénaires de courants de pensées scientifiques et philosophiques. Donner une vue d'ensemble de ces différentes théories est donc nécessaire afin de replacer la vision par ordinateur dans son contexte.

2.4. Perception visuelle de la parole

De nombreuses recherches ont été menées pour tenter de comprendre comment la parole est reconnue. Dans la mesure où les paramètres acoustiques de chaque phonème sont changés (en fonction du contexte, des phénomènes de coarticulation), le signal acoustique ne peut être à lui seul l'objet de la perception de la parole. Prenons par exemple la consonne [d] selon la nature du contexte vocalique, le son émis par la consonne peut avoir des profils acoustiques très différents. Malgré ces différences acoustiques, tout le monde est capable de reconnaître qu'il s'agit de la consonne [d]. Le seul point commun à toutes ces productions, c'est l'image articulatoire : la langue est toujours dans la même position, derrière les incisives. C'est le geste articulatoire qui forme l'unité de base à la perception de la parole.

Dans un article récent, Devlin recense les études faisant intervenir la neuro-imagerie pour valider la théorie motrice de la perception de la parole, c'est-à-dire que la production et la perception de la parole sont intimement liées. Il en résulte que la perception passive de la parole implique les aires cérébrales dévolues à la production de la parole. Percevoir la parole, c'est alors percevoir les mouvements articulatoires qui la produisent. La théorie motrice de la perception de la parole permet alors d'entraîner le labiolecteur à perfectionner ses praxies bucco-faciales et ses schémas articulatoires afin de mieux percevoir la parole [31].

2.5. Perception audio-visuelle de la parole

La perception audio-visuelle de la parole est plus que de la lecture labiale. En effet, nous utilisons les informations visuelles à notre disposition pour écouter quelqu'un : le contexte, les gestes, les mimiques, la posture, les changements dans le regard liés à l'articulation. La vision des mouvements articulatoires correspond en une translation en codes phonologiques d'informations purement optiques [32].

Au niveau du visage, la lecture se porte sur la région des yeux, de la bouche, de la région nasale surtout du côté droit. Le pourcentage de temps passé à regarder l'une ou l'autre région dépend des sujets et des tâches. Lorsqu'il s'agit d'un monologue, le regard portera surtout sur les yeux. Cela sera également le cas lors de l'analyse de l'intonation, de la prosodie. Par contre, les yeux seront moins regardés lors des tâches de lecture labiale de mots. Dans des conditions d'écoute plus difficile, la bouche sera plus investie.

2.6. Lieu d'intégration audio-visuelle

L'intégration audio-visuelle se produit déjà à un niveau bas du système nerveux : dans le colliculus supérieur. Dans cette structure, des neurones répondent à différentes modalités sensorielles avec pour certains une réponse de type intégrative. Ces neurones multimodaux sont distribués dans plusieurs réseaux, dans le cerveau, entres autre dans la région pariétale, autour du sulcus temporal supérieur et dans les loges frontaux. Samson a décrit une aire auditive spécifique pour le traitement du langage activé lors d'une analyse acoustique de mots, en modalité auditive ou visuelle. Il s'agit d'une région multisensorielle d'intégration intervenant après le cortex auditif primaire. Cette zone s'active en présence de langage mais ne s'active pas en présence de bruit. Elle se retrouve de part et d'autres du gyrus temporal supérieur, dans le sulcus temporal supérieur, bilatéralement [33].

Le cortex auditif sensible pour le langage semble correspondre à une structure de décodage de la parole. Elle élaborerait des représentations neuronales d'objets sonores qui sont spécifiques à la voix et au langage. Ces objets sonores intègrent dans cette structure multimodale des informations visuelles liées à l'analyse des mouvements de lecture labiale. Il s'agirait d'une étape indispensable pour mettre en route les réseaux neuronaux du traitement ultérieur du langage dans l'hémisphère gauche. L'expérience auditive précoce est nécessaire au développement d'un réseau bien structuré, bien cohérent. Giraud a démontré, chez les adultes sourds porteurs d'un implant cochléaire et chez les normo-entendants, que l'étude de la parole active le cortex visuel et non pas uniquement le cortex auditif. Il y a d'emblée activation audio-visuelle. L'activité des aires auditives et visuelles est modulée par l'interaction des stimulations auditives et visuelles [34].

L'intégration audio-visuelle favorise les informations auditives ou visuelles en fonction du type d'analyse réalisée. Lorsque la non-congruence porte sur les informations spatiales, les informations visuelles vont dominer. Lorsqu'elle porte sur les informations

fréquentielles, les informations auditives vont dominer. Il s'agit d'une réelle intégration et non pas une facilitation avec modification des activités cérébrales modulées. L'intégration audio-visuelle du langage concerne les entendants comme les sourds.

•

2.7. Effet McGurk

L'effet McGurk est un phénomène perceptif qui montre une interférence entre l'audition et la vision lors de la perception de la parole [35]. Il suggère que la perception de la parole est multimodale. Pour montrer l'effet McGurk, classiquement on présente une vidéo montrant une personne prononçant une syllabe [ga] alors que la bande sonore diffuse l'enregistrement d'une autre syllabe [ba]. On a alors l'impression d'entendre une troisième syllabe intermédiaire [da] comme il est illustré dans la figure 2.2. Afin de comprendre ce phénomène, il faut se référer aux propriétés articulatoires des consonnes. Les phonèmes [b], [d] et [g] sont en Français des consonnes occlusives orales sonores. Cependant, leur point d'articulation diffère, il est : labial pour le [b] ; dental pour le [d] ; et vélaire ou vélo-palatal pour le [g].

L'effet McGurk est robuste : on a beau connaître l'effet, on y reste sensible. Cela diffère de certaines illusions d'optique, qui disparaissent une fois que l'on connaît le mécanisme. Il est cependant parfois difficile de le mettre en évidence en situation de silence, puisque le percept auditif est très bien audible. Il peut parfois être nécessaire d'ajouter un peu de bruit afin de dégrader le percept auditif et de favoriser l'émergence de ce phénomène d'intégration audio-visuelle. L'effet McGurk est utilisé afin de produire des programmes de reconnaissance de la parole plus précis en se servant d'une caméra vidéo et d'un logiciel de lecture sur les lèvres.

L'effet McGurk est nommé ainsi d'après son découvreur, Harry McGurk et a été d'abord décrit dans McGurk et MacDonald [36].

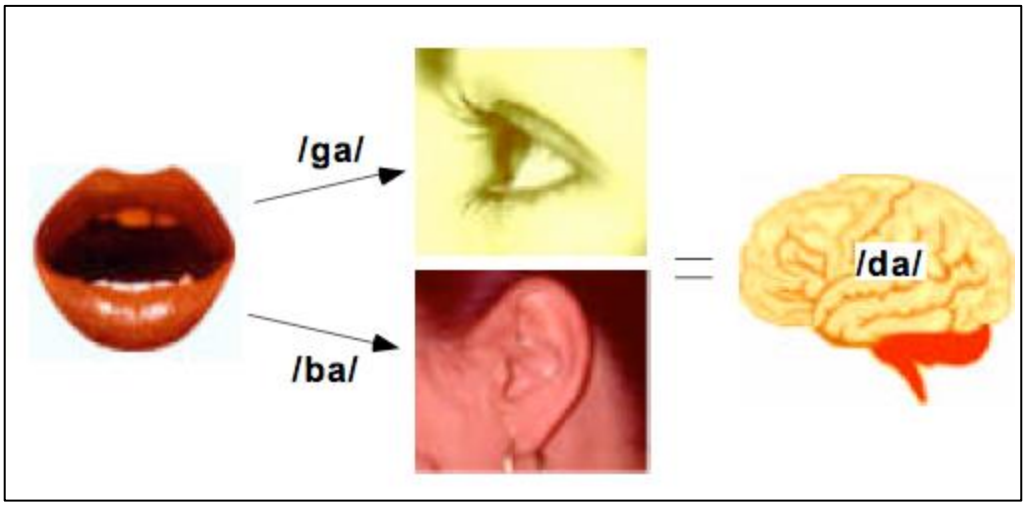


Figure 2.2 : Illustration de l'effet McGurk [37]

Ce percept illusoire peut s'expliquer par le fait que les informations auditives et visuelles de la parole dans les conditions habituelles, sont complémentaires et concordantes. Plus récemment, Massaro et Stork ont montré que cet effet pouvait également être observé pour des phrases entières [38]. La présentation visuelle de la séquence « my gag kok me koo grive » doublée de la séquence sonore « my bab pop me poo brive » induit la perception de la phrase « my dad taught me to drive ». Dans ce cas, l'illusion perceptive est d'autant plus forte que les deux séquences isolées n'ont aucun sens et que seule leur fusion permet d'obtenir une phrase intelligible, cohérente pour le sujet.

De manière naturelle l'interaction entre les perceptions auditive et visuelle de la parole opère en coopération dans les trois situations suivantes [39] :

- localisation et focalisation de l'attention sur un locuteur particulier dans un environnement où d'autres parlent en même temps (effet « cocktail-party ») ;
- redondance entre les informations acoustique et visuelle lorsque les deux modalités sont bien perçues, entraînant un gain d'intelligibilité systématique quel que soit la qualité de décodage dans chaque canal ;
- complémentarité entre les informations acoustique et visuelle lorsque du bruit ambiant dégrade la perception auditive pure.

Summerfield a comparé les réponses de sujets pour la reconnaissance de séquences comportant des consonnes en contexte vocalique [VCV], en condition auditive seule et en condition visuelle seule [40]. L'arbre de confusion des réponses auditives montre une organisation globalement inverse de son équivalent visuel (figure 2.3), ce qui est bien

perçu acoustiquement ne l'est pas visuellement et vice versa. Notamment, les résultats montrent un discernement visuel entre [p], [t] et [k] plus efficace qu'en acoustique. A l'inverse une forte confusion visuelle entre [p], [b] et [m], car ils sont caractérisés par une même fermeture bilabiale, disparaît au niveau acoustique. Walden a rapporté des résultats similaires avec des sujets spécialement entraînés à la lecture labiale [41]. Une des propositions de Summerfield sur cette complémentarité est d'associer les articulateurs visibles (lèvres, dents et langue) à la production des sons de fréquence élevée, sons provoqués par des mouvements rapides comme lors de certaines consonnes occlusives. Ils correspondent acoustiquement à des turbulences de faible intensité sonore dont la sensibilité au bruit acoustique est alors corrigée par l'information visuelle apportée par leur articulation. A l'inverse, la position des articulateurs non visibles (langue, velum, larynx) produisent des sons constants, de forte intensité, à des fréquences basses caractéristiques notamment du mode d'articulation (nasal ou oral) et des voyelles [42].

On peut aussi expliquer cette complémentarité à travers les résultats présentés par Fant la résonance de la cavité arrière (non visible) correspond généralement au premier formant, alors que le second formant correspond plutôt à la cavité avant. Si le premier formant présente une bonne stabilité, le second varie davantage. Les phénomènes de coarticulation sont largement exploités dans la perception audiovisuelle de la parole. La vision des lèvres, auxquelles il est lié, renforce alors la stabilité de la perception [43].

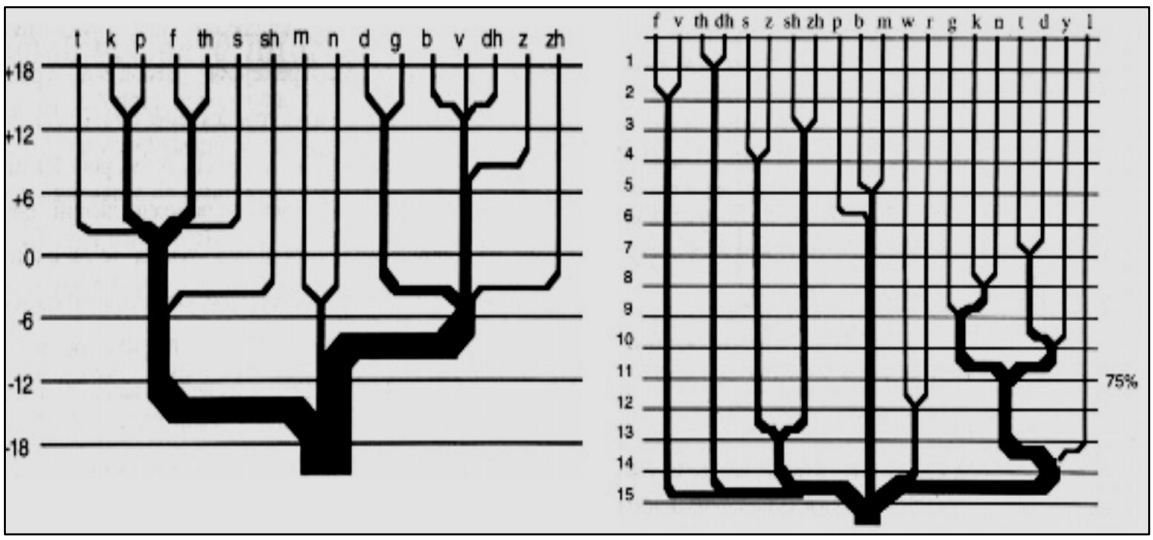


Figure 2.3. Arbres de confusion auditive et visuelle des consonnes [44]

2.8. Intelligibilité de la parole audiovisuelle

La lecture labiale chez certains déficients auditifs prouve la capacité du visage d'un locuteur à porter de l'information linguistique [39]. Cette faculté se retrouve chez des sujets ne présentant aucune perte auditive. Bien sûr, la perception auditive reste alors prépondérante sur la perception visuelle tant que le signal acoustique est suffisamment clair. Par contre, en présence de bruit, l'information visuelle contribue de manière significative à augmenter l'intelligibilité du signal de parole par effet à la fois de redondance et de complémentarité. La bimodalité intrinsèque de la perception de la parole a été illustrée à travers de nombreuses expériences d'intelligibilité en milieu acoustiquement dégradé [24]**, [**44-49].

Dans toutes les conditions, toutes les informations du visage sont utilisées. L'intelligibilité dans le bruit est meilleure lorsqu'on peut voir tout le visage et non pas que la bouche. Le terme de lecture labiale n'est donc pas adéquat, l'aide apportée par la "lecture labiale" est très connue chez les sourds puisqu'ils peuvent détecter les mots avec un pourcentage de succès de 90 % [39].

Si on constate une augmentation de la compréhension des deux modalités incluant des informations visuelles même avec un bruit faible, plus le bruit sera fort par rapport au signal audio et plus le gain en compréhension sera important. En outre, la différence entre les modalités « lèvres seules » et « visage entier » suggère que le cerveau intègre d'autres indices visuels que le seul mouvement des lèvres (figure2.4)**.**

Dans toutes les conditions, toutes les informations du visage sont utilisées. L'intelligibilité dans le bruit est meilleure lorsqu'on peut voir tout le visage et non pas que la bouche.

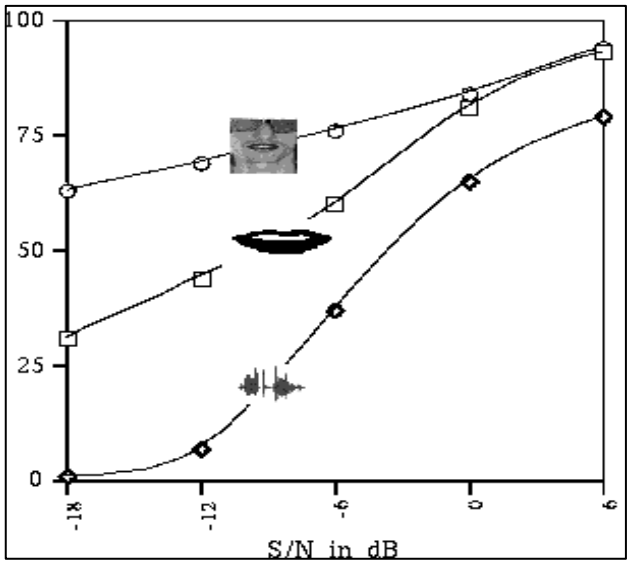


Figure 2.4. Comparaison de l'intelligibilité de la parole bimodale en condition bruitée en ajoutant successivement les lèvres puis tout le visage du locuteur [39].

2.9. Reconnaissance automatique de la parole

La démocratisation progressive des ordinateurs a conduit au développement d'interfaces de plus en plus intuitives entre l'Homme et la Machine [50]. La souris et les interfaces graphiques font partie de cet effort visant à rendre accessibles et ergonomiques les ressources des ordinateurs personnels. La commande vocale découle de la même volonté et constitue l'une des utilisations les plus évidentes des techniques de reconnaissance automatique de la parole.

N'utilisant d'abord que le canal audio pour des résultats mitigés, les chercheurs de ce domaine ont rapidement tenté d'exploiter l'apport du visuel pour la compréhension de l'oral une fois que celui-ci fut démontré et quantifié. De plus en plus de méthodes ont utilisé cette information pour augmenter leurs performances au cours des dix dernières années et plus particulièrement les scores de reconnaissance en milieu bruité. Historiquement, ce sont d'ailleurs les besoins de la reconnaissance de la parole qui ont engendré les premiers travaux sur la segmentation des lèvres.

Si le premier algorithme exploitant l'information visuelle se contentait d'une description très générique des lèvres (hauteur, largeur, surface) et fusionnait les données audiovisuelles de façon séquentielle (le visuel servant uniquement à résoudre les ambiguïtés de l'audio), les techniques de fusion de données se sont sophistiquées et ont

commencé à prendre en compte toute la finesse de description des lèvres que pouvaient fournir les algorithmes de segmentation [51].

2.10. La lecture labiale

La lecture labiale ne sert pas qu'aux malentendants ; chacun de nous a la capacité de lire sur les lèvres et peut s'en servir notamment en situation de communication bruitée. L'entraînement à la lecture labiale est long et difficile. Il demande un grand effort d'attention de la part de la personne qui s'engage dans la voie de cet apprentissage. Mais ceci ne représente que l'aspect déchiffrage des composants du message. En plus de ce travail de déchiffrage des mots, un travail de suppléance mentale est nécessaire à la construction du message. Il faut que les mots aient un sens. Cette lecture ne peut être de qualité que si le sujet connaît la langue qu'il déchiffre. Ceci est essentiel et indispensable.

La lecture labiale peut être considérée comme une aide à la compréhension de messages utilisant, pour véhiculer, la langue parlée. En effet, lorsque nous parlons, les phénomènes articulatoires sont dépendants des lois phonétiques et articulatoires. Il y a des influences réciproques des différents phonèmes les uns sur les autres, et dans ces influences il y a des dominantes, certains phonèmes dominent les autres, les commandent. Il est donc essentiel lorsqu'on essaie de communiquer oralement et de lire sur les lèvres, de connaître ces interactions et en particulier, l'usage des liaisons. C'est quelque chose que l'on apprend à l'école mais les liaisons se sont établies naturellement pour les enfants qui entendent pendant le temps d'appropriation du langage. Elles sont utilisées dans tout discours et ont besoin d'être connues de la personne sourde.

2.11. Limites de la lecture labiale

La lecture labiale rencontre rapidement des limites. On estime en effet que les « bons » lecteurs labiaux, notamment les adultes « devenus sourds » n’ont un accès direct qu’à 30% des énoncés émis par leurs interlocuteurs. La suppléance mentale demandée est de ce fait extrêmement importante pour la compréhension et est intimement liée au niveau langagier acquis mais aussi aux capacités de reconnaissance phonologiques, à des informations non langagières tirées du contexte de l’énonciation, aux références culturelles.

La lecture labiale s'appuie sur le déchiffrage des mouvements des lèvres, l'observation des dents, des mouvements de la langue et des mâchoires ainsi que sur les mouvements du visage. Les difficultés à décrypter le message sont dues à [52]:

- la place de la langue qui, selon les phonèmes se réalise sur quatre points d'appui différents (voile du palais, luette, palais, dents). On ne la voit que lorsqu'elle prend appui sur les dents [k], [g], [r] ;
- la coarticulation altère l'image labiale d'un phonème en fonction de la prononciation de celui qui le précède ou qui le suit ; exemples : si on dit :« la », on voit le [l] et si on dit : « lou », on ne le voit pas. Pour la phrase « Qui est là ? », on ne perçoit pas le « qui » visuellement. Les phonèmes [l-t-d- n] ne sont pas visibles associés à [u], dans « loup, toux, doux, nous » ;
- certains phonèmes sont invisibles : la prononciation de [r], [k], [g],[s], [z], [t], [d], [n] [z], [ʃ] et [j] n'induit pas de mouvement labial. Le jeune sourd perçoit le mot « pourquoi » en « pou a » et la question « Qui est-ce ? » peut devenir pour lui « i è ? ». Certains mots n'ont aucune image labiale comme « tisse », « coule ». Ainsi des mots à valeur syntaxique ne sont visibles que si l'on ralentit le débit de la parole et qu'on accentue l'articulation : « **le** chat est **sous** la table ».
- certains phonèmes sont des sosies labiaux c'est-à-dire que leur image labiale est identique, c'est le cas de [p] / [b] /[m], [d] / [t] / [n], [s] / [z], [ʃ] / [j] ; on ne peut pas distinguer le « o » et le « on » puisque leur différence tient au fait que ce dernier est nasalisé ;
- des mots ou des expressions entières peuvent être des sosies labiaux ; comme cocorico/coquelicot, menthe à l'eau/pantalon, « Mets ton manteau. » / « Prends ton ballon».

En pratique, la lecture labiale est malheureusement assez inefficace pour la majorité des sourds et des situations en raison des fautes de prononciation et des ambiguïtés du discours. C'est la raison pour laquelle des techniques spéciales, des langues ou langages et des codes sont utilisés pour communiquer avec des sourds ou entre sourds (langue maternelle labialisée + compléments gestuels codés ou langue des signes, et ce quelle que soit la langue maternelle parlée). Notons que pour une meilleure compréhension des sosies, la technique du Langage Parlé Complété (LPC) a été mise au point. Cette méthode d'aide à la lecture labiale code manuellement les différents phonèmes de la langue française. La

position de la main renseigne sur les sons vocaliques tandis que différentes configurations digitales codent les sons consonantiques : les Sosies labiaux sont alors codés par des signes différents qui évitent toute confusion.

2.12. Langue française Parlée Complétée

Le LPC (*Cued Speech* en Anglais) n'est pas une langue mais un moyen pour les sourds, et notamment les enfants sourds, de recevoir la langue française par la vue LPC, comme l'entendant la reçoit par l'ouïe. La main du locuteur, placée près du visage complète le mouvement des lèvres, permettant ainsi de lever l'ambiguïté existant entre plusieurs phonèmes (sons élémentaires) correspondant au même mouvement des lèvres.

La lecture sur les lèvres consiste à identifier les sons en fonction des déformations subies par la bouche. Cependant, la lecture labiale ne donne que des informations incomplètes qui, sans le son, ne peuvent être levées qu'avec le contexte de la conversation (on estime que la lecture labiale permet de percevoir seulement le tiers du message oral). Par exemple, les sons [pa], [ba] et [ma] sont produits de la même façon au niveau de la forme des lèvres, ce sont des sosies labiaux.

Le code LPC associe des gestes des doigts et de la main à la parole. Chaque syllabe est définie par une configuration des doigts, représentant une consonne, et par une position de la main près du visage, représentant une voyelle. Il existe 8 configurations pour coder les consonnes et 5 positions pour coder les voyelles (figure 2.5). Le LPC fournit ainsi 40 combinaisons différentes qui permettent de lever toutes les confusions de la lecture labiale. Pris isolément, le code ne donne qu'une information partielle sur le message (à l'instar de la lecture labiale), mais la combinaison de l'image labiale et de la clé manuelle permet de visualiser la totalité du message oral.

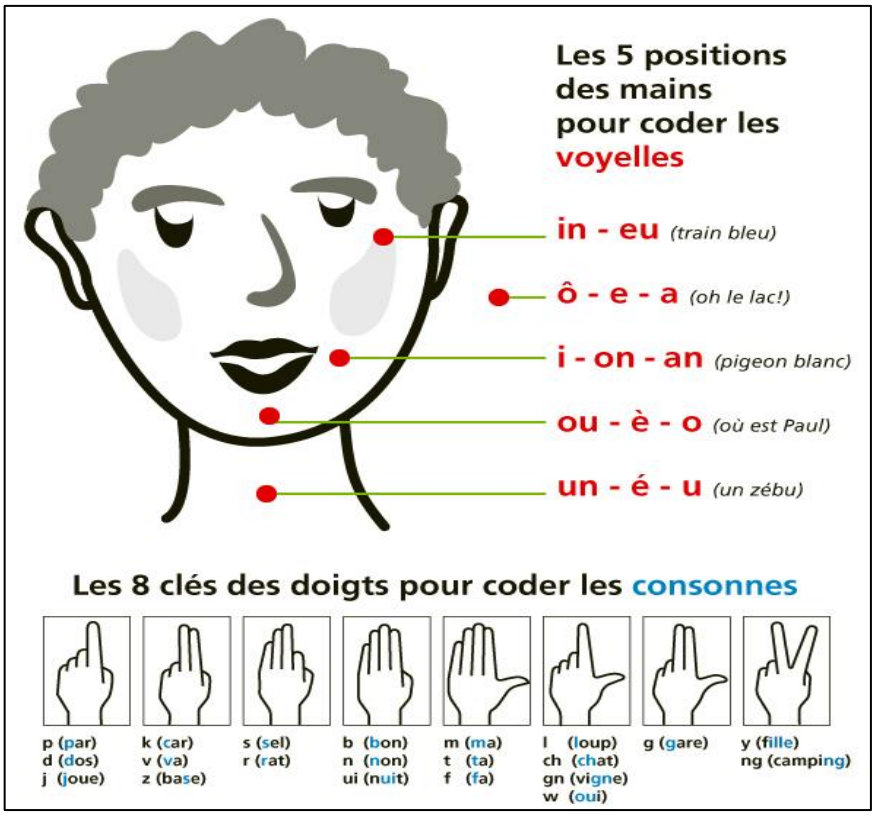


Figure 2.5 : Configuration des doigts, représentants les consonnes et les voyelles

2.13. Analyse labiale : Etat de l'art

Nous présentons la segmentation des lèvres en regroupant les méthodes selon l'approche [50] :

- Apparence ou pixel : les lèvres sont segmentées en exploitant les valeurs des pixels dans les espaces couleurs ;
- forme : méthode détectant les contours des lèvres en utilisant des modèles plus ou moins complexes de la forme de la bouche ;
- forme et apparence : méthode utilisant des modèles pour décrire à la fois la forme et l'apparence.

Il est à noter que ces approches ne répondent pas forcément aux mêmes objectifs de précision et de finesse et ne sont pas nécessairement mutuellement exclusives, ainsi une méthode de type pixel peut par exemple être utilisée comme prétraitement d'une méthode utilisant un modèle de lèvres.

2.13.1 Méthode avec une approche pixel

Cette famille de méthodes regroupe toutes les approches exploitant la distribution des couleurs présentes dans une image, sans considération de forme ou de contour. Elles prennent donc comme postulat que les lèvres correspondent à un groupe de pixels de caractéristiques homogènes dans un espace couleur donné et vont tenter de segmenter l'image en des régions lèvres et autres.

L'espace couleur le plus fréquemment employé est le système RVB (Rouge Vert Bleu), mais le YCbCr (où Y est la luminance, et Cb et Cr sont des composantes chromatiques du Bleu et du Rouge), le TLS (Teinte, Saturation, Luminance), qui est le système se rapprochant le plus de la perception humaine et les espaces perceptifs (Luv, Lab) font également partie des espaces utilisés régulièrement pour cette tâche. .

Les techniques les plus élémentaires de cette famille reviennent ainsi à faire des seuillages sur une grandeur colorimétrique jugée pertinente. Ce principe est utilisé dans des travaux, où l'on procédait à un simple seuillage de l'image de luminance [51], et encore récemment dans d'autres travaux où un deuxième seuillage est effectué sur la composante R du RVB [52]. Pour s'affranchir de la dépendance à l'éclairage des simples niveaux de gris ou du RVB, de nombreux auteurs ont proposé d'utiliser d'autres espaces couleurs, où la séparation entre les lèvres et le reste du visage serait facilitée grâce à l'utilisation de composantes chromatiques [53].

Dans le travail de X. Zhang [54], les auteurs exploitent par exemple la teinte ; tandis que R. Hsu introduit une grandeur synthétique $(Cr/Cb) - Cr^2$ où il utilise une combinaison des différentes composantes du format YCrCb pour obtenir une zone de détection labiale [55].

De nombreuses méthodes de classification statistique ont également été utilisées afin d'effectuer cette segmentation de façon plus sophistiquée que par un simple seuillage. Par exemple, W. Liew traite de l'utilisation des techniques d'agrégation floues, où les pixels sont classés à partir d'une carte d'appartenance aux lèvres dans l'espace perceptif [56].

2.13.2 Méthode avec une approche forme

Cette famille de méthodes vise à effectuer la segmentation des lèvres par une approche contour. Les valeurs des pixels sont utilisées pour faire converger la recherche sur les

contours recherchés mais de façon indirecte, comme par exemple, par l'intermédiaire des images gradients.

On peut alors séparer les méthodes selon qu'un modèl est utilisé ou non pour décrire les lèvres et le cas échéant, le degré de complexité et la nature du modèle utilisé. On obtient alors trois catégories :

- les méthodes sans modèle spécifique de lèvres ;
- les méthodes utilisant un modèle analytique ;
- les méthodes utilisant un modèle statistique.

2.13.2.1 Méthodes sans modèle de lèvres

Les exemples les plus communs de ce type d'approche sont les méthodes faisant appel aux contours actifs ou snacks. Cette méthode populaire a été introduite et désigne des courbes paramétriques flexibles ayant la capacité de se déformer de façon à converger sur les contours d'un objet quelconque [57]. La méthode des snacks pouvant s'adapter potentiellement à tous types d'objets, son application aux contours des lèvres a été envisagée à plusieurs reprises, mais cette méthode a rencontré quelques problèmes.

Le problème de l'initialisation est en effet très pointu, et limité quand il s'agit d'envisager des applications totalement automatiques et pouvant s'appliquer à tous types de prises de vue, sans contrôle des conditions. En outre, mis à part la contrainte des deux paramètres d'élasticité et de courbure (dont le réglage optimal pour toute situation peut se révéler ardu, voir impossible, à trouver), le snack a une forme totalement libre et peut donc converger vers une forme qui ne ressemblera parfois nullement à des lèvres réalistes.

Un exemple parmi d'autres applications des contours actifs à la zone labiale est donné dans certains travaux [4]. Néanmoins, si les résultats se révèlent très satisfaisants si les conditions sont bonnes (éclairage contrôlé, bon contraste entre la couleur de la peau et celle des lèvres), les performances décroissent lorsque ces conditions favorables ne sont plus réunies, la phase d'initialisation du snack manquant de robustesse ce qui nécessite de compliquer la méthode en procédant en plusieurs phases de convergence (figure 2.6).

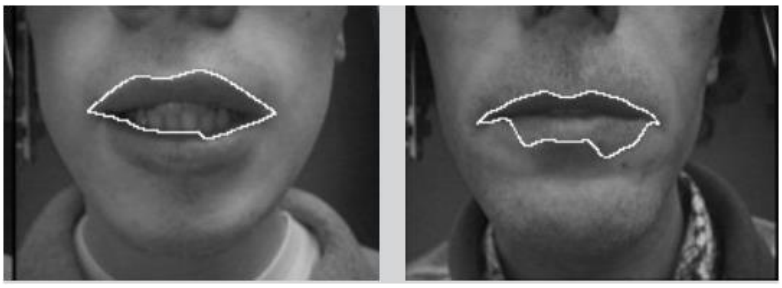


Figure 2.6 : Résultats d'une mauvaise initialisation et un mauvais choix de paramètres

2.13..2.2 Méthodes avec des modèles de lèvres analytiques

Dans l'objectif d'avoir un contour détecté plus conforme à la réalité par rapport aux méthodes n'utilisant que l'apparence ou n'ayant pas de modèle de la forme des lèvres, de très nombreux auteurs ont proposé d'utiliser des modèles paramétriques ou patrons déformables (deformable templates). Les lèvres seront alors décrites par un certain nombre de points de contrôle pertinents et de courbes qui constituent une forme prototype pouvant se déformer en jouant sur un jeu de paramètres de contrôle. L'une des difficultés de ces approches est de trouver le bon dosage de flexibilité : un modèle peu flexible donnera toujours une forme de lèvre plausible mais échouera parfois à segmenter des formes de bouches ou des configurations moins habituelles alors qu'un modèle trop flexible donnera des formes non réalistes dans certains cas. Une fois le modèle analytique déterminé, il convient encore de déterminer des critères pour paramétrer correctement le modèle sur une image inconnue.

A titre d'exemple, dans son article, Horbet et al ont décidé d'améliorer l'approche par contours actifs en rajoutant une force intérieure baptisée : force template. A chaque itération de la convergence du snack, la force template ramène le contour actif sur une forme admissible de lèvres. L'utilisation du template fournit en quelque sorte une connaissance a priori des formes possibles et contraint le snack à des déformations encadrées [58].

Dans un souci d'avoir un modèle suffisamment flexible afin de pouvoir modéliser n'importe quelle configuration de lèvres, N. Eveno a proposé d'utiliser quatre courbes paramétriques cubiques pour décrire le contour extérieur des lèvres en imposant conditions et limites aux dérivées des courbes au niveau des points saillants [5].

Après une initialisation où un snack simplifié (sans force intérieure) donne une première estimation du contour supérieur et la détection d'un point sur le contour inférieur, les

paramètres des cubiques sont optimisés en maximisant le flux de vecteurs gradients à travers les différentes courbes.

Les méthodes de cette famille donnent parfois d'excellents résultats, en fonction de la pertinence de la façon dont les auteurs auront paramétré la flexibilité d'une bouche déformable. L'étape de création et de calibration d'un modèle analytique étant néanmoins très longue, des chercheurs ont obtenu des modèles directement à partir d'un ensemble de données réelles, en utilisant une approche statistique [6].

2.13.2.3 Méthode basée sur un modèle statistique de la forme

Cette catégorie de méthodes correspond presque exclusivement aux Modèles Actifs de Forme (ASM) [59]. Les ASMs correspondent en fait à l'application des Modèles de Distribution de Points (Point Distribution Model: PDM) [60] pour segmenter un objet sur une image.

Le principe des PDMs sont construits à partir d'exemples d'entraînements : les contours de l'objet que l'on veut modéliser seront représentés par un nombre N de points qui doivent être étiquetés manuellement sur un assez grand nombre M d'images. On va alors disposer de M vecteurs ***xi*** contenant les coordonnées des points (dans un espace à deux ou trois dimensions selon les cas) qui, après une indispensable normalisation (par exemple grâce à la transformation procrustéenne généralisée), constitueront la base d'apprentissage du PDM.

On calcule alors le vecteur moyen :

$$\overline{x} = \frac{1}{M} \sum_{i=1}^{M} xi \qquad (2.1)$$

Puis l'on procède à l'Analyse en Composante Principale (ACP) de la matrice de covariance centrée. Toute forme pourra alors être approximée grâce à l'équation :

$$x = \bar{x} + P_S \, b_S \qquad (22.)$$

Qui correspond à la somme pondérée des n vecteurs propres p_i les plus significatifs qui correspondent à une portion choisie de la variance totale et qui sont rangés dans la matrice $Ps = [p_1,\ldots,p_n]$ avec $bs = [b_1,\ldots,b_n]$ un vecteur contenant les poids affectés à chaque mode propre.

Par opposition aux modèles paramétriques qui sont bâtis de façon empirique par des experts, les modèles statistiques ne nécessitent donc pas de réflexion sur la paramétrisation de patron ou de dosage de flexibilité. Le modèle sera, en effet, naturellement capable de se déformer de façon à reproduire toute forme présente dans la base d'apprentissage, mais la limite en est justement la phase d'apprentissage. Le temps économisé à paramétrer son patron déformable est en partie perdu lors de l'étiquetage manuel de la base d'apprentissage. En outre une configuration absente de l'apprentissage sera probablement impossible à segmenter ultérieurement sur une image inconnue, le modèle manquant de flexibilité en dehors de ce qu'il a appris à faire.

soigneusement faits, les ASMs se révèlent performants et robustes et, s'ils ont d'abord été utilisés dans le domaine biomédical, leur champ d'application s'est étendu à de nombreux domaines, dont l'application labiale [6].

2.13.3. Méthodes avec une approche combinant forme et apparence

Cette famille de méthodes comprend tout d'abord un certain nombre de techniques présentant des caractéristiques les plaçant à la frontière entre les méthodes présentées au 2.13.1 et 2.13.2.

Y. Tian et al présentent une méthode est où un modèle de forme analytique à plusieurs états (ouvert, relativement fermé, étroitement fermée) est utilisé en parallèle à une description de la distribution colorimétrique des lèvres par une Mixture Gaussienne [61].
Un algorithme de suivi de mouvement inspiré par la méthode Lucas-Kanade combinant les informations de forme et de couleurs permet alors d'effectuer la segmentation [62].
Très rapidement après la création des ASMs, la volonté de modéliser non seulement la forme mais également l'apparence a en effet donné naissance à des méthodes utilisant une description des Niveaux de Gris présents sur l'image.

Parmi les très nombreux travaux sur les AAMs (Modèles Actifs d'Apparence) appliqués au visage humain, nous pouvons encore citer parmi les approches originales de P. Daubias où est proposée une construction semi-automatique du modèle d'apparence. Les mouvements labiaux sont d'abord enregistrés alors que le locuteur a été préparé : ses lèvres ont été teintées en bleu de façon à capturer parfaitement la forme, ce qui permet de construire le PDM. Un deuxième enregistrement est ensuite effectué, sans maquillage bleu, durant lequel le locuteur doit effectuer les mêmes mouvements que précédemment. Ainsi,

la forme étant censée être connue, on peut alors prélever l'apparence correspondante. La principale limitation de cette stratégie est que les mouvements labiaux entre les deux enregistrements ne pourront jamais être parfaitement identiques, ce qui entraînera des erreurs sur le modèle statistique, même si les résultats présentés suggèrent que cette imprécision est acceptable [63].

1.14. Conclusion

Ce chapitre nous a permis d'introduire certains concepts de la perception visuelle de la parole sur le plan physiologique, acoustique et phonétique. Aussi nous avons introduit l'intérêt de la lecture labiale pour des sujets présentant un handicap dans l'audition. Nous avons présenté aussi une étude brève sur le LPC. Par la suite, nous avons présenté les principales méthodes et approches pour effectuer la segmentation des lèvres, permettant de faire une analyse labiale.

CHAPITRE 3 :
Segmentations et Contours Actifs

3.1 Introduction

Dans ce chapitre nous allons présenter la notion générale de segmentation d'images, puis nous abordons le principe des contours actifs avec l'algorithme Greedy, qui est basé sur la rapidité d'exécution, et qui possède plusieurs points communs avec d'autres algorithmes. Nous présentons un état de l'art sur des hypothèses de travail qui seront utilisées dans la segmentation des lèvres, une méthode pour augmenter le contraste entre les pixels de la peau et des lèvres, nous étudions l'espace couleur, le gradient hybride, et la pseudo-teinte. Des différentes formes de snacks utilisées pour la segmentation labiale seront exposées. Enfin nous proposons des résultats de segmentations des lèvres.

3.2 Images et segmentation

La segmentation des images est une étape de base dans le traitement d'images, qui a pour but de rassembler des pixels entre eux, suivant des critères prédéfinis. Les pixels sont ainsi regroupés en régions, qui constituent une partition de l'image. Il peut s'agir par exemple de séparer les objets du fond, ou bien extraire des figures de cette dernière, qui emploie les propriétés de base des valeurs de Niveau de Gris pour détecter les points et les frontières d'isolation des formes et des figures.

La segmentation est un traitement de bas-niveau qui consiste à créer une partition de l'image *A* en sous-ensembles Ri, appelés régions [64], tels que:

$$\begin{aligned} &\forall i \quad : \quad R_i \neq \emptyset \\ &\forall i, j \,;\, i \neq j \quad : \quad R_i \cap R_j = \emptyset \\ &A = \bigcup_i R_i \end{aligned} \tag{3.1}$$

Une région est un ensemble connexe de points image (pixels) ayant des propriétés communes (intensité, texture,…) qui les différencient des pixels des régions voisines. Il n'y a pas de méthode unique de segmentation d'une image, le choix d'une technique est lié :

- à la nature de l'image (présence de bruit, de zones texturées etc.) ;

- aux opérations situées en aval de la segmentation (Reconnaissance des Formes, interprétation, diagnostic, etc.) ;
- aux primitives à extraire (Contours, segments de droite, régions, formes, etc.) ;
- aux contraintes d'exploitation (Complexité algorithmique, fonctionnement en temps réel, taille de la mémoire disponible en machine.

Les principales approches de segmentation d'image sont (figure 3.1) :

- la segmentation basée sur les contours ;
- la segmentation basée sur les régions ;
- la segmentation basée sur la classification ou le seuillage des pixels ;
- la segmentation basée sur la coopération entre les trois premières méthodes.

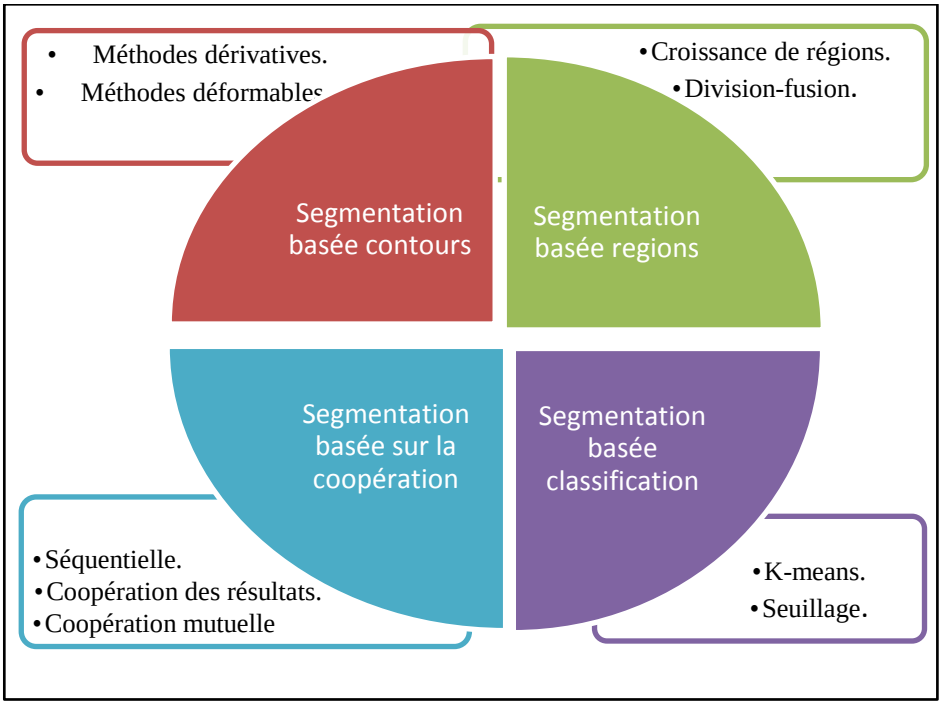


Figure 3.1 : Principales méthodes de segmentation d'images

3.2.1 – Segmentation basée sur les contours

La segmentation basée sur les contours est une approche qui consiste à délimiter les différents objets constituant l'image par leurs frontières (contours) en détectant les régions

de fortes variations d'intensité lumineuse, par conséquent, plus la variation est forte, plus le contour sera significatif.

Parmi les méthodes de segmentation basée sur les contours, nous pouvons citer les méthodes :

- dérivatives ;
- déformables.

- Les méthodes dérivatives consistent à estimer les dérivées de la fonction d'intensité, en appliquant une convolution par un filtre dérivateur. Parmi ces méthodes nous avons celles par différences finies et par filtrage optimal.

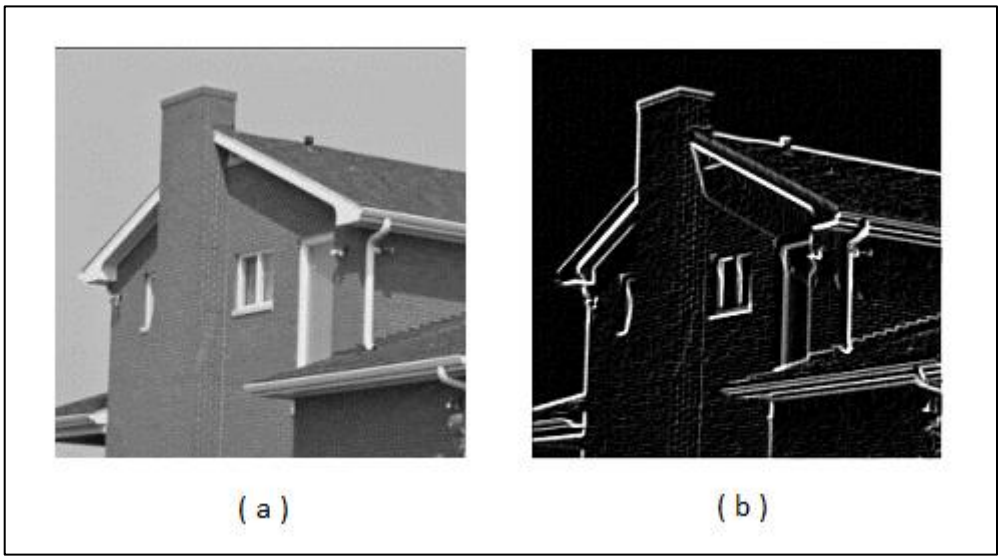


Figure 3.2 : Segmentation d'image par une méthode dérivative :
(a) image originale, (b) résultat de la segmentation.

Les méthodes déformables (souvent appelées contours actifs (CA) ou snacks en Anglais) sont basées sur une structure dynamique sous forme de courbes ou surfaces qui évoluent sous des contraintes précises d'un état de départ (contour initial) vers un état final qui représente le contour de l'objet que l'on souhaite segmenter.

Il existe plusieurs méthodes de segmentation basées sur cette approche : les CA classiques (snacks), les géodésique, les levels- set et les contours paramétriques.

3.2.2 Segmentation basée sur les régions

Les méthodes de segmentation basées sur les régions consistant à regrouper les pixels ayant des propriétés communes dans des régions. Selon des critères pouvant être la valeur de niveau de gris, de la couleur, de la texture, ou une combinaison de plusieurs informations [65].

Parmi les méthodes de segmentation basées sur les régions, nous avons la segmentation par :

- croissance de régions ;
- division-fusion.

La Segmentation par croissance de régions comporte deux étapes :

- initialisation : elle consiste à sélectionner de façon aléatoire ou automatique les germes des régions les plus représentatives de l'image ;
- la croissance des germes : les germes initiaux s'accroissent pour former des régions homogènes en respectant les contraintes de segmentation (les contraintes d'homogénéité), de forme géométrique ou de taille.

La Segmentation par division-fusion comporte deux étapes :

- la division : l'image initiale est divisée en régions (généralement en quatre quadrants). Ensuite, chaque région est analysée individuellement. Si celle-ci ne vérifie pas le critère d'homogénéité, alors elle est divisée en quatre blocs. Ce processus est appliqué à toutes les sous-régions jusqu'à ce que tous ces derniers soient indivisibles ;
- la fusion : cette étape consiste à fusionner les couples de régions voisins s'ils vérifient le critère d'homogénéité.

3.2.3. Segmentation basée sur la classification

Les méthodes de segmentation basées sur la classification des intensités des pixels sont les plus simples à mettre en œuvre. Car ils ne prennent en compte qu'un seul critère de segmentation, qui est le niveau de gris (ou la couleur) des pixels. Ils permettent de diviser l'image en « n » classes d'intensité, tels que les pixels de chaque classe appartiennent à un intervalle précis de niveaux de gris.

Avec cette approche de segmentation les classes peuvent être déterminées par les méthodes de seuillage, ou la méthode de nuées dynamiques (K-means) :

- Segmentation par Seuillage est le procédé de segmentation le plus simple. L'hypothèse qui sous-tend cette approche consiste à supposer que tout objet dans une image se différencie de l'arrière-plan. La seule difficulté avec cette approche de segmentation est l'éclairage non uniforme de l'image ;
- Segmentation par la méthode K-means : cette méthode est la plus utilisée parmi les méthodes de segmentations basées sur la classification, du fait de sa simplicité de mise en œuvre. Elle consiste à segmenter l'image en K différentes classes. Les valeurs de chaque classe résultant de cette méthode sont aussi proches et homogènes que possible les unes des autres, et aussi loin et hétérogènes que possible des valeurs des autres classes [65].

3.2.4. Segmentation basée sur la coopération

Comme il n'existe pas de méthode parfaite pour la segmentation d'image, les chercheurs qui s'intéressent au domaine de la segmentation ont développé une nouvelle approche qui se base sur la coopération des trois autres approches vues précédemment.

L'approche de segmentation par coopération région-contour donne de meilleurs résultats [66]. En effet, elle résout dans plusieurs cas les problèmes de l'approche contours et de l'approche régions en éliminant les fausses détections des contours.

3.3. Les contours actifs

Les CA sont introduits par l'équipe Kass, Witkin et Terzopoulos [67], tirent leur origine des modèles élastiques [68]. Les CA (ou snacks) tiennent leur nom de leur aptitude à se déformer comme des serpents. Depuis, les snacks sont devenus un sujet très important pour le traitement d'images. Les domaines d'utilisation des snacks sont nombreux tant en 2D qu'en 3D tels que : la Reconnaissance de Formes, la simulation, le suivi de scènes, la segmentation d'images, etc.

Les CA (ou snacks) sont des courbes déformables évoluant au gré de la minimisation de la fonctionnelle d'énergie E(v) associée. Ils se déplacent au sein de l'image d'une position initiale vers une configuration finale qui dépendra de l'influence respective des divers termes d'énergie en présence. L'énergie des snacks comprend principalement un terme d'énergie interne appelé énergie de régularisation ou de lissage et un terme d'énergie externe. Les snacks s'appuient sur les contraintes de forme pour guider la recherche des objets souhaités.

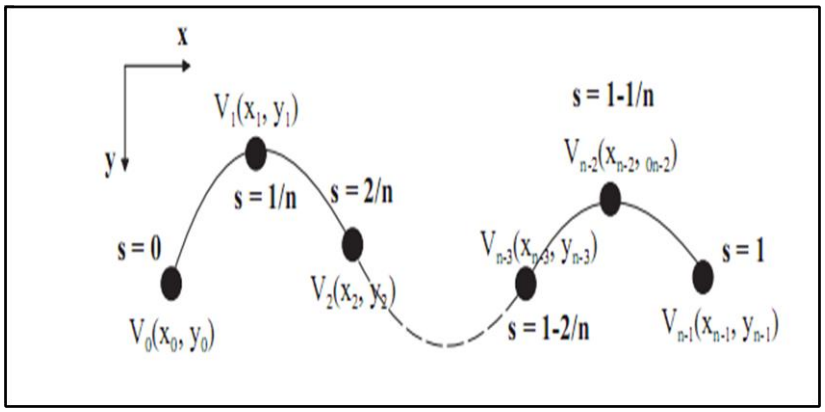


Figure 3.3 : Contours actifs : coordonnées cartésiennes et abscisses curvilignes pour un snack de n points [64]

Soit $v(s,t)$ la position d'un point de la courbe à un instant t, où s est l'abscisse curviligne à l'instant t, et (x, y) les coordonnées cartésiennes d'un point de l'image (Figure 3.3).

$$\begin{cases} v: [0,1] \to [0,\infty] \\ v(s,t) =^{t} \big(x(s,t), y(s,t)\big) \quad \forall (s,t) \in [0,1] \times [0,\infty] \end{cases} \tag{3.2}$$

3.3.1 Les différentes énergies

La fonctionnelle d'énergie E(v) attachée au CA est composée de trois énergies.

$$E(v): v \to Einterne(v) + Eexterne(v) + Econtexte(v) \tag{3.3}$$

3.3.1.1 L'énergie interne

L'énergie interne gère la cohérence de la courbe. Elle maintient la cohésion des points et la raideur de la courbe, autrement dit, elle gère la régularisation du CA qui est donnée par la formule :

$$E_{interne}(v) = \int_0^1 \left(\frac{\alpha}{2}(s) \parallel v'(s) \parallel^2 + \frac{\beta}{2}(s) \parallel v''(s) \parallel^2\right) ds \tag{3.4}$$

Les termes v', v'' sont les dérivées première et seconde de v par rapport à s, le terme de 1er ordre correspond à la tension. Il prend une valeur importante quand la courbe se distend. Lorsque $\alpha = 0$, la courbe peut présenter des discontinuités. Nous parlerons donc d'énergie de continuité. Le terme du 2ème ordre correspond à la courbure. Il prend une valeur importante lorsque la courbe s'incurve rapidement c'est-à-dire pour l'obtention de coins. Lorsque $\beta = 0$, la courbe peut prendre une forte convexité, par contre lorsque β est grand, la courbe tendra vers un cercle si elle est fermée ou une droite si elle est ouverte.

3.3.1.2 L'énergie externe

L'énergie externe correspond à l'adéquation aux données. Cette énergie externe prend en compte les caractéristiques de l'image. Rappelons, dans le cas classique, que ce sont les contours de formes qui sont recherchés, donc les points de fort gradient ou des points ayant une propriété de position par rapport à une couleur donnée :

- ***Le Gradient*** **:** Pour la recherche des zones de fort contraste dans l'image, est introduite la fonction :

$$E_{externe}(v) = -\int_0^1 \| \nabla I((v(s))) \|^2 \, ds \quad (3.5)$$

Où $\nabla I((v(s)))$ représente le gradient de l'image I en v(s)

- ***L'intensité :*** Cette énergie, au contraire, permet de sélectionner les zones sombres ou claires selon le signe choisi.

$$E_{intensité}(v) = \mp \int_0^1 \big(I\big(v(s)\big) - i_0\big)^2 \, ds \quad (3.6)$$

La valeur i_0 introduit ou non, un certain seuillage. On peut ainsi favoriser la position du contour dans une zone donnée.

- ***Gradient Vecteur Flow "GVF" :*** C. Xu constate la médiocrité de la qualité de la convergence de la courbe de contour actif vers le contour souhaité dans les zones à forte concavité, ce chercheur introduit un nouveau potentiel. Il s'agit d'une nouvelle force externe qui traduit la diffusion isotopique d'un flux externe. Il définit " GVF " comme le champ de vecteurs [69]:

. $$V(x,y) =^t [u(x,y)v(x,y)] \quad (3.7)$$

Ce champ minimise la fonctionnelle d'énergie. Ce nouveau potentiel est d'un intérêt certain lorsque l'objet à segmenter est unique, cela peut poser un problème dans le cas d'objets multiples dans des images réelles, la diffusion du gradient pouvant créer des interférences entre les zones d'influence des différents objets.

3.3.1.3. L'énergie de contexte

L'énergie de contexte, parfois appelée énergie de contrainte, permet d'introduire des connaissances a priori sur ce que nous cherchons. Entre autre, nous plaçons, sous cette rubrique, l'énergie ballon introduite par D. Cohen [70]. Les snacks, de par leur discrétisation

ont une tendance naturelle à se rétracter. La minimisation de l'énergie implique une minimisation de distance. La force ballon va tendre à gonfler le contour actif ou accélérer sa rétraction selon le signe de la force introduite. De plus, cette force va permettre de dépasser les contours présentant un faible gradient et ainsi de sortir du bruit pour atteindre une frontière plus fortement marquée. Il s'agit d'une force normale au contour en chaque point.

$$F_{Ballon}(v(s)) = K\vec{n}(s) \tag{3.8}$$

Où $\vec{n}(s)$ est un vecteur unitaire normal à la courbe en $v(s)$.

L'intensité de l'énergie ballon est un scalaire généralement négatif proportionnel à l'aire intérieure du contour.

3.4 Implémentation des CA classiques

Pour l'implémentation des CA classiques, il existe trois approches principales qui peuvent être recensées :

- l'approche Variationnelle introduite par Kass et al. qui tire son avantage des développements mathématiques de l'analyse numérique, c'est la méthode la plus développée, la plus courante et la plus déclinée [67].
- l'approche par Programmation dynamique introduite par Amini et al**.** qui utilisent les avancées du domaine de l'informatique [71].
- la troisième méthode pour la mise en œuvre des CA classiques, est l'utilisation de l'algorithme glouton (ou l'algorithme Greedy) introduit par Williams et Shah [72].

3.4.1. Différences finies

Les dérivées d'une fonction par rapport à une variable peuvent être approximées par les différences finies. Il a été montré par plusieurs chercheurs.

En utilisant la méthode des différences finies on peut approximer la dérivée première de la fonction d'énergie qui représente la continuité de la courbe :

$$\| V_i - V_{i-1} \|^2 = (X_i - X_{i-1})^2 + (Y_i - Y_{i-1})^2 \tag{3.9}$$

Et la dérivée seconde qui représente la courbure :

$$\| V_{i-1} - 2V_i + V_{i+1} \|^2 = (X_{i-1} - 2X_i + X_{i+1})^2 + (Y_{i-1} - 2Y_i + Y_{i+1})^2 \quad (3.10)$$

Il est à remarquer, qu'il existe une autre variante d'approximation de la dérivée première qui évite une forte rétraction du contour actif au cours de son évolution temporelle, elle a été proposée par Williams et Shah. Ces derniers utilisent la différence de la distance moyenne $\bar{d}$ de tous les points du contour par rapport à la distance qui sépare deux points consécutifs du contour. La nouvelle formule de la continuité est la suivante [72] :

$$\| V_i - V_{i-1} \|^2 = \bar{d} - |(X_i - X_{i-1})^2 + (Y_i - Y_{i-1})^2| \quad (3.11)$$

Mais Williams et Shah avaient tort, car certains points dont la distance était plus grande que la moyenne, celle-ci donnait une continuité négative. Ils étaient ainsi favorisés par rapport à d'autres dont la distance était plus proche de la moyenne. Pour remédier à ce problème, ces chercheurs ont proposé une nouvelle formule pour calculer la continuité des points d'un CA :

$$\| V_i - V_{i-1} \|^2 = \left| \bar{d} - ((X_i - X_{i-1})^2 + (Y_i - Y_{i-1})^2) \right| \quad (3.12)$$

3.4.2. Approche Variationnelle

Dans la méthode des contours actifs, il s'agit de minimiser une fonctionnelle d'énergie, composée d'une énergie interne, d'une énergie externe, éventuellement d'une énergie de contexte. La recherche du contour est limitée au cas d'une courbe plane :

$$s \longmapsto v(s) = (x(s), y(s)) \quad (3.13)$$

Où s est abscisse curviligne.

Le contour initial v(0) est défini par l'utilisateur et la courbe évolue avec une certaine vitesse. Le problème est de trouver cette vitesse avec laquelle la courbe évolue vers un minimum local correspondant aux contours des objets ou des régions à segmenter. Dans cette méthode on définit un modèle comme un espace de formations admissibles Ad (Plan). Nous voulons minimiser l'énergie E de $Ad\ vers\ \mathbb{R}$:

$$v \mapsto E(v) = \int_{\mathbb{R}} \alpha \left| v'(s) \right|^2 + \beta \left| v''(s) \right|^2 + P\big(v(s)\big) ds \quad (3.14)$$

où P(v(s)) est un potentiel de forme

Cette équation peut s'écrire $AV = F$

Où A est une matrice fonction de α et β, V représente les vecteurs de positions v_i et F les forces $F(v_i)$ en ces points.

V et F sont des matrices colonnes. Berger présente les différents cas des CA à extrémités fixes, à extrémités libres et le cas d'un modèle fermé, qui est de la forme la plus courante et la plus utilisée [73].

3.4.3. Programmation dynamique

La programmation dynamique est une méthode classique de résolution de problème d'optimisation. Son application aux CA est due à Amini et al. Cette approche peut être une alternative intéressante au calcul variationnel. Amini considère l'équation classique [71] :

$$E_{tot} = \int_0^1 E_{ext}\big(v(s)\big) + 0.5(\alpha(s)|v_s(s)|^2 + (\beta(s)|v_{ss}(s)|^2)ds = \int_0^1 (E_{ext} + E_{int}) \tag{3.15}$$

En représentant la fonction à intégrer par $F_v(s, v_s, v_{ss})$, la solution d'Euler Lagrange donne :

$$F_v - \frac{\partial}{\partial s} F_{v_s} + \frac{\partial^2}{\partial s^2} F_{v_{ss}} = 0 \tag{3.16}$$

En discrétisant avec $E_{int}(v_i) = 0.5(\alpha_i|v_i - v_{i-1}|^2 + \beta_i|v_{i+1} - 2v_i + v_{i-1}|^2)$ (3.17)

D'où $E_{tot} = \sum_{t=0}^{n-1} E_{int}(v_i) + E_{ext}(v_i)$ (3.18)

Il est possible de l'exprimer par :

$$E_{tot}(v_1, v_2, \dots, v_n) = E_1(v_1, v_2, v_3) + E_2(v_2, v_3, v_4) + \cdots + E_{n-2}(v_{n-2}, v_{n-1}, v_n) \tag{3.19}$$

Où $E_{i-1}(v_{i-1}, v_i, v_{i+1}) = E_{ext}(v_i) + E_{int}(v_{i-1}, v_i, v_{i+1})$ (3.20)

On se ramène donc à un problème d'optimisation d'une fonction numérique de plusieurs variables. Ces derniers seront les positions des différents points du snack.

La convergence de la minimisation de l'énergie est garantie, mais la complexité est élevée.

3.4.4. L'algorithme GREEDY

L'algorithme Greedy est une méthode itérative qui consiste à minimiser l'énergie totale de chaque point du CA, et de déplacer les points du CA pour approcher la frontière d'un objet tout en essayant de conserver certaines caractéristiques, telles que la courbure et la répartition des points du CA.

L'algorithme Greedy tente de trouver un meilleur positionnement pour la courbe afin de minimiser les dérivées par rapport aux contraintes utilisées. Pour chaque point de la courbe, il choisit un nombre de voisins pour lesquels on calcule l'énergie, ensuite on déplace le point sur le voisin qui possède l'énergie la plus faible. La méthode s'arrête lorsque le résultat satisfait un critère d'arrêt qui peut être :

- l'atteinte d'un nombre maximal d'itérations ;
- la stabilité du CA (lorsqu'il ne sera plus possible d'améliorer le positionnement de la courbe par rapport la frontière d'un objet) ;
- lorsque le critère de déformation (nombre de points qui ont été déplacées) devient constant ou inférieur à un seuil entre deux itérations successives.

3.5 Segmentation labiale

Le but est d'effectuer la localisation et la segmentation de la bouche sur des images de visage. L'étude porte sur le cas d'images de visage en couleurs dans lesquelles la bouche est visible. Beaucoup d'algorithmes de modélisation du contour externe de la bouche reposaient sur l'étude de la luminance. Le problème, lorsque l'on se base sur l'information de luminance, est la dépendance par rapport aux variations d'illumination de l'image. Suivant la direction de la source lumineuse par rapport au sujet, des réflexions spéculaires ou des ombres pourront apparaître sur les lèvres. Dans le cas d'une source de lumière située au-dessus du sujet, on pourra voir apparaître des ombres sous la lèvre supérieure et sous la lèvre inférieure.

Afin de traiter la segmentation et leur approche d'une manière générale, nous spécifions la segmentation que nous utilisons dans notre travail qui est la segmentation des lèvres.

La séparation des lèvres du reste du visage a toujours été au centre de la problématique du suivi labiale. Il est vrai qu'avec des couleurs presque identiques, il s'avère délicat d'obtenir un contour labial sans utiliser le moindre artifice. Pour contrer cela de nombreux auteurs se sont tournés vers l'étude des différents espaces couleur afin de pouvoir trouver le plus discriminant d'entre - eux pour effectuer cette segmentation.

En premier lieu, nous analysons les mélanges chromatiques associés aux lèvres et à la peau. Nous proposons d'utiliser une grandeur colorimétrique permettant d'effectuer une bonne séparation des lèvres et de la peau. De plus, nous introduisons un gradient hybride qui combine à la fois les informations de luminance et de chrominance, et qui facilite la localisation de la frontière supérieure des lèvres.

3.5.1 Analyse chromatique des lèvres et de la peau

De nombreuses études ont montré que l'utilisation de la couleur améliore significativement les performances des algorithmes d'analyse faciale. En effet, la peau est caractérisée plus par sa couleur que par sa luminance [74]. Ainsi, le point de départ de la plupart des méthodes d'analyse labiale est la détermination d'un espace couleur dans lequel la luminance et la chrominance sont exprimées séparément de manière explicite. De nombreux auteurs choisissent le système RGB, qui est le meilleur espace pour caractériser les lèvres par rapport à la peau [75].

3.5.2 Caractéristiques des lèvres dans le système RGB

L'espace couleur RGB est un système de couleur additif basé sur la théorie trichromatique. En combinant les trois primitives Rouge, Verte et Bleue, il est possible d'obtenir presque toutes les couleurs visibles.

Plusieurs études ont proposé de travailler dans l'espace couleur RGB pour extraire des informations labiales [76 - 79].

Les composantes chromatiques R, G et B ont des distributions de valeurs larges aussi bien pour les pixels peau que pour les pixels lèvres. Pour chacune des trois composantes, les distributions associées aux lèvres et à la peau se chevauchent fortement. En conséquence, il est difficile de définir deux zones distinctes, l'une correspondante aux pixels peau et l'autre aux pixels lèvres. Même si les lèvres sont généralement vues plus rouges que la peau, l'histogramme de la figure 3.4 montre que la composante R est prédominante par rapport aux composantes G et B, mais qu'elle est aussi importante pour les lèvres que pour la peau. En outre, les pics des distributions de R et de G sont plus éloignés pour la peau que pour les lèvres, ce qui explique que la peau apparaît plus jaune que les lèvres.

L'espaces *TLS* (Teinte, Luminance, Saturation) est un espace de représentation des couleurs plus proche de la perception humaine. Plusieurs formules mathématiques ont défini des espaces différents comme *HSV* (*Hue, Saturation, Value*), *HSI* (*Hue, Saturation, Intensity*) ou *HSL* (*Hue, Saturation, Lightness*), mais ils décrivent trois informations similaires :

- la teinte : le type de couleur ;
- la saturation ou la pureté : l'intensité de la couleur;
- la luminance ou la brillance.

L'espace HSV est très utilisé pour caractériser la peau [80] et [81].

On a constaté que la teinte des lèvres est relativement constante et bien séparée de celle de la peau, par conséquent, on peut l'utiliser pour localiser les lèvres [75].

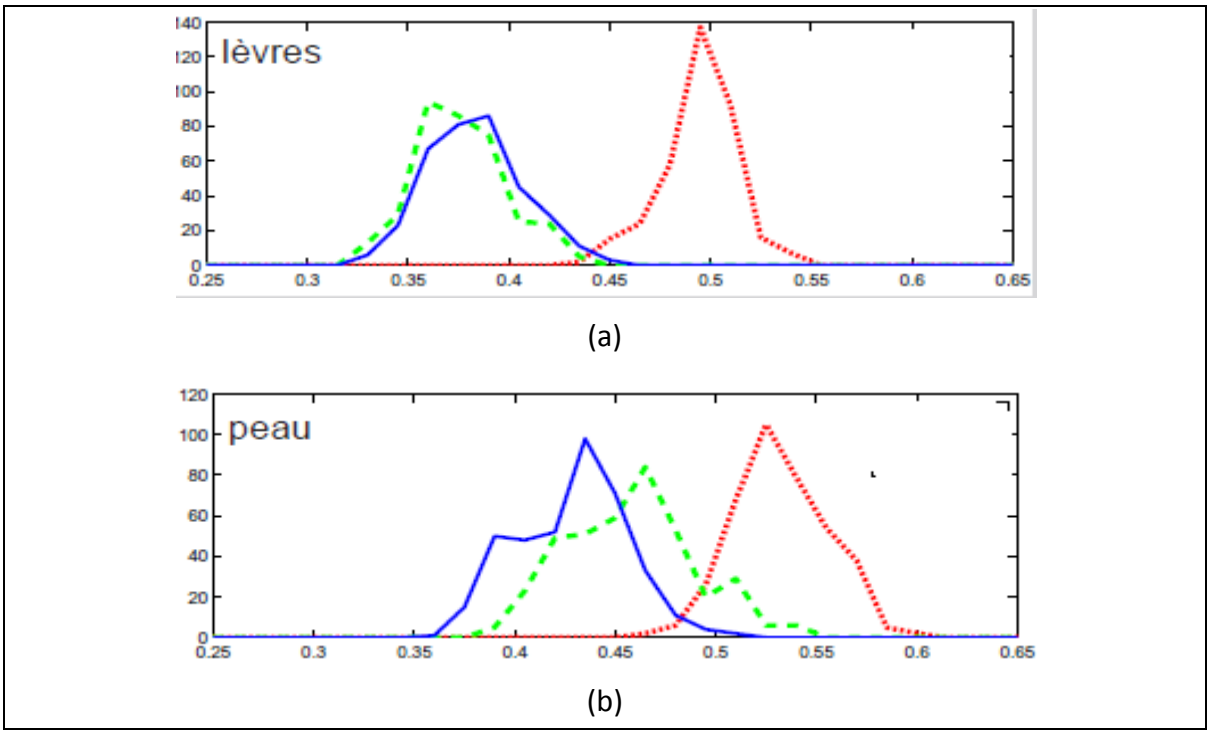


Figure 3.4. Histogramme de comparaison dans le système RGB entre la lèvre et la peau
a) lèvres b) peau

5.3.2.1 Calcul de la luminance (image en niveau de gris)

Vu que les filtres et les algorithmes sont faits pour des images à 256 niveaux de gris, une image couleur ne peut pas être traitée comme elle est ; elle doit être convertie en image à niveaux de gris. Les images en couleur, possédant trois composantes trichromatiques :

verte, bleue et rouge sont transformées en images possédant une seule composante grise, représentative des trois anciennes composantes, pour ne pas détériorer trop fortement l'image et donc les contours. La valeur de cette dernière est calculée selon la relation :

$$Y = 0{,}59\,V + 0{,}30\,R + 0{,}11B \tag{3.21}$$

C'est l'équation fondamentale de la luminance.

5.3.2.2 Pseudo-teinte

Les distributions des pixels lèvres et peau sont relativement concentrées, mais elles se chevauchent beaucoup. Nous remarquons qu'il est difficile de distinguer les lèvres par rapport au reste du visage. Des études portées sur des propriétés de l'espace couleur classique et des grandeurs colorimétriques sont destinées à augmenter le contraste entre la peau et les lèvres. Notre étude se base sur les grandeurs RGB qui sont adaptées au problème de la séparation des lèvres et de la peau. Nous avons constaté que la pseudo-teinte h et que la teinte H offraient les meilleures performances pour séparer la peau et les lèvres

La teinte usuelle H est exprimée par le quotient suivant :

$$H\,(x,y) = \frac{R(x,y)}{G(x,y)} \tag{3.22}$$

La lèvre est caractérisée par une valeur importante sur la couleur rouge devant la couleur verte, de nombreux auteurs ont utilisé cette propriété pour définir une grandeur colorimétrique caractéristique des lèvres. Par exemple, le quotient R/G a des valeurs plus importantes pour les lèvres que pour la peau. Il est utilisé pour localiser et segmenter les lèvres [82 - 84]. Cependant, lorsque G est faible, le quotient R/G peut prendre des valeurs très importantes, l'image est souvent bruitée. Pour résoudre ce problème, Hulbert et Poggio ont proposé d'utiliser une pseudo-teinte bornée entre 0 et 1 [85].

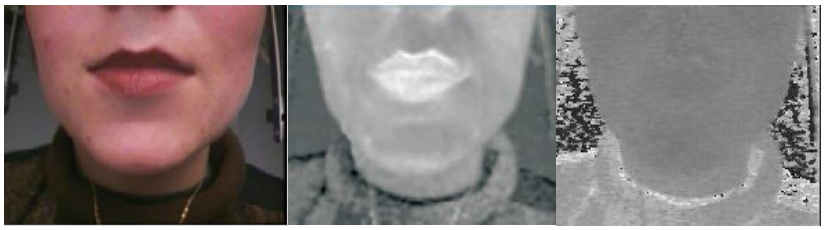
Nous utilisons la pseudo-teinte h, l'avantage de cette caractéristique est qu'elle présente un fort contraste entre les lèvres et la peau, ce qui correspond à un vecteur gradient de grande norme au niveau du contour [19].

La Pseudo-teinte h(x, y) est une amélioration de la teinte usuelle, h est calculée en un point (x,y) de l'image I comme :

$$h(x,y) = \frac{R(x,y)}{R(x,y)+G(x,y)} \tag{3.23}$$

où R et G sont les composantes RGB

La valeur de la pseudo teinte est comprise entre 0 et 1. Cette valeur de pseudo-teinte est grande pour les lèvres que pour la peau. La figure 3.5 montre La Pseudo-teinte et la teinte d'une image.



(a) (b) (c)

Figure 3.5. Teinte usuelle et la Pseudo teinte d'une image :
(a) Image de départ, (b) pseudo-teinte et (c) teinte

La pseudo-teinte est utilisée pour son pouvoir discriminant, elle est combinée avec la luminance en considérant que la lumière vient d'en haut et que la frontière supérieure est une zone de forte luminance.

3.5.2.3 Gradient hydride

Des études plus récentes, calculent le gradient à partir du plan intensité car la région de la bouche est caractérisée par des changements d'illumination entre les lèvres et la peau. Par exemple, dans les conditions d'éclairage les plus courantes, la source de lumière vient d'en haut et la frontière supérieure de la bouche est un contour avec une forte luminance alors que la lèvre supérieure est plus sombre. De la même manière, la frontière inférieure de la bouche est un contour avec une faible luminance alors que la lèvre inférieure est bien éclairée (Figure 3.6), [86 – 89].

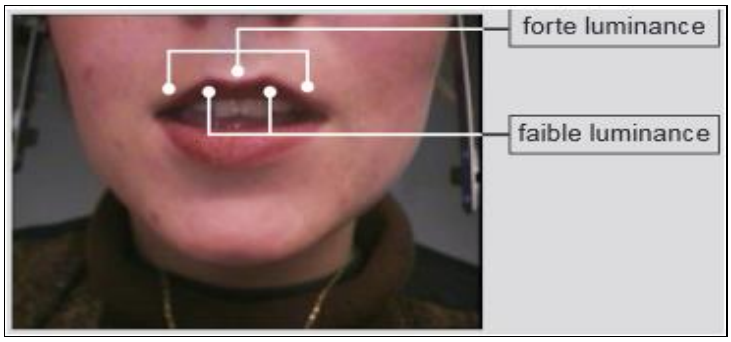


Figure 3.6 : Caractéristiques de luminance des différentes zones des lèvres

Lorsque l'on désire extraire les contours des lèvres, il est nécessaire d'accentuer la frontière entre les lèvres et la peau. Pour obtenir un fort gradient sur les contours des lèvres, il faut que l'espace de représentation de l'image permette une variation importante des valeurs des pixels lèvres et peau. En plus de leur couleur spécifique, les lèvres possèdent une structure particulière et génèrent des zones d'ombre caractéristiques.

Les espaces couleurs standards sont peu utilisés pour calculer des gradients directement à partir des composantes trichromatique ; Pour combiner les informations de chrominance et de luminance des lèvres, N. Eveno propose le gradient hybride Rtop (x, y). Pour le pixel (x, y) d'une image I qui est construit à partir de la pseudo-teinte h(x,y) pour extraire le contour extérieur des lèvres. Rtop permet d'accentuer le contour extérieur supérieur, il est calculé comme suit [19] **:**

$$\mathrm{Rtop}(x,y) = \nabla[hn(x,y) - Ln(x,y)] \tag{3.24}$$

Telle que :

∇ [ ] **:** est l'opérateur gradient (Sobel, Perwitt,……).

$h_n(x, y)$: pseudo teinte normalisée au point (x, y) et calculée par :

$$h_n(x,y) = \frac{h(x,y) - \min(h)}{\max(h) - \min(h)} \tag{3.25}$$

$L_n(x, y)$: luminance normalisée au point (x, y) et calculée par :

$$L_n(x,y) = \frac{L(x,y) - \min(L)}{\max(L) - \min(L)} \tag{3.26}$$

Où : h(x, y) et L(x, y**)** sont respectivement la pseudo-teinte et la luminance

Min ( ), Max ( ) sont respectivement les minimas et les maximas calculés sur toute l'image. Il est à noter que cette normalisation est indispensable pour donner à la pseudo-teinte et à la luminance des valeurs comparables.

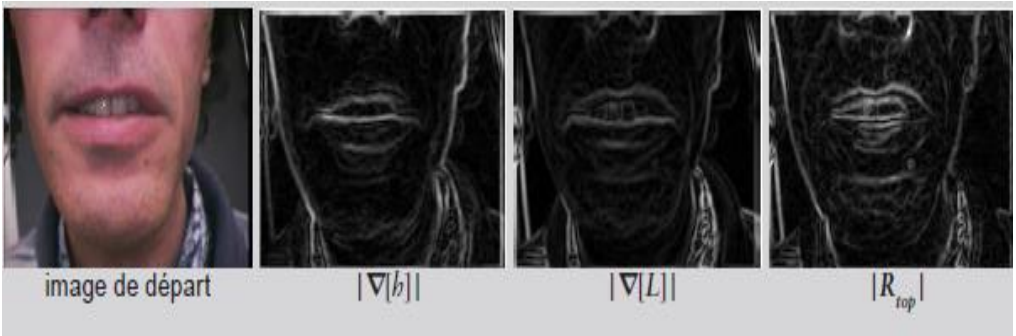


Figure 3. 7 : Comparaisons de différents types de gradients pour la localisation du contour supérieur de la bouche [19].

Le gradient hybride permet une bonne détection de la frontière supérieure des lèvres par rapport au gradient de la luminance ou de la pseudo- teinte pris séparément.

3.6 Modélisation des contours de la bouche

En ce qui concerne la segmentation labiale, les contours actifs sont largement utilisés car ils proposent d'importantes propriétés de déformation et un résultat de segmentation réaliste. Ils permettent également une implémentation facile et une convergence rapide, ce qui les rend particulièrement efficace pour des applications de suivi des contours des lèvres.

3.6.1 Différentes formes de snacks utilisées pour la segmentation labiale

La bouche est une caractéristique faciale hautement déformable et la première étape consiste à choisir un modèle paramétrique suffisamment flexible pour représenter fidèlement les contours des lèvres quelle que soit la forme visible dans l'image. Les lèvres peuvent prendre de multiples configurations qui sont autant de formes différentes [90].

Le choix d'un modèle est un compromis entre la complexité algorithmique et la déformabilité. Pour obtenir un grand nombre de degrés de liberté (et donc un choix plus vaste de formes possibles), il faut soit utiliser un grand nombre de courbes, soit utiliser des courbes polynomiales d'ordre élevé. Les deux solutions entraînent une augmentation du nombre de paramètres qui risque de rendre lente et difficile la minimisation de l'énergie. A

l'opposé, la convergence sera très rapide si le modèle n'est composé que de 2 ou 3 courbes simples. Mais dans ce cas, la rigidité du modèle rendra la segmentation approximative. Un grand nombre de modèles de bouche a déjà été proposé dans la littérature. On peut distinguer 3 modèles principaux.

3.6.1.1 Modèle asymétrique

C'est un modèle symétrique à deux (2) paraboles [91]. Les courbes les plus utilisées pour la conception des modèles du contour intérieur et extérieur des lèvres sont sans aucun doute les paraboles. Les coordonnées (x, y) d'une parabole peuvent être calculées par l'équation 3.27, où h est la hauteur de la courbe et w la largeur (figure 3.8):

$$y = h\left(1 - \frac{x^2}{w^2}\right) \qquad (3.27)$$

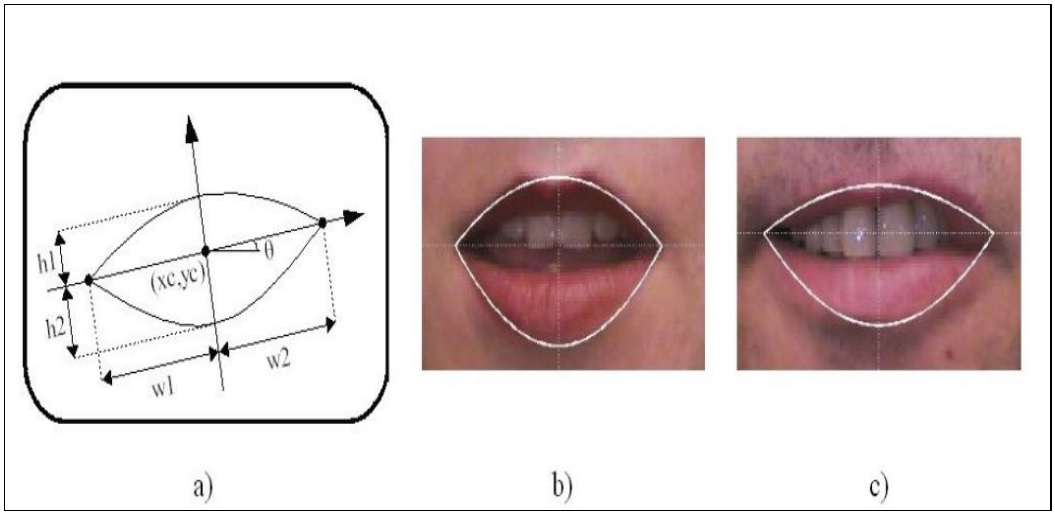


Figure 3.8. a) Modèle extérieur à 2 paraboles ;
b) et c) exemples de résultat de forme de lèvres [42]

3.6.1.2 Modèle quartique

C'est un modèle symétrique à 3 paraboles, aplati et tordu [92**],** (figure 3.9). Le premier modèle extérieur, proposé par Yuille *et al.* [93], utilise des quartiques. Les coordonnées (x, y) d'une quartique peuvent être calculées avec l'équation 3.28.

$$y = h\left(1 - \frac{x^2}{w^2}\right) + 4q\left(\frac{x^4}{w^4} - \frac{x^2}{w^2}\right) \qquad (3.28)$$

Où *h* est la hauteur de la courbe, *w* la largeur et le paramètre *q* contrôle la dérivée de la quartique.

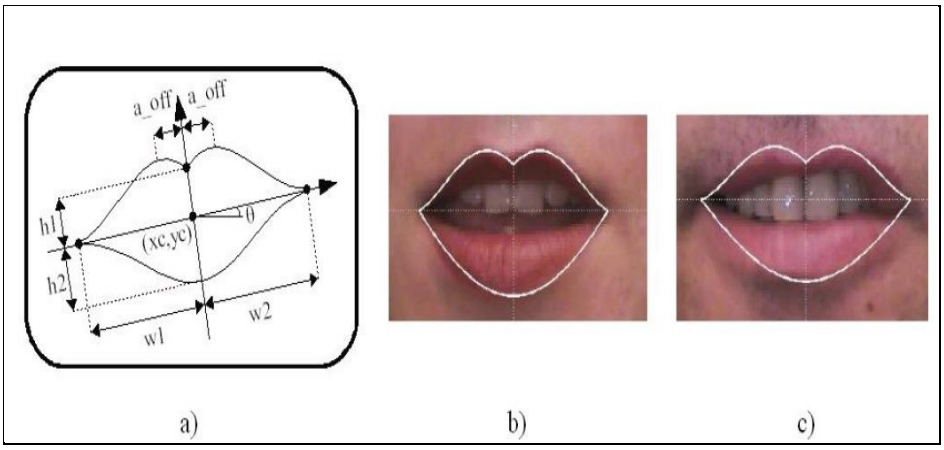


Figure 3.9. a) Modèle extérieur à 3 quartiques ;
b) et c) exemples de résultat de forme de lèvres [42]

3.6.1.3 Modèle cubique

C'est un modèle composé de quatre cubiques reliant six PC. Certains chercheurs utilisent quatre courbes cubiques et une ligne brisée pour relier six points clefs et modéliser le contour extérieur des lèvres (Figure 3.10), [19,10].

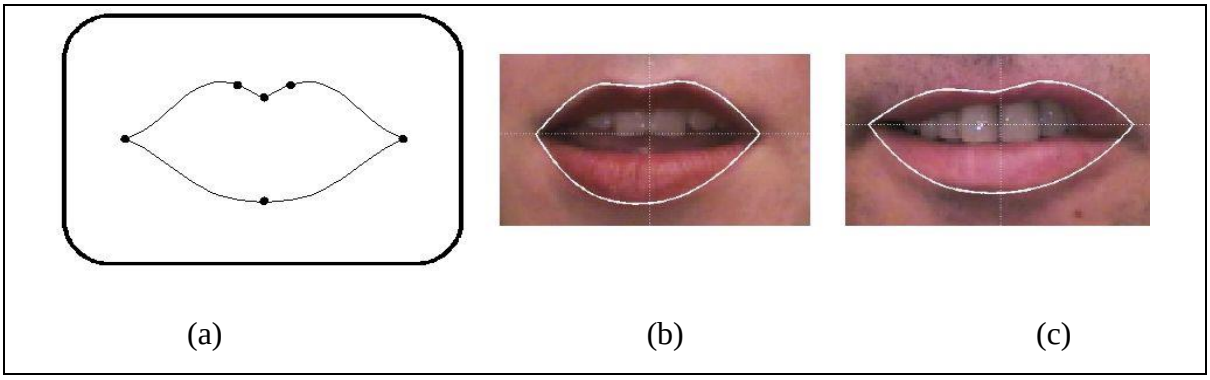


Figure 3.10. a) Modèle extérieur à 4 cubiques,
b) et c) exemples de résultat de forme de lèvres [42]

Le modèle est suffisamment flexible pour représenter des formes de lèvres très variables et la ligne brisée est une description fidèle de la forme en «V» de l'arc de Cupidon. Bouvier *et al.* utilisent le même modèle, à l'exception du contour extérieur bas qui est remplacé par deux courbes de Bézier [7].

Depuis l'introduction des CA par Kass et al. différents modèles de snack ont été proposés pour améliorer leur convergence. Dans le cadre de la segmentation des lèvres par le modèle déformable analytique nous présentons quelques résultats des différents travaux, utilisant trois cas de modèles de segmentation de lèvres. [57].

Le premier modèle est constitué de 2 paraboles (figure 3.11)

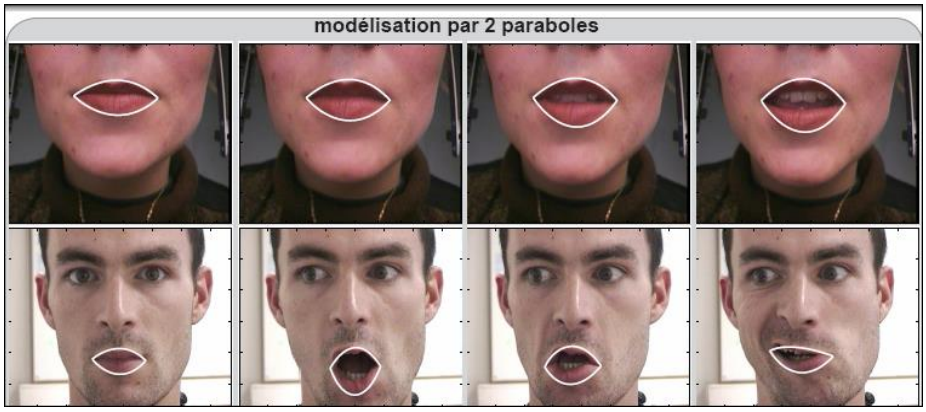


Figure 3.11 : Segmentations réalisées en utilisant un modèle à deux paraboles [19]

Le deuxième modèle est un peu complexe et représente le contour par 3 quartiques. Il est présenté à la figure 3.12.

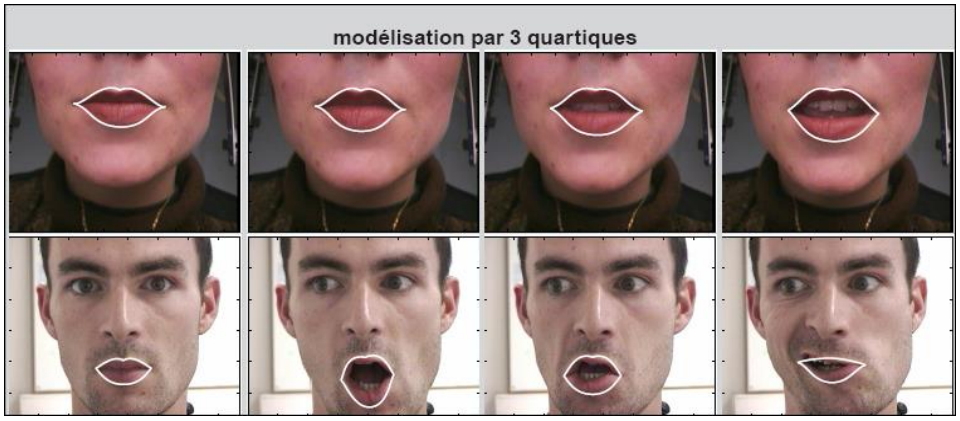


Figure 3.12 : Segmentations réalisées en utilisant un modèle constitué de quartiques [19].

Le troisième modèle est le modèle cubique qui apporte une amélioration très nette de la précision par rapport aux modèles des lèvres présentés précédemment. Nous exposons quelques résultats de la segmentation des lèvres extérieures utilisant le modèle cubique.

Figure 3.13 : Segmentations réalisées en utilisant un modèle cubique [19]

3.6.2 Autres résultats de segmentation des lèvres

Nous proposons les résultats d'une segmentation des lèvres en mouvements par les Contours Actifs Hybrides (CAH) et par l'algorithme Greedy.

3.6.2.1 Segmentation par les CAH

Les CAH combinent l'approche contour-région pour segmenter les lèvres en statique et en dynamique [94].

Il existe deux grandes approches par contours actifs. On peut alors distinguer les contours actifs basés contours où les fonctionnelles introduites sont composées de termes sur les contours. Ce sont les premiers modèles introduits dans la littérature. Puis ont été introduits les contours actifs basés régions où les fonctionnelles à minimiser sont composées des combinaisons linéaires de termes basés contours et d'autres termes sur les régions concernées.

Afin d'améliorer les performances de la segmentation, et l'intégration d'informations sur le contour et la région, certains auteurs utilisent les contours actifs. Dans [95] Lankton et al. proposent un modèle hybride pour la segmentation des images. Ils ont combiné le modèle géodésique basé sur le gradient de l'image et la méthode de Chan-Vese [96]. Ils présentent un flux basé sur une nouvelle fonctionnelle d'énergie qui est capable de produire des segmentations robustes et précises des images. Cette approche est une hybridation des contours actifs géodésiques locaux et plus globalement des contours actifs

basés régions. Ils présentent également une nouvelle dérivation mathématique utilisée pour mettre en œuvre cette approche.

On prend l'équation d'énergie des contours actifs géodésiques :

$$E = \oint_{C(s)} f(I)ds \tag{3.29}$$

Où la fonction *f* doit être positive et décroissante. On choisit *f* vérifiant certaines conditions et utilise des informations sur des régions locales c'est-à-dire le voisinage de chaque point, et cela pour la rendre similaire à l'énergie basé région [96], On trouve :

$$E = \oint_{C(s)} \int_{x\in\Omega} (I\chi(x,s) - u_\ell(s))^2 + \int_{x\in\overline{\Omega}} (I\chi(x,s) - v_\ell(s))^2 \, ds \tag{3.30}$$

Où :

- s : est l'abscisse curviligne.
- $u_l(s)$ et $v_l(s)$: sont les moyennes arithmétiques de l'intensité des points dans le voisinage local intérieur et extérieur du point C(s).
- X : une fonction qui prend la valeur « 1 » pour les points dans un voisinage autour du point C(s) et zéro ailleurs. Ce voisinage est divisé en deux par le contour (voisinage intérieur et voisinage extérieur).

$$X(\mathrm{x},\mathrm{s}) = \begin{cases} 1 \; avec \; x \in B(C(s)) \\ 0 \qquad\qquad Ailleurs \end{cases} \tag{3.31}$$

Sachant que B(C(s)) est un disque autour du point C(s).

- Ω : L'ensemble des points à l'intérieur du contour.

Avec les moyennes locales :

$$u_\ell(s) = \frac{S_{I_\ell}(s)}{A_{I_\ell}(s)} \qquad v_\ell(s) = \frac{S_{E_\ell}(s)}{A_{E_\ell}(s)} \tag{3.32}$$

Où $S_I(s)$ et $S_E(s)$ sont respectivement les sommes de l'intensité du voisinage intérieur et extérieur autour du point *C(s)*.

$$S_{I_\ell}(s) = \int_{x\in\Omega} I\chi(x,s)dA \qquad S_{E_\ell}(s) = \int_{x\in\overline{\Omega}} I\chi(x,s)dA \tag{3.33}$$

$A_I(s)$ et $A_E(s)$ sont respectivement les aires de voisinage intérieur et extérieur autour du point *C(s)*.

$$A_{I_\ell}(s) = \int_{x\in\Omega} \chi(x,s)dA \qquad A_{E_\ell}(s) = \int_{x\in\overline{\Omega}} \chi(x,s)dA \tag{3.34}$$

La figure 3.14 montre la façon de représenter le voisinage d'un point, à l'intérieur et à l'extérieur du contour :

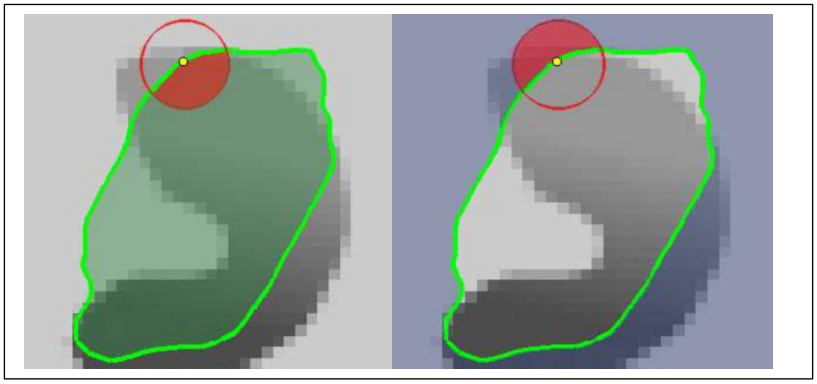


Figure 3.14 Diagramme montrant le voisinage local. Le disque représente le voisinage X [95].

Les CAH montrent une amélioration remarquable pour la détection des lèvres en mouvement (figure 3.15).

Figure 3.15 : Segmentation dynamique des lèvres par les CAH

3.6.2.2 Segmentation par l'algorithme Greedy

Pour la détection des contours extérieurs et intérieurs des lèvres, plusieurs travaux utilisent la position du snack extérieur pour initialiser le snack intérieur. Lorsque l'objectif est de suivre les contours des lèvres dans une séquence vidéo, l'initialisation du snack dans l'image courante est réalisée en utilisant des informations sur la forme des lèvres obtenues

dans l'image précédente. Par exemple la position du CA final de l'image précédente est utilisée directement comme initialisation dans l'image suivante [97]; [89].

Nous rappelons la fonctionnelle du CA utilisé pour la segmentation :

$$E^*_{Snake} = \alpha E_{Cont} + \beta E_{Cour} + \gamma E_{Intensité} + \delta E_{Ballon} + \varepsilon E_{Distance} \quad (3.35)$$

$\alpha, \beta, \gamma, \delta$, et ε sont des paramètres attribués à chaque énergie de la fonctionnelle du CA. Le réglage de ces paramètres varie d'une image à une autre.

Dans ce contexte, nous présentons le résultat d'une segmentation avec les lèvres en mouvement par des CA classiques (snack) en utilisant l'algorithme Greedy. En premier lieu une segmentation externe et interne des lèvres en statique a été effectuée, puis une segmentation dynamique pour le suivi des lèvres.

Nous constatons pour une segmentation dynamique que les résultats obtenus sont très réalistes (figure 3.16), car ils correspondent à une parfaite segmentation externe et interne des lèvres.

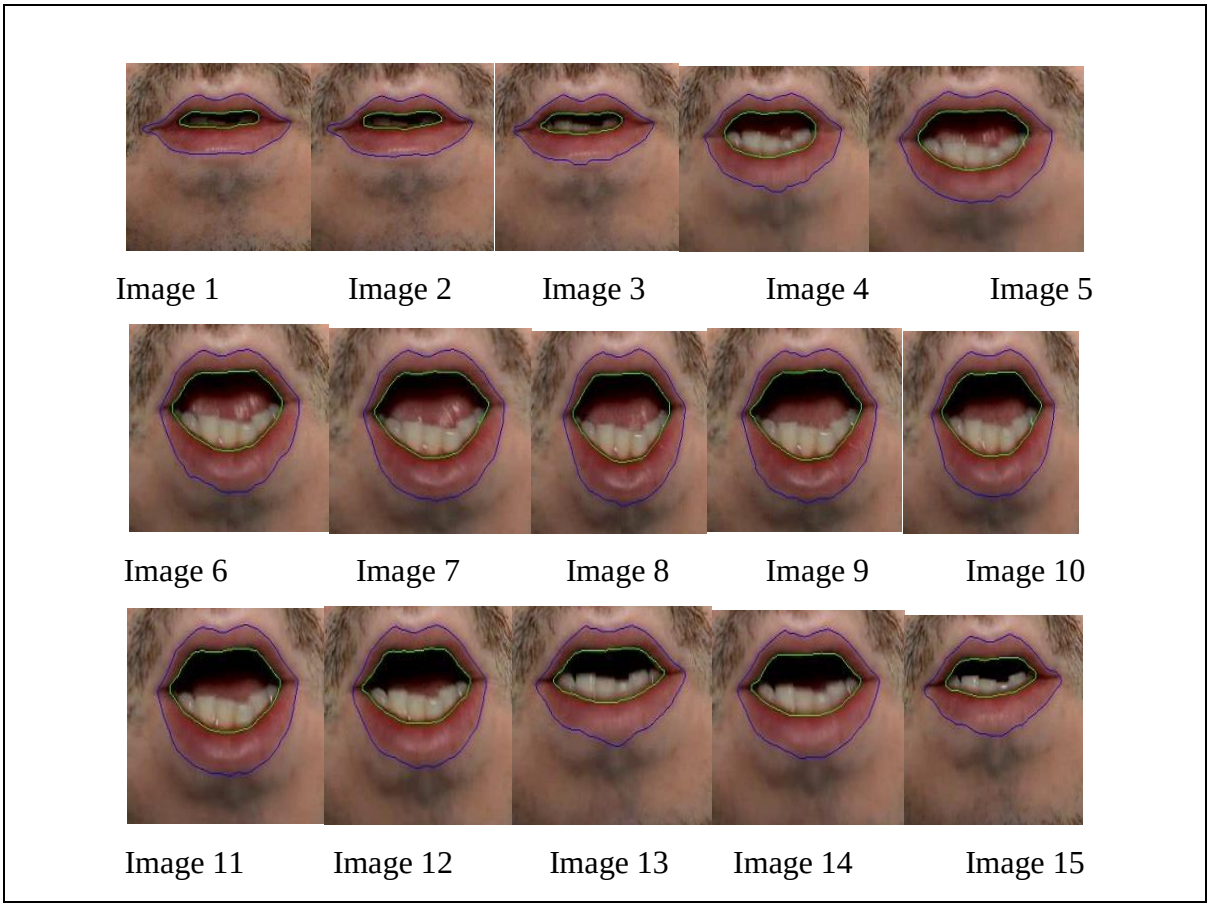


Figure 3.16**:** Résultats de la segmentation externe et interne des lèvres sur une séquence d'images

Sur cette séquence d'images, les deux contours actifs (externe et interne) ont bien suivi l'évolution des vrais contours des lèvres, malgré la présence de bruits et de moustache, de plus la lumière n'est pas distribuée uniformément (les parties hautes des images sont plus éclairées que les parties basses).

Nous affichons dans le tableau 3.1 les valeurs des différents paramètres permettant la convergence des deux snacks vers les lèvres internes et externes.

Tableau 3.1 **:** Les différents paramètres des snacks utilisés pour segmenter la séquence

Teinte utilisée	Pseudo-Teinte + lissage avec le filtre Median 3x3 (x4)					
	Alpha	Beta	Gamma	Delta	Epsilon	Max_iteration
Snack externe	0.01	3	- 4.5	3	- 0.13	200
Snack interne	0.01	1	- 1	- 4		45

3.7. Conclusion

Au cours de ce chapitre, nous avons exposé les différentes méthodes de segmentation, puis nous avons présenté les contours actifs classiques, implémentés grâce à l'algorithme glouton « Greedy » et des énergies plus puissantes et plus avantageuses utilisées dans la segmentation des images. Nous avons présenté les résultats d'un ensemble de méthodes permettant de modéliser précisément la zone de la bouche avec la meilleure robustesse possible. Nous avons aussi présenté les résultats d'une méthode fiable qui permettent une segmentation fidèle des contours externes et internes de la bouche.

Une approche combinée région-contour est introduite dans le but d'obtenir une segmentation de la bouche. Ainsi, nous avons montré l'application des CA dans la segmentation des lèvres.

CHAPITRE 4 :
Segmentation Statique et Dynamique des Lèvres

4.1. Introduction

Nous avons vu dans les trois chapitres précédents l'essentiel de notions et de techniques auxquelles nous aurons recours durant le reste de notre travail. Dans ce chapitre nous abordons le problème de la saillance des contours des lèvres, pour cela nous proposons la méthode de pente optimale pour augmenter la robustesse de la modélisation des contours de la bouche, surtout dans l'optique de la lecture labiale. Nous traitons les deux approches de la segmentation statique et dynamique des lèvres. Pour l'approche de segmentation statique, nous utilisons les caractéristiques des lèvres et la méthode du contour actif pour la recherche des Points Caractéristiques (PC) ; pour la segmentation dynamique, nous appliquons la méthode de la mise en correspondance puis le recalage de ces PC.

4.2. Modèle choisi

Nous avons opté pour notre application d'utiliser le modèle cubique grâce à sa simplicité dans la mise en œuvre, et sa flexibilité de déformation sur les formes des lèvres **[19]**.

En effet, Le modèle cubique que nous avons choisi est composé de (Figure 4.1) :

- Six points caractéristiques :
  - deux points de commissures P1 et P5.
  - trois points hauts P2, P3 et P4 formant le cupidon.
  - point bas P6.
- Et quatre cubiques, nous notons les cubiques comme suit : (cub1: entre les points P1 et P2, cub2 : entre P4 et P5, cub3 : entre P5 et P6, cub4 : entre P6 et P1).

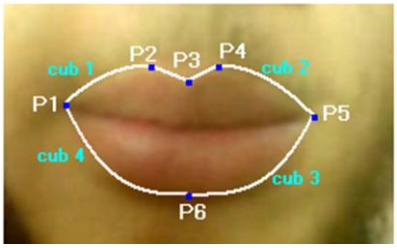


Figure 4.1 : Modèle cubique à six PC et quatre cubiques

4.3. Segmentation statique

Les contours actifs (CA) ou snacks, ont prouvé leur efficacité dans de nombreux problèmes de segmentation. Depuis leur apparition, de nombreuses améliorations ont été proposées dans la littérature. Mais aucune n'a vraiment réglé les problèmes de l'initialisation qui doit être faite suffisamment près du contour final et du réglage des paramètres basé sur des essais successifs et rarement réutilisable sur d'autres images [64]. La méthode présentée ici permet, dans une certaine mesure, de pallier à ces inconvénients. Notre travail consiste à déterminer les contours dans une séquence d'images dynamiques. Pour cela, nous déterminons les PC de la forme appropriée de la bouche en état statique; en seconde étape, nous construisons le contour de l'image prédéfinie. Enfin, nous déterminons les contours des lèvres de l'image dynamique.

4.3.1 Détermination des PC

Les PC donnent des indices importants sur la forme des lèvres. Ils sont utilisés comme points d'ancrage pour la construction du modèle paramétrique. Dans ce qui suit, nous présentons les méthodes qui nécessitent la détermination des 6 PC.

4.3.1.1 Détermination du point P3

Pour déterminer les PC, nous avons proposé de sélectionner un rectangle qui entoure la forme des lèvres, ensuite nous plaçons le point P3 manuellement dans la bonne position, (figure 4.2). Nous déterminons ensuite le reste des points par différentes méthodes. Nous commençons d'abord par les points hauts qui constituent le cupidon supérieur, ceux-ci peuvent être déterminés par une nouvelle méthode de snack, qui est le « jumping snack », dans la partie suivante nous allons voir le principe général de cette méthode et comment déterminer les points P2 et P4.

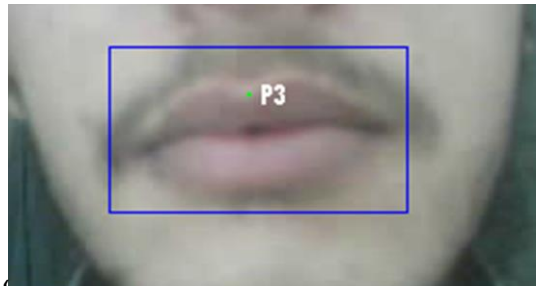


Figure.4.2 : Début du processus de segmentation sélection d'un rectangle et un PC P3.

4.3.1.2 Détermination des points P2, P4

Pour localiser Les points hauts de cupidon supérieur de la bouche, nous utilisons un nouveau type de contour actif que nous appelons «jumping snack» (ou «serpent bondissant) car sa convergence est une succession de phases de croissance et de saut [98]. Le principe général de jumping snack est d'initialiser un point P3 entre P2 et P4 en le plaçant convenablement au-dessus de la bouche. Le jumping snack grandit ensuite à partir de ce point jusqu'à ce qu'il atteigne un nombre prédéterminé de points (le nombre d'itérations). Cette phase de croissance est assez comparable au growing snack proposé par Berger dans le sens où le contour est initialisé avec un seul point et est progressivement prolongé à chacune de ses extrémités [73]. Durant cette phase des points terminaux sont ajoutés aux extrémités droite et gauche qui sont situées à une distance horizontale constante Δ, et aussi nous réduisons la zone de recherche dans un intervalle angulaire [θ_{inf}, θ_{sup}]. Les points terminaux qui seront choisis sont notés M_{i+1} et M_{-i-1} points qui ont un flux maximum du gradient hybride à travers les deux segments $M_i M_{i+1}$, $M_{-i} M_{-i-1}$.

Les deux flux moyens sont définis par :

$$\Phi_{i+1} = \frac{\int_{Mi}^{Mi+1} Rtop \, .dn}{|MiMi+1|} \quad (4.1)$$

$$\Phi_{-i-1} = \frac{\int_{M-i-1}^{M-i} Rtop.dn}{|M-iM-i-1|} \quad (4.2)$$

Où dn est la normale au segment

La figure suivante montre la manière de propagation de jumping snack.

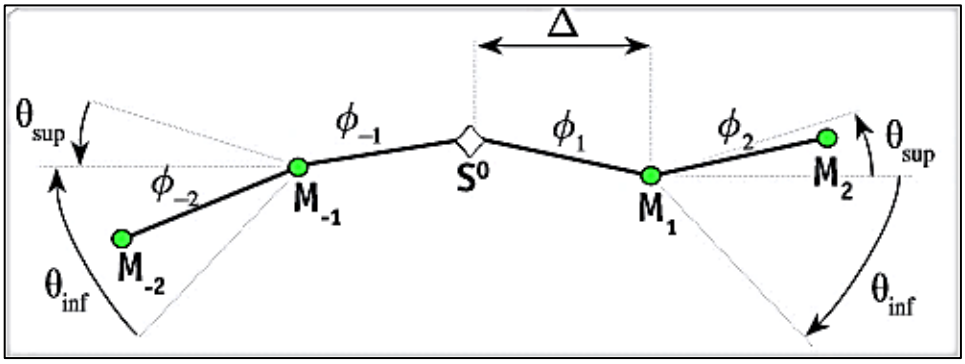


Figure 4.3 : Propagation de jumping snack (S^0 équivalent à P3)

Pour que le jumping snack converge sur le cupidon supérieur de la lèvre au premier temps (itération 0), nous allons forcer le jumping snack de monter vers le haut, ceci se fait par la

diminution de la plage de recherche [θ_{inf}, θ_{sup}] ; nous admettons pour la plage de recherche seulement deux angles : ($\theta_{inf}=\frac{\pi}{9}$, $\theta_{sup}=\frac{2\pi}{9}$).

Dans l'itération 1 aussi nous divisons la plage [0,$\frac{2\pi}{9}$)] en quatre angles égaux.la plage de recherche devient donc : [$\frac{\pi}{18},\frac{2\pi}{18},\frac{3\pi}{18},\frac{4\pi}{18}$].

Par la suite, dans les autres itérations on admet la propagation vers le haut et vers le bas. Et dans ce cas, la plage de recherche devient 8 valeurs égales distribuées dans l'intervalle suivant : [$\frac{-4\pi}{24},\frac{\pi}{3}$], après des tests on peut prendre la plage [$\frac{-4\pi}{15},\frac{\pi}{3}$] qui est aussi convenable.

D'après notre test sur l'échantillon des images qu'on a utilisé, alors le nombre d'itérations nécessaires pour converger les deux snacks (à gauche et à droite de P3) est environ de 6 à 7 itérations. Et la distance horizontale delta (Δ) est environ de 4 à 7 pixels.

Comme les deux points P2 et P4 sont positionnés symétriquement par rapport à l'axe qui passe par le point P3, alors il suffit de propager un seul snack, soit le snack qui est situé à droite ou à gauche de P3 À la fin de ce processus, le snack va positionner sur le cupidon supérieur de la lèvre, c'est évident que le point P4 se trouve dans le haut de snack c'est - à - dire que le point P4 ayant une plus petite valeur de y. Dans la figure 4.4, nous représentons la propagation du snack de droite. À partir d'une position P3 (point vert) le snack droit se propage vers le haut selon le paramètre de saut Δ=5 et le nombre d'itérations 6. Le point P4 se trouve dans le haut du snack. P2 est trouvé par symétrie de P4.

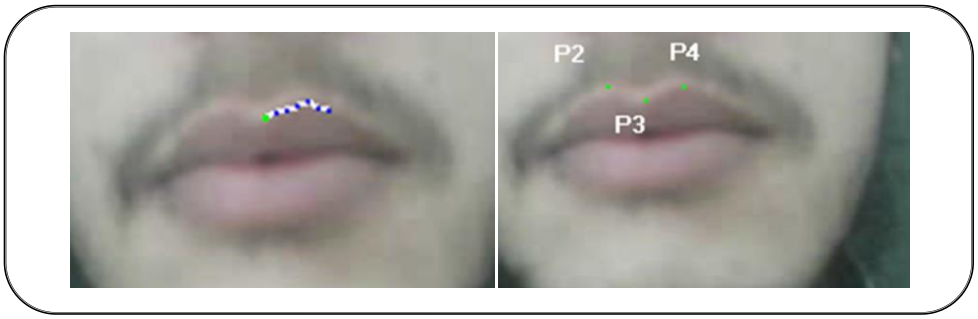


Figure.4.4 : Détermination des points P2 et P4

4.3.1.3 Détermination des points de commissures P1 et P5

Les zones les plus sombres du plan de luminance apparaissent au niveau des commissures de la bouche ainsi à l'intérieur de la bouche, qu'elle soit ouverte ou fermée. Il apparaît donc intéressant de déterminer pour chaque colonne de l'image la position dont le niveau de gris est minimum. Un chaînage des pixels les plus sombres donne le minimum de luminance [4]. La figure 4.5 représente le minimum de luminance calculé sur toute l'image.

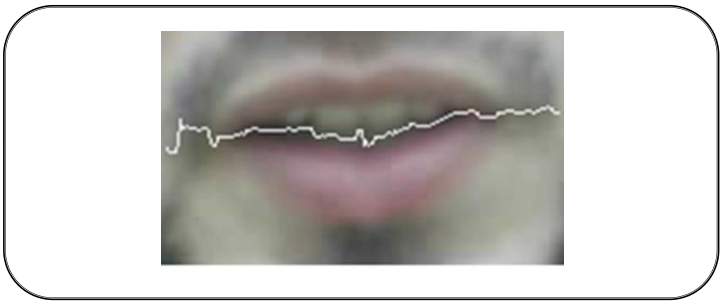


Figure 4.5 Représentation de minimas de luminance (L_{min}) sur toute l'image

Pour éviter la recherche sur toute la colonne de la zone sélectionnée au début de processus de segmentation, nous avons proposé de diviser horizontalement cette zone en 3 rectangles égaux ; comme il est évident que les points intéressants (P1 et P5) sont situés dans le rectangle qui est au milieu [10].

Jusqu' à maintenant nous avons déterminé la zone qui contient les deux commissures (rectangle 2). Pour une meilleur localisation des deux PC, nous avons proposé de diviser aussi le rectangle 2 en trois rectangles : l'un sur la droite (pour P5) et l'autre sur la gauche (pour P1). La largeur de chaque rectangle extrême est environ de 15 à 20 pixels (figure 4.6).

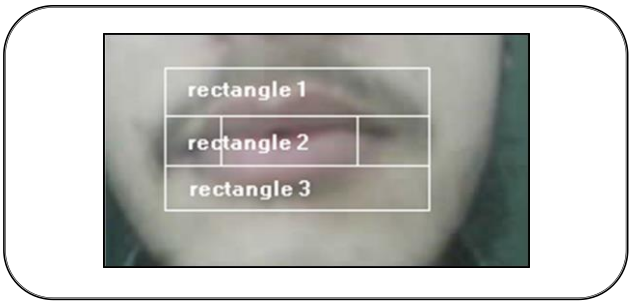


Figure 4.6 : Localisation des points P1 et P5

Ensuite, nous calculons le minimum de luminance (Lmin) sur les deux petits rectangles (figure 4.7)

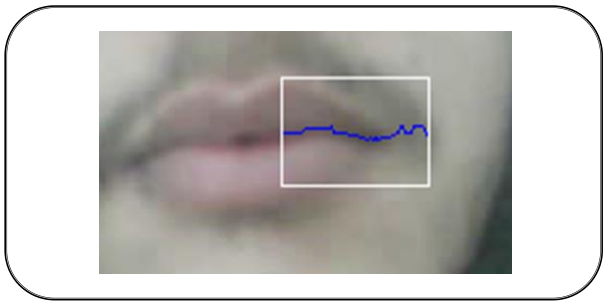


Figure.4.7 : Minima de luminance LminD (ligne en bleu dans rectangle blanc)

L'estimation de points P1 et P5 se fait par la maximisation de flux moyen de gradient hybride à travers les courbes supérieures ; et la pseudo teinte sur les courbes inférieures [4].

Pour estimer les PC P1 et P5, nous nous intéressons dans notre algorithme seulement aux courbes situées en haut. En effet, pour trouver P5 (respectivement P1), nous parcourons le minimum de luminance droite LminD (respectivement LminG), et à chaque fois nous associons chaque point de LminD (respectivement LminG) une courbe supérieure reliant ce point par le point P4 (respectivement P2).

Le point P5 (respectivement P1) est le point qui maximise le flux moyen de gradient hybride à travers ces courbes. La figure 4.8 représente le déroulement de la recherche du point P5 (en bleu) en maximisant le flux moyen du gradient hybride, chaque fois en examinant un candidat de courbes(en blanc) à travers LminD (en vert).

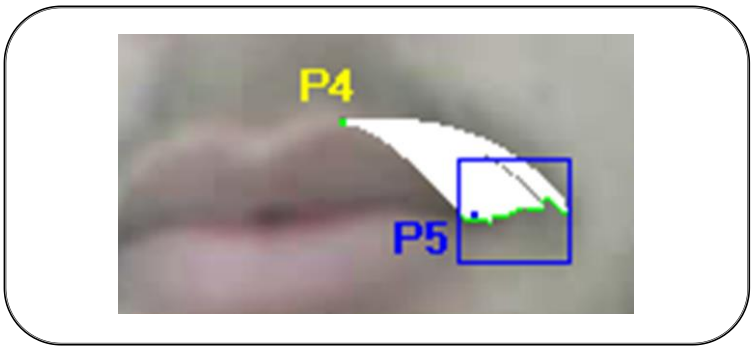


Figure 4.8 : Détermination du point P5

4.3.1.4 Détermination du point bas P6

Pour déterminer P6, nous réduisons la zone de recherche par un rectangle R0 de hauteur environ 15 pixels à partir du bas vers le haut, et de largeur environ de 15 pixels, telle que la ligne verticale qui passe par le point P3 passe au centre de ce rectangle (figure 4.9).

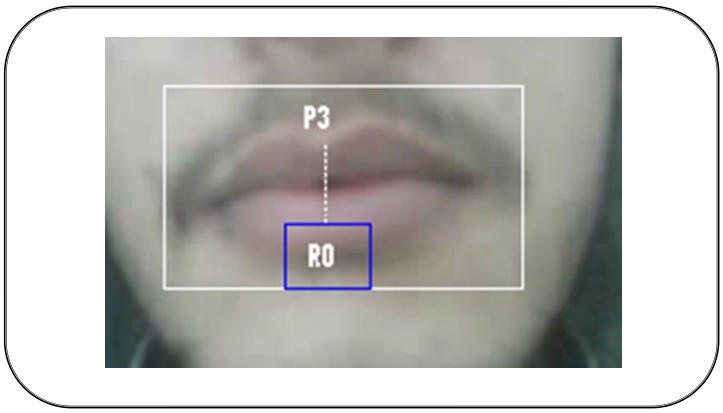


Figure.4.9 : Détermination du point P6 par sélection d'un rectangle (en bleu)

Dans le rectangle R0 de l'image I, nous cherchons la ligne i correspondant à la valeur maximale de la somme suivante :

$$S = \sum_{j=1}^{N} (Tab [j][i]) \quad (4.3)$$

Tel que : i, j : ligne et colonne courante.

N : nombre de lignes de l'image I.

Tab : un tableau correspondant aux valeurs de la pseudo-teinte de l'image I.

La ligne qui correspond à la valeur maximale de la somme S, sa position désigne l'ordonnée du point P6. L'abscisse de P6 est la même que celle de P3.

Finalement le résultat de recherche des six PC est donné par la figure 4.10.

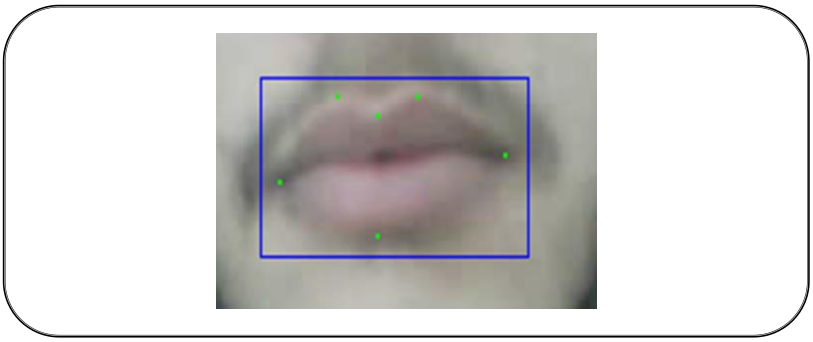


Figure 4.10 : Résultat de recherche des six PC

4.3.2 Tracé du contour final de lèvres

Après détermination des six PC, il reste donc à relier ces points par des cubiques qui approchent bien le contour des lèvres. Pour cela nous proposons l'implémentation de la méthode de la pente optimale pour le calcul de ces cubiques.

La pente correspond à l'inclinaison d'une surface ou d'une ligne par rapport à l'horizontale. Elle peut être mesurée selon un angle en degrés, radians ou grades ou encore en pourcentage, c'est- à- dire selon le rapport hauteur sur longueur multiplié par 100. La pente s'avère souvent utile dans bien des domaines, notamment et surtout dans le bâtiment (inclinaison des toits, positionnement des cellules photovoltaïques), inclinaison des routes, etc.

Nous détaillons maintenant le calcul de la pente optimale appliquée au tracé des contours des lèvres. Cette pente nous permet de tracer des cubiques qui segmentent convenablement les lèvres.

Soit P le point d'intersection de la ligne verticale qui passe par le point P2 et la ligne horizontale qui passe par P1 (figure 4.11).

Pour tracer la cubique entre le point P1 et P2 : soit le point M appartenant au segment [P1 P] tel que :

$$\overline{P1M} = 0.7\ \overline{P1P} \qquad (4.4)$$

Soit le point M1 le point d'intersection entre la ligne verticale qui passe par le point M et la ligne horizontale qui passe par le point P2.

β : est l'angle situé entre la ligne (P1M1) et la ligne (P1M).

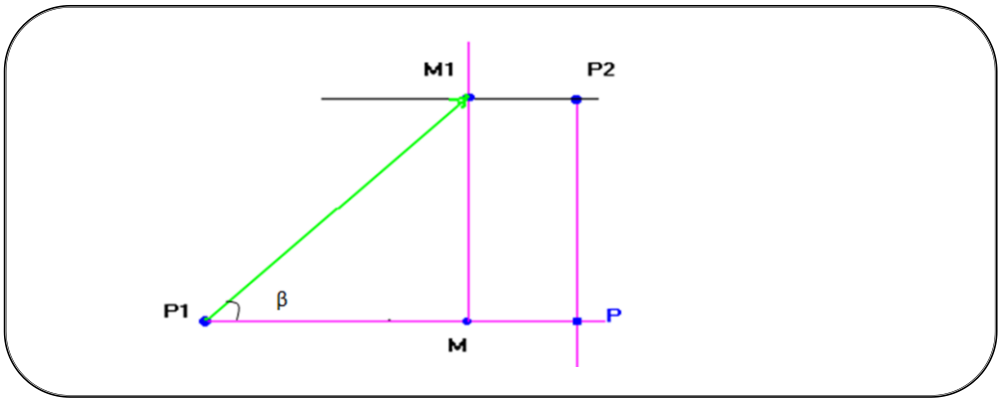


Figure 4.11 : Pente recherchée entre les deux lignes P1M1 et P1M

La pente correspond à la tangente de cet angle.

$$\text{Pente} = \text{tangente}(\beta) = \frac{\overline{\text{M1M}}}{\overline{\text{P1M}}} \qquad (4.5)$$

À partir de cette pente, nous traçons une cubique qui est Cub1 entre les deux points P1 et P2. Cette cubique a pour asymptote oblique P1M1 et M1P2 comme asymptote horizontale.

Même procédé pour les cubiques inférieures sauf que nous utilisons l'égalité suivante:

$$\overline{\text{P1M}} = 0.3\overline{\text{P1P}} \qquad (4.6)$$

Les autres cubiques sont calculées par le même procédé qui est déjà implémenté pour la cubique un (Cub1). Cependant, les constructions des cubiques dépendent des positions des PC (P1, ……, P6). Alors, nous avons besoin de déterminer la bonne position de ces points. La figure 4.12 montre le tracé de contour final des lèvres.

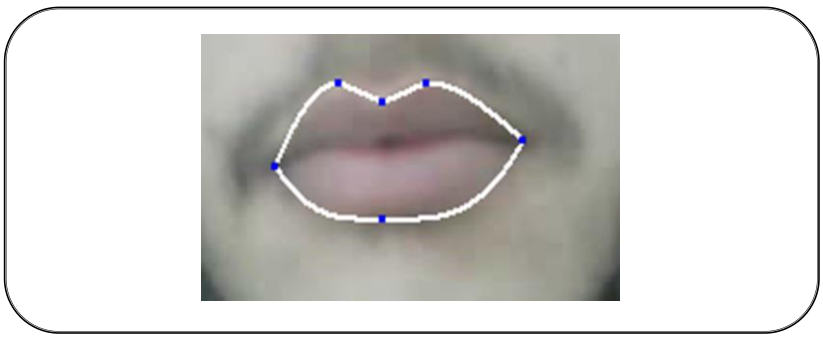


Figure 4.12 : Calcul des PC et tracé des cubiques selon la méthode de la pente optimale

Nous résumons notre algorithme de segmentation statique dans la figure 4.13.

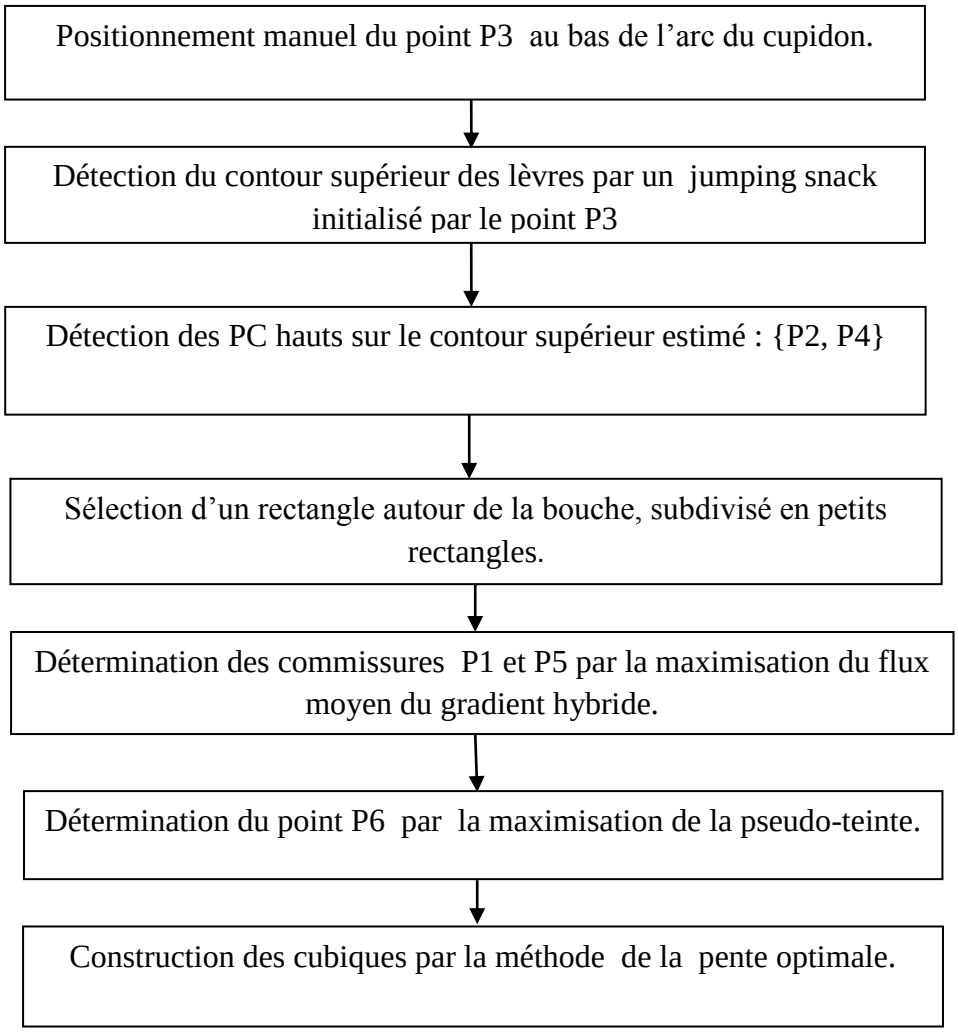


Figure 4.13 : Résumé de notre algorithme de segmentation statique

4.4 Segmentation dynamique

La méthode de segmentation que nous proposons comporte deux étapes principales : l'initialisation et le suivi. Dans l'étape d'initialisation, le contour était détecté en utilisant une segmentation statique. Lors du suivi, les résultats obtenus dans les images fournissent des informations supplémentaires susceptibles de rendre la segmentation dynamique robuste.

Comme nous avons proposé d'utiliser le modèle formé de six points caractérisant les lèvres et quatre cubiques reliant ces points et deux segments reliant le cupidon supérieur. Pour faire suivre le mouvement des formes de lèvres il suffit de suivre les six PC (P1,… ,P6), et les quatre cubiques (cub1,… ,cub4) [11,12].

4.4.1 Suivi des PC

L'estimation de mouvement est un problème fondamental en traitement d'image appliqué à des séquences vidéo. Dans ce domaine, de très nombreuses méthodes ont été proposées. Parmi lesquels les méthodes de mise en correspondance (block-matching techniques) tentent d'estimer le mouvement d'une région de l'image courante en minimisant la distance avec une région candidate de l'image suivante. En général, cette mesure de similarité est obtenue par une somme des différences inter-pixels élevée au carré.

4.4.2 Algorithme de mise en correspondance

Dans cette méthode, nous supposons que le voisinage du point suivi dans l'image I_t se retrouve dans l'image suivante I_{t+1} par une translation(ou un déplacement d(x)) [62].

$$I_t(x) = I_{t+1}\big(x - d(x)\big) \qquad (4.7)$$

Où d(x) est le vecteur déplacement du pixel de coordonnée x (x est un vecteur dans R^2).

Considérons un voisinage R de taille N×N dans l'image de référence prise au temps t. Le but est de retrouver dans l'image suivante la région la plus ressemblante à R. Nous notons $I_t(x)$ et $I_{t+1}(x)$ les valeurs des niveaux de gris dans ces 2 images. Pour cela, il faut minimiser une fonction coût égale à la somme des différences inter-pixels au carré :

$$\varepsilon\big(d(x)\big) = \sum_{x\epsilon R}\big[I_t(x) - I_{t+1}\big(x - d(x)\big)\big]^2 . w(x) \qquad (4.8)$$

où w(x) est une fonction de pondération. En général, w(x) est constant et vaut 1.

Dans toute la suite de cette section, nous supposons que le voisinage considéré ne subit pas de déformation. Par conséquent, la valeur du déplacement est la même pour tous les pixels de R. Comme il est évident qu'un test exhaustif de toutes les régions possibles est très coûteux en temps de calcul, alors nous restreignons la recherche uniquement sur le voisinage de R, soit S cette région du voisinage.

4.4.2.1. Algorithme 1 utilisant tous les points de la région S

Pour rechercher la région qui ressemble à R, nous faisons un test exhaustif sur tous les points de la région S (région prise de l'image I à l'instant t+1). I_t et I_{t+1} ce sont des niveaux de gris de deux images successives.

4.4.2.2 Application de l'algorithme 1 sur un point de S

L'algorithme un suit généralement bien le mouvement de points. Pour minimiser le temps de recherche, nous avons proposé d'utiliser une collection de points de la région S. Dans l'algorithme utilisant un échantillon de points, nous allons voir comment restreindre les tests sur un candidat de points de S.

4.4.2.3. Algorithme 2 utilisant un échantillon de points

Il est évident que la recherche sur tous les points de S prend beaucoup de temps. Cependant, on a besoin d'exclure les points ayant une grande différence au niveau de gris par rapport au PC recherché. Pour cela, nous prenons le PC comme référence de la région R (généralement le centre de R), puis nous recherchons les points qui ressemblent à ce point dans la région S. La validation de ressemblance se fait par le calcul de différence inter-pixel au carré de chaque point de S avec le centre de R.

Ajouté ce point à l'échantillon, lorsque la différence est inférieure à une valeur donnée (ex. : ε=0.0001). Finalement, on applique l'algorithme un (1) sur l'échantillon trouvé.

Dans la figure 4.14 on présente l'estimation de la Position d'un Point en Mouvement : (a) Position d'un Point dans l'image à l'instant t (■), (b) Estimation de l'échantillon des points (□) et (c) Position Correspondante du Point dans l'Image Suivante (■), (ε =0,003).

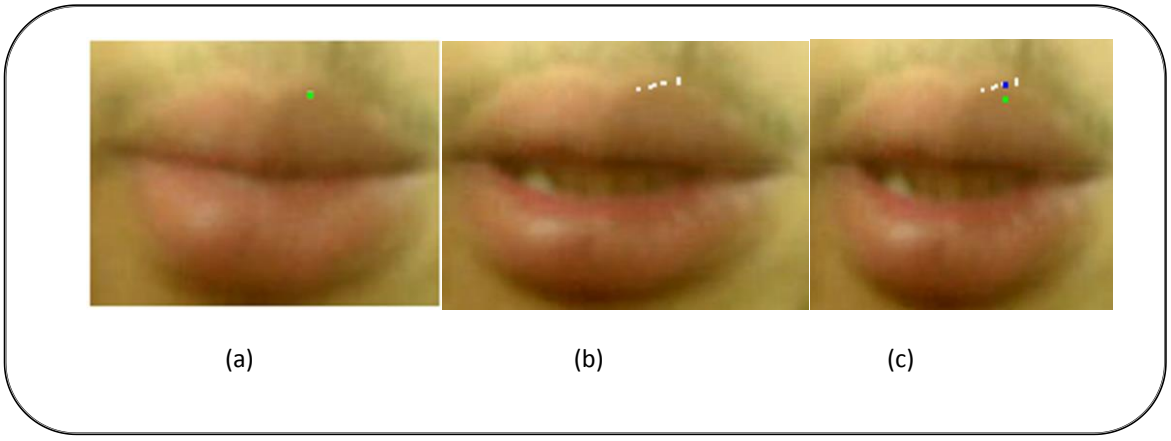


Figure 4.14 : Estimation de la Position d'un Point en Mouvement

Cet algorithme suit bien le mouvement des points, prend moins de temps de recherche, mais tout cela n'est pas suffisant pour valider une estimation, ceci revient au choix du seuil de ressemblance, si celle-ci est plus petit, alors il peut exclure des points ayant une région qui ressemble bien à la région R, aussi s'il est de valeur près de 1, alors il peut

ralentir la recherche. D'après notre test sur notre algorithme, le bon choix d'erreur, c'est entre [0.003-0.007].

Dans la vidéo de la figure 4.15, nous remarquons bien que le mouvement des PC dans les premières images est généralement bien par rapport à celui des dernières images avec un décalage de certains points à leurs positions désirées, cela revient à l'accumulation des erreurs ou le changement brusque de luminance, etc.

Figure 4. 15 : PC et estimation de leurs Mouvements dans une Séquence Vidéo (ε =0.005).

En effet, le suivi par le calcul des erreurs n'est pas suffisant pour valider une estimation notamment dans une vidéo, donc nous avons besoin de trouver une méthode de rectification des mauvaises estimations. Comme nous ne connaissons pas la position où l'estimation n'est pas bonne, donc nous faisons réajuster les PC pour chaque dizaine d'itérations. Dans la partie suivante, nous allons voir comment réajuster les positions des points.

4.4.3 Recalage des points

À première vue, les deux algorithmes semblent fournir des résultats corrects d'une image à la suivante. En réalité, les estimations de positions ne sont pas parfaites et l'accumulation progressive des erreurs mène souvent, après quelques images, à des résultats très imprécis.

4.4.3. 1 Recalage des points hauts P2, P4

Pour ajuster la position des points hauts et bas, nous détectons les contours supérieurs des lèvres en utilisant les snacks ouverts. Les CA sont des courbes ν définies paramétriquement par ν(s) = (x(s), y(s)), (où s est l'abscisse curviligne) qui peuvent se déformer progressivement de manière à s'approcher au plus près des contours d'un objet.

Cette déformation est guidée par la minimisation d'une fonctionnelle d'énergie comprenant une énergie interne permettant de régulariser le contour et une énergie externe attirant le snack vers les contours. La résolution des équations régissant le comportement du snack conduit à la relation dynamique suivante :

$$v_i = (A + yI)^{-1}(yv_{i-1} + F_{ext}(v_{i-1})) \quad (4.9)$$

Où v_i et v_{i-1} sont les positions du snack aux itérations i et i-1 respectives.

Le coefficient y est appelé coefficient d'amortissement et contrôle la vitesse de déplacement du snack.

I : matrice identité

F_{ext} : Force extérieure s'appliquant sur le CA.

Pour le recalage des points P2 et P4, le snack doit être attiré par le contour supérieur des lèvres. La force extérieure qui s'applique sur le snack du haut dérive donc du gradient hybride:

$$F_{ext} = \nabla(|gradien\_hypride|^2 \quad (4.10)$$

A est la matrice de rigidité, elle est de taille Ns×Ns, où Ns est le nombre de points du snack, il est fonction des coefficients d'élasticité et de courbure α et β. Or, comme le réglage de ces coefficients est un des points délicats des CA.

Pour simplifier le problème, nous utilisons un snack sans force intérieure. Cela conduit à une relation dynamique plus simple ne comportant pas de matrice de rigidité :

$$v_i = v_{i-1} + \frac{1}{y}F_{ext}(v_{i-1}) \quad (4.11)$$

On n'a autorisé que les déplacements haut et bas de points de snack.

Comme n'existe plus les forces internes de contour (énergie interne) ou autres forces externes, alors le paramètre y n'a aucune influence sur la propagation de snack on peut l'éliminer (on pose $y = 1$).

Et comme il n'existe plus de contrainte d'élasticité ni de courbure, les snacks obtenus sont bruités et irréguliers. Mais notre but n'est pas d'extraire les contours en entier. Nous avons seulement besoin de les estimer dans les voisinages de P'2(t), P'4(t)

Les points qui se situent sur le cupidon supérieur de lèvre ayant une forte valeur de gradient hybride donc, le point recherché doit vérifier cette condition.

4.4.3.2 Recalage de P3

Nous supposons que les deux points P2 et P4 sont bien recalés. Le point central haut P3 est trouvé en supposant qu'il est situé équidistant des points P2 et P4 (figure 4.18). Nous testons donc une dizaine de positions le long de la médiatrice de P2 P4. Le meilleur candidat maximise le flux moyen de gradient hybride Rtop à travers la ligne brisée [P2, P3, P4] [**19].**

4.4.3.3 Recalage de P1 et P5

La position des commissures ne peut pas être ajustée en utilisant les CA, car elles sont généralement situées dans des zones de faible gradient. Nous considérons quelques commissures possibles au voisinage de P'1 et P'5, le long de la ligne des minima de luminance Lmin (figure 4.16). Il s'agit ensuite d'associer à chacun de ces candidats un couple de courbes approchant au mieux le contour des lèvres en utilisant la méthode de la pente optimale [10].

Les meilleures commissures maximisent le flux moyen de gradient hybride à travers les courbes supérieures (les cubiques un et deux) des lèvres et minimisent le gradient de la pseudo teinte pour les courbes inférieures (les cubiques trois et quatre) [4].

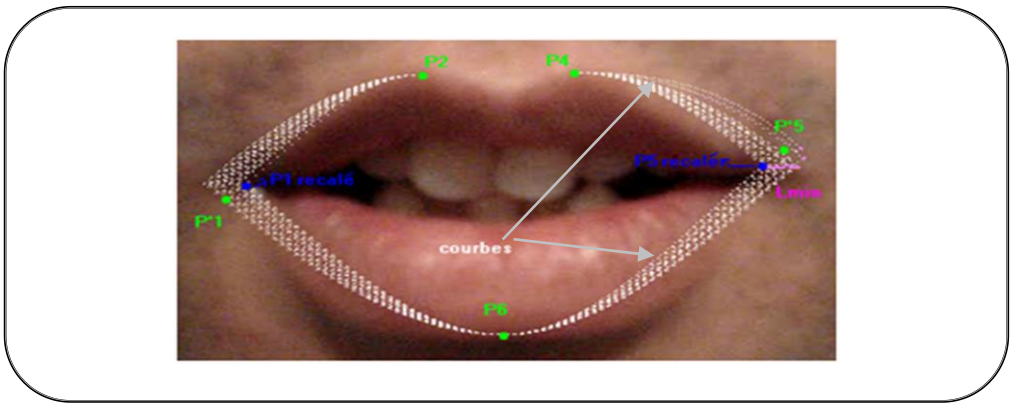


Figure 4.16 : Représentation du recalage des points P1 et P5

4.4.3.4 Recalage de P6

La position du P6 est généralement située dans des zones de fort gradient de la pseudo teinte. Alors, le recalage de P6 se fait comme dans le cas statique. Nous testons un candidat de lignes au voisinage de P6 (nous prenons généralement un rectangle contenant P6)). L'ordonné y de P6 c'est la ligne qui maximise le flux moyen de gradient pseudo teinte.

Dans le cas général P3 et P6 apparaissent sur la même colonne .Alors ils ont la même abscisse x.

Dans la figure 4.17, on présente le déroulement du test d'un candidat de lignes (lignes blanches) au voisinage de P6 (point rouge). Le point P6 recalé (point magenta) situé sur la ligne qui maximise le pseudo teinte (ligne verte)

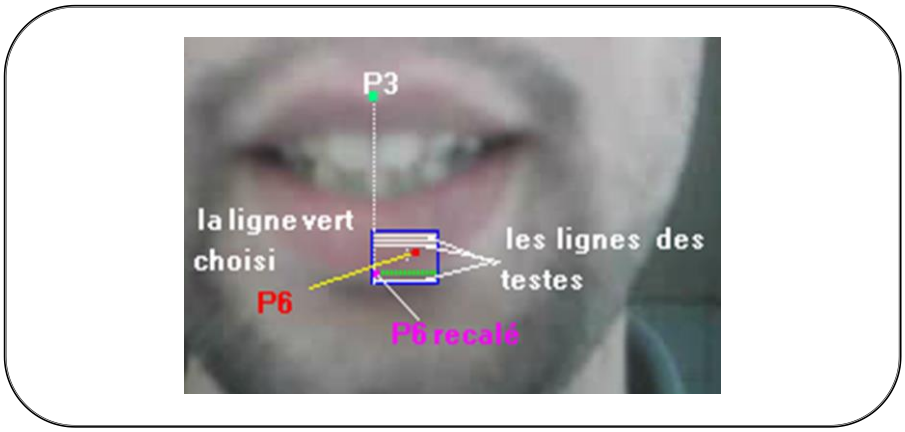


Figure 4.17 : Test d'un candidat de lignes au voisinage de P6

La méthode de recalage que nous avons implémentée précédemment est l'une des estimations qui donne des résultats généralement précis.

La figure 4.18 représente le suivi des mouvements des 6 points avec recalages seulement sur les points P3, P1 et P5, cela après chaque huit itération.

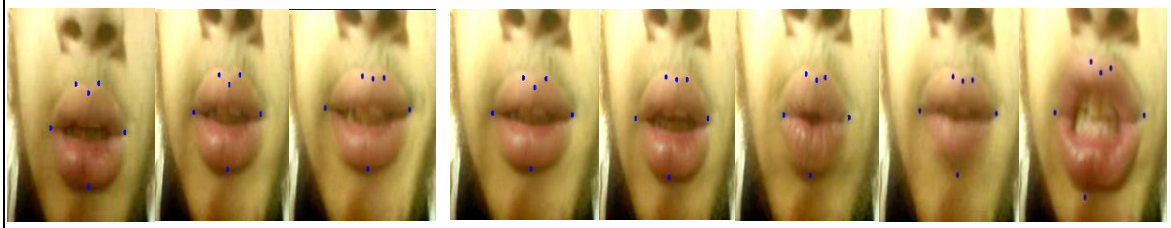


Figure 4. 18: Suivi des six PC avec recalages des points P1, P3 et P5.

Nous constatons que le point P6 à quitter sa position correcte car il n'est pas soumis au recalage. Et comme nous avons constaté précédemment le calcul des erreurs par la méthode de mise en correspondance n'est pas suffisant pour valider une estimation dans une vidéo. Le point P6 devait être soumis à un recalage.

Nous avons vu comment estimer les mouvements des PC et leurs recalages. Il reste donc de relier ces points par des courbes qui contournent bien les lèvres. Comme nous avons déjà vu dans l'état statique, la création des courbes. Dans la partie suivante, nous allons voir comment estimer les mouvements de snacks (courbes).

4.4.4 Les mouvements des snacks

Pour déterminer la position des snacks à l'instant t, nous avons besoin d'utiliser la segmentation réalisée dans l'image de la figure 4.12. Le modèle de lèvres obtenu à l'image t-1 est étiré de manière à coïncider avec les PC estimés. Pour cela, nous appliquons une déformation linéaire aux cubiques de l'image t-1, notées γ k(t-1) (k allant de 1 à 4), le déplacement de chacun de leur point est obtenu par une moyenne pondérée des déplacements des 2 PC associés.

Par exemple, les déplacements de P1 et P2 sont utilisés pour calculer le vecteur déplacement de chaque point Q de γ1(t-1) :

$$d_Q = d_{p_1}\left(1 - \frac{|p_1(i-1)Q(i-1)|}{|p_1(i-1)p_2(i-1)|}\right) + d_{p_2}\left(1 - \frac{|p_2(i-1)Q(i-1)|}{|p_1(i-1)p_2(i-1)|}\right) \qquad 4.12$$

Où d_Q, d_{P1}et d_{P2} sont les vecteurs déplacements des points Q, P1 etP2 respectivement :

$$\begin{cases} d_Q = \overrightarrow{Q(i\text{-}1)Q(i)} \\ dp_1 = \overrightarrow{P_1(i\text{-}1)P_1'(i)} \\ dp_2 = \overrightarrow{P_2(i\text{-}1)P_2'(i)} \end{cases} \qquad (4.13)$$

i : représente le temps.

Cette transformation permet d'obtenir la courbe déformée γ1(t) (figure 4.19).

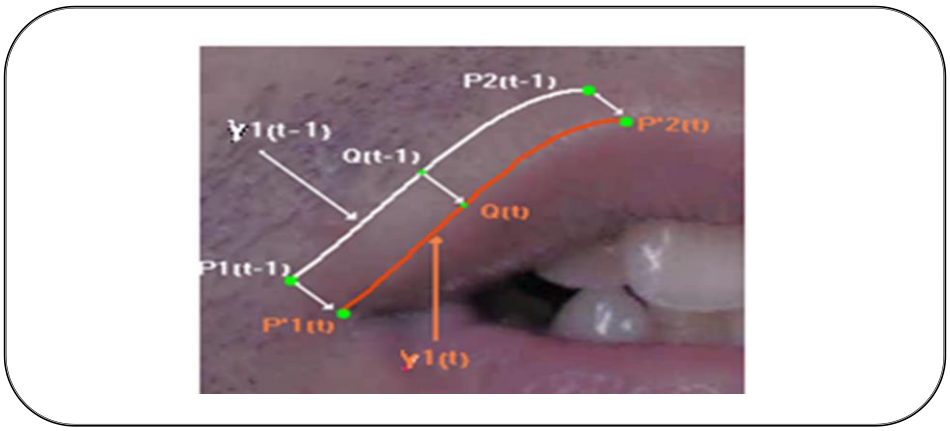


Figure 4. 19 : Estimation de la position de la cubique γ1(t-1) à un instant t

Dans la figure 4.20, nous appliquons cette translation sur les quatre cubiques. Nous obtenons un modèle proche des contours des lèvres.

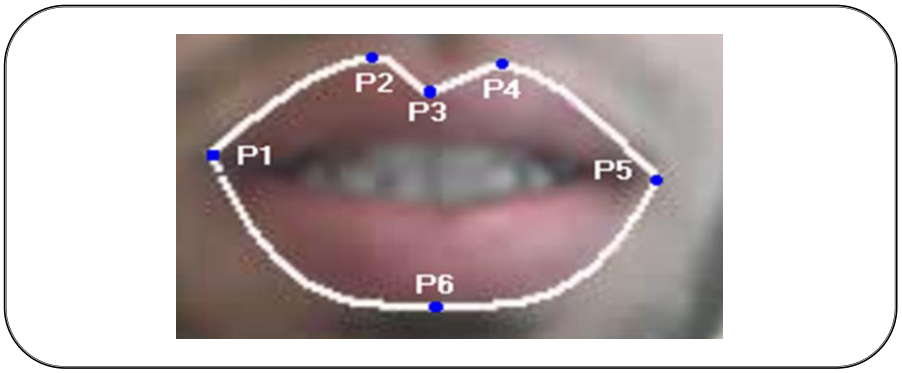


Figure 4. 20 : Application des déformations sur les quatre cubiques

Nous résumons l'algorithme de segmentation statique que nous avons proposé dans la figure 4.21.

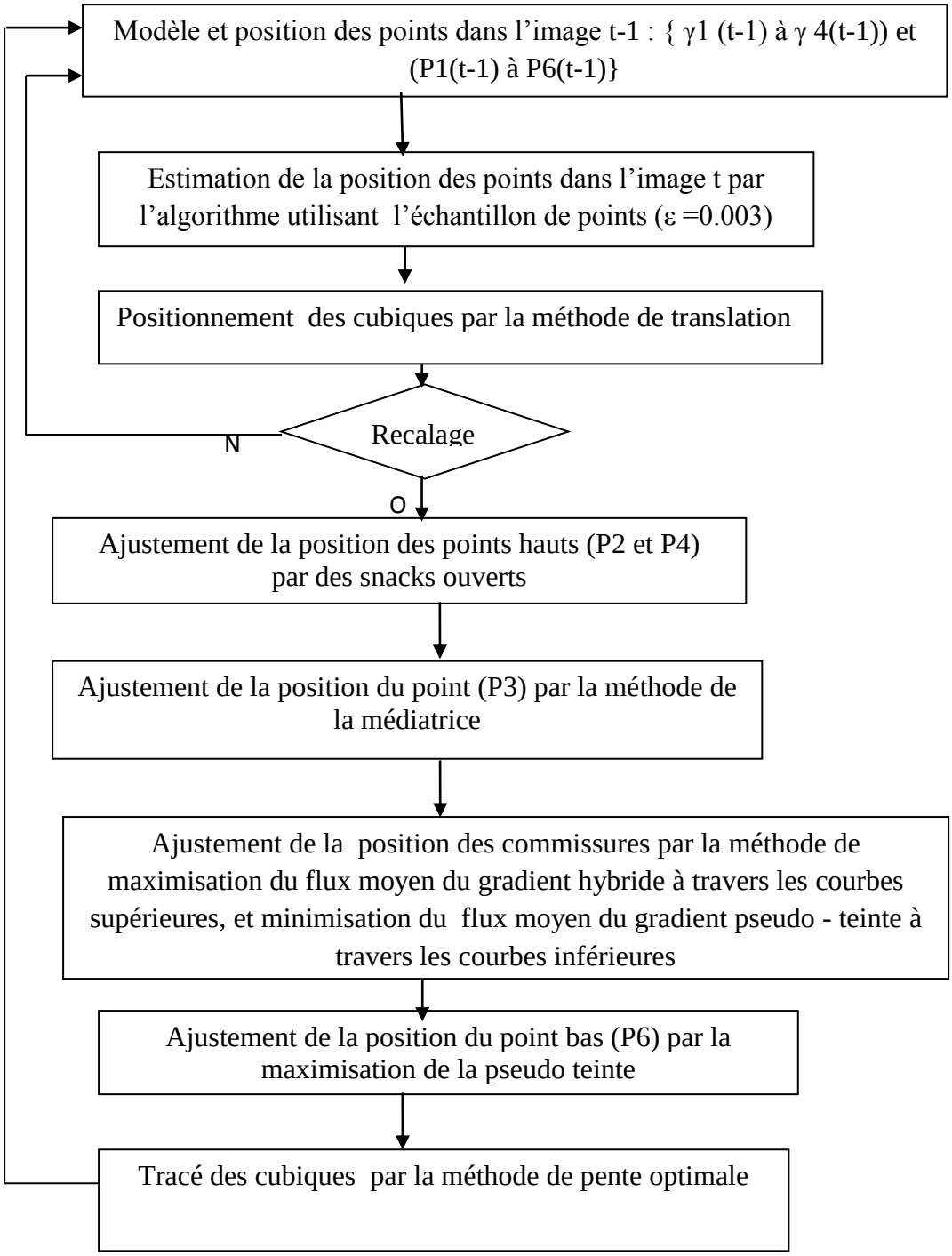


Figure 4.21 : Résumé de notre algorithme de segmentation dynamique

4.5 Calcul des paramètres labiaux

Après avoir déterminé les PC externes des lèvres, nous pouvons aisément calculer les paramètres labiaux (Ouverture, Etirement, Surface labiale) externe des lèvres [12].

- Ouverture externe, noté Oex, est déterminée à l'aide des points P3 et P6 :

$$O_{ex} = \text{Distance}_{\text{euclidienne}}(P3, P6) \quad (4.15)$$

- Etirement externe, noté Eex, est déterminée à l'aide des points P1 et P5 :

$$E_{ex} = Distance_{euclidienne}(P1, P5) \quad (4.16)$$

Après le calcul de l'ouverture et l'étirement des lèvres, nous devons calculer :

- Surface labiale (aire labiale) qui représente le nombre de pixels limité par le contour des lèvres extraites ;
- Temps d'exécution est le temps de calcul de gradient, du suivi, et de recalage. C'est un paramètre très important dans la lecture labiale, c'est le temps qui nous renseigne sur la segmentation d'une seule image de la séquence.

Les six PC ont une très grande utilité dans le domaine de la lecture labiale, car ils déterminent les paramètres labiaux nécessaires pour la reconnaissance et la restitution de la parole.

4.6 Discussion des résultats

Dans cette partie, nous présentons les résultats obtenus avec notre algorithme de segmentation et du suivi des mouvements des lèvres. Dans la segmentation statique, nous avons implémenté deux méthodes : la méthode semi-automatique et la méthode manuelle pour segmenter les lèvres. Dans la segmentation dynamique nous avons implémenté aussi les méthodes de Mise en Correspondances pour le suivi des PC. La minimisation des erreurs de suivi se fait à l'aide d'un algorithme de recalage afin de suivre convenablement les contours des lèvres. Pour s'assurer de la performance de nos algorithmes, nous avons utilisé aussi une initialisation manuelle des PC dans le but d'expertiser notre méthode. Les résultats obtenus ont montré la rigueur de nos algorithmes, et cela en comparant les

graphes obtenus par l'initialisation manuelle avec ceux de l'initialisation semi-automatique.

Nos résultats ont été obtenus dans un environnement matériel et logiciel tels qu'une caméra mobile de résolution 5 Méga pixels pour acquérir les vidéos films, un microordinateur de processeur Intel dual coré 1.46 GHZ avec une mémoire vive de 1 Go, équipé des logiciels suivants : Windows XP service pack2 (32 bits), le langage de programmation est le visuel C# 2008, et d'autres logiciels tels que le Directx et le Krypton. Notre application ne dépend pas de la taille de l'image, mais il faut prendre le temps en compte. La meilleure taille d'images pour traiter en temps réel : 320x240. Nous testons notre logiciel sur plusieurs séquences vidéos contenant les différents locuteurs, entre autres ceux possédant des barbes, et des moustaches, pour voir l'efficacité de nos algorithmes.

4.6.1 Méthode de la segmentation semi-automatique

Les différentes étapes de la détermination des PC et le suivi des contours des lèvres sont présentées dans la figure 4.22.

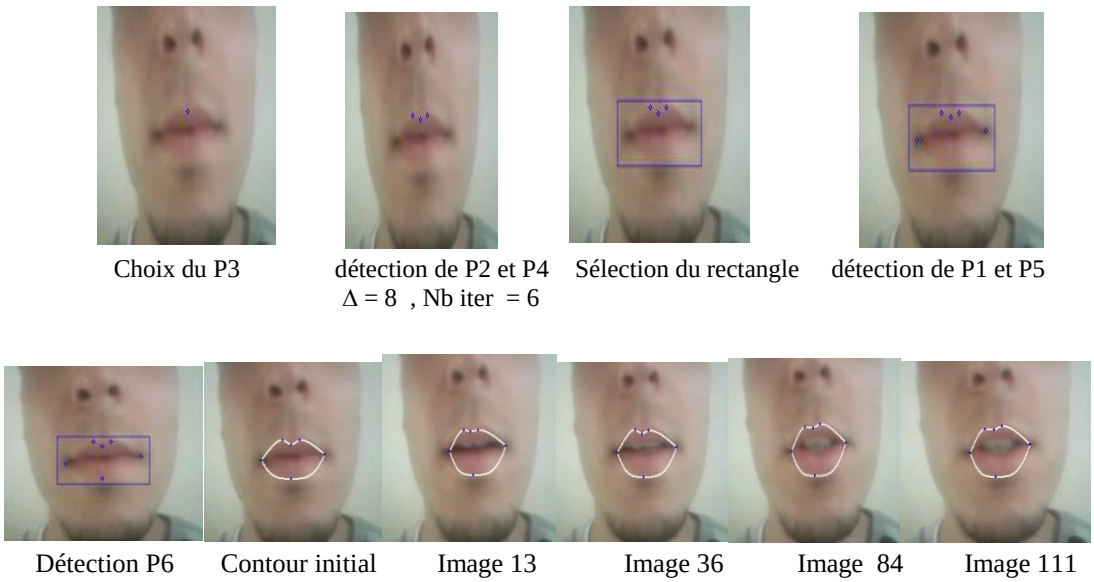


Figure 4.22 : Suivi des contours des lèvres par la méthode semi-automatique

Le tableau 4. 1 représente les paramètres labiaux, avec le temps d'exécution (temps de calcul du gradient et le suivi des lèvres en mouvement plus le recalage), et cela pour les différentes images de la séquence.

Tableau 4.1 : Paramètres labiaux avec le temps d'exécution pour chaque Image de la Séquence Vidéo.

	Images	Etirement	Openning	Surface	Execution Time
▶	0	112	44	4434	54
	1	111	46	4631	219
	2	111	47	4716	188
	3	110	48,01042	4714	165
	4	101	50,03998	4433	156
	5	101	50,01	4433	149
	6	101	51,0098	4507	165
	7	101	52,00961	4586	197
	8	101	57,14018	4706	201
	9	101	57,14018	4706	192
	10	101	57,14018	4708	175
	11	101	56,14268	4640	162
	12	102	55,22681	4584	159
	13	102	54,23099	4510	231
	14	103	52,23983	4340	129
	15	103	52,23983	4340	130
	16	102	47,26521	4035	172
	17	102	46,27094	3959	126
	18	102	46,17359	3944	133
	19	102	46,17359	3945	136
	20	105	57,21888	4850	163
	21	105	59,21149	5018	127
	22	106	59,13544	5073	125
	23	106	57,07889	4964	131

	Images	Etirement	Openning	Surface	Execution Time
	23	106	57,07889	4964	131
	24	108	55,14526	4956	181
	25	108	55,14526	4956	132
	26	108	54,14795	4878	125
	27	108	54,14795	4914	131
	28	107	54,14795	4864	128
	29	108	54,14795	4907	127
	30	109	54,23099	4943	126
	31	109	54,23099	4943	139
	32	109	54,08327	5017	173
	33	109	55,08176	5092	141
	34	108	56,0803	5126	127
	35	108	56,0803	5126	129
	36	108	56,0803	5112	276
	37	108	56,0803	5112	188
	38	108	55,08176	4966	157
	39	107	55,14526	4920	142
	40	107	54,23099	4900	183
	41	107	53,23533	4821	130
	42	107	52,23983	4743	125
	43	107	52,23983	4749	135
	44	108	53,23533	4807	128
	45	108	55,22681	4961	134
	46	107	56,22277	4989	131

	Images	Etirement	Openning	Surface	Execution Time
	46	107	56,22277	4989	131
	47	107	55,22681	4913	134
	48	104	55,3263	4777	170
	49	103	55,22681	4731	132
	50	102	55,22681	4681	149
	51	100	55,14526	4591	130
	52	98	55,14526	4491	137
	53	95	55,08176	4354	127
	54	92	55,03635	4218	133
	55	90	54,00926	4049	117
	56	102	55,03635	4577	175
	57	101	56,00893	4595	123
	58	101	55,03635	4529	120
	59	102	55,00909	4581	115
	60	102	55,00909	4580	119
	61	106	57,00877	4880	106
	62	106	57,00877	4880	124
	63	112	59,03389	5375	121
	64	110	57,03508	5261	189
	65	111	58,03447	5394	136
	66	112	59,03389	5528	156
	67	113	62,07254	5850	138
	68	114	62,07254	5896	135
	69	115	63,07139	6042	143

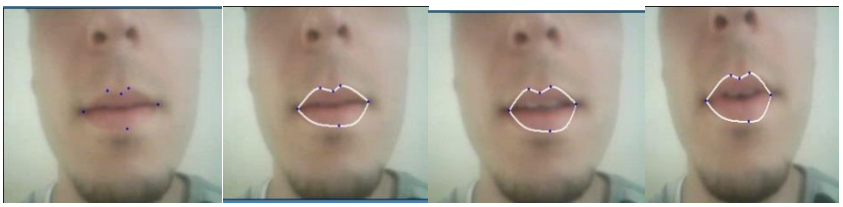
	Images	Etirement	Openning	Surface	Execution Time
	69	115	63,07139	6042	143
	70	116	63,12686	6073	148
	71	117	62,1289	6081	139
	72	109	57,21888	5199	198
	73	111	55,08176	5079	124
	74	110	56,14268	5131	135
	75	109	57,21888	5167	149
	76	107	59,21149	5232	158
	77	104	60,13319	5167	154
	78	104	60,13319	5167	150
	79	103	58,21512	4975	148
	80	105	56,14268	5022	194
	81	103	56,0803	4926	135
	82	101	57,03508	4855	128
	83	100	61,07373	5058	174
	84	100	61,07373	5058	227
	85	101	62,1289	5188	191
	86	103	61,13101	5203	177
	87	105	61,13101	5296	156
	88	99	62,20129	5107	206
	89	101	62,20129	5212	156
	90	101	62,20129	5212	164
	91	101	62,20129	5212	196
	92	101	62,1289	5221	160

	Images	Etirement	Openning	Surface	Execution Time
	92	101	62,1289	5221	160
	93	101	62,1289	5221	182
	94	101	62,1289	5222	169
	95	100	62,20129	5175	167
	96	97	58,30952	4829	211
	97	91	56,32051	4401	158
	98	89	56,22277	4361	151
	99	91	57,31492	4475	183
	100	91	58,54912	4554	189
	101	91	59,4138	4620	162
	102	92	60,53098	4692	162
	103	92	60,53098	4692	168
	104	92	62,39391	4723	219
	105	93	62,39391	4769	190
	106	94	62,39391	4774	190
	107	96	62,514	4884	178
	108	97	63,50591	4994	161
	109	97	63,50591	4994	163
	110	96	62,39391	4824	214
	111	95	60,53098	4657	343
	112	101	55,57877	4798	236
	113	101	56,56854	4871	166
	114	100	56,56854	4815	153
	115	100	56,56854	4815	202

Les extremums des trois paramètres labiaux et du temps d'exécution à l'intérieur du tableau, sont des grandeurs importantes pour la Lecture Labiale.

4.6.2 Méthode de la segmentation manuelle

Nous présentons dans la figure 4.23 les différentes étapes du suivi des contours des lèvres.



Sélection manuelle Contour initial Image 18 Image 39

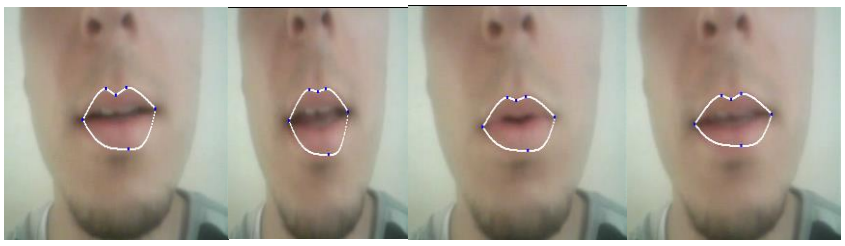


Image76 Image 89 Image108 Image114

Figure 4. 23 : Suivi des contours des lèvres par la méthode manuelle

Les résultats de l'étude sont présentés dans le tableau 4.2 contenant les paramètres labiaux (étirement, ouverture, et surface), avec le temps d'exécution (temps de calcul du gradient ajouté au suivi et au temps de recalage), et cela pour les différentes images de la séquence.

Tableau 4.2 : Paramètres Labiaux avec le Temps d'Exécution pour chaque Image de la Séquence Vidéo

Images	Etirement	Openning	Surface	Execution Time
0	116	48,37355	4828	34
1	115	51,47815	5056	373
2	115	52,46904	5117	160
3	114	52,46904	5074	173
4	106	51,08816	4711	130
5	106	51,08816	4711	129
6	106	52,03845	4773	123
7	106	53,03772	4858	1485
8	98	57,31492	4621	188
9	98	57,31492	4621	133
10	98	57,31492	4624	156
11	98	56,22277	4543	130
12	99	55,3263	4501	136
13	99	54,33231	4433	140
14	99	51,35173	4216	143
15	99	51,35173	4216	129
16	101	46,27094	4011	238
17	101	45,27693	3943	122
18	101	45,17743	3906	122
19	101	45,17743	3903	125
20	104	56,32051	4806	110
21	104	58,30952	4955	130
22	105	58,21512	5001	122
23	105	57,21888	4921	126
Images	Etirement	Openning	Surface	Execution Time
23	105	57,21888	4921	126
24	104	53,08484	4743	170
25	104	53,08484	4743	126
26	104	52,08647	4673	128
27	104	52,03845	4683	134
28	103	52,03845	4642	127
29	104	52,03845	4683	126
30	105	52,03845	4721	129
31	105	52,03845	4721	125
32	98	51,08816	4495	166
33	98	52,08647	4554	132
34	97	53,08484	4583	135
35	97	53,08484	4583	133
36	97	53,08484	4581	137
37	97	53,08484	4581	130
38	97	52,15362	4449	138
39	97	52,23983	4449	130
40	103	57,14018	4718	176
41	103	56,14268	4648	129
42	103	55,08176	4562	122
43	103	55,14526	4567	143
44	104	56,14268	4675	123
45	104	57,14018	4744	132
46	103	57,21888	4740	132

Images	Etirement	Openning	Surface	Execution Time
46	103	57,21888	4740	132
47	103	56,22277	4688	141
48	102	55,14526	4615	166
49	101	55,14526	4569	143
50	100	55,14526	4515	131
51	98	55,14526	4417	135
52	96	55,22681	4334	126
53	93	54,14795	4114	129
54	90	54,08327	3985	137
55	87	53,08484	3793	122
56	97	50,15974	4190	185
57	95	50,24938	4113	129
58	95	49,36598	4049	128
59	95	49,25444	4044	122
60	95	49,25444	4044	124
61	99	50,24938	4281	109
62	99	50,24938	4281	128
63	105	52,34501	4692	123
64	109	55,3263	4886	187
65	110	56,32051	5013	138
66	111	57,42822	5154	132
67	112	60,29925	5470	134
68	113	60,29925	5516	137
69	114	61,29437	5653	136

Images	Etirement	Openning	Surface	Execution Time
69	114	61,29437	5653	136
70	115	60,40695	5661	141
71	115	59,4138	5607	142
72	106	51,35173	4678	172
73	108	49,25444	4530	119
74	107	50,24938	4573	127
75	106	51,35173	4617	139
76	104	53,33854	4690	142
77	101	54,45181	4635	159
78	101	54,45181	4635	139
79	100	52,61179	4455	141
80	107	49,81967	4627	187
81	105	50,01	4539	128
82	103	50,01	4481	131
83	102	54,12947	4689	187
84	102	54,12947	4689	132
85	104	55,10898	4865	141
86	106	54,91812	4957	137
87	107	54,91812	5001	138
88	106	54,74486	4955	183
89	108	54,58938	5050	143
90	108	54,58938	5050	148
91	108	54,74486	5052	150
92	108	54,74486	5056	146

Images	Etirement	Openning	Surface	Execution Time
92	108	54,74486	5056	146
93	108	54,74486	5056	140
94	107	54,74486	5010	148
95	107	54,74486	5010	135
96	104	50,80354	4689	194
97	101	49,24429	4419	139
98	99	50,21952	4372	136
99	100	51,1957	4453	140
100	100	52,17279	4530	151
101	101	52,95281	4640	138
102	102	54,12947	4805	141
103	102	54,12947	4805	143
104	107	57,87055	5053	193
105	108	58,85576	5179	137
106	109	57,87055	5188	138
107	111	58,0517	5285	150
108	113	58,0517	5382	139
109	113	58,0517	5382	145
110	112	57,07013	5221	153
111	112	55,10898	5053	157
112	105	51,78803	4634	176
113	104	51,97115	4605	144
114	103	51,97115	4542	126
115	103	51,97115	4542	134

Les extremums des trois paramètres labiaux et du temps d'exécution à l'intérieur du tableau, sont des grandeurs importantes pour la Lecture Labiale.

4.6.3 Interprétation des résultats

Dans les figures (fig. 4.24, fig.4.25) suivantes, nous représentons la méthode de la segmentation manuelle par les courbes rouges et la méthode de la segmentation semi-automatique par les courbes bleues, et cela pour les quinze premières images de la séquence vidéo :

- l'étirement : nous constatons que les deux courbes sont pratiquement proches l'une de l'autre, et que la différence n'est pas significative, elle est de 3 à 5 pixels. Cette différence est due à l'emplacement approximatif de P1 et P5 dans l'image initiale de la méthode manuelle (figure 4.23). L'application de nos algorithmes semble montrer que la méthode semi-automatique se rapproche mieux de la réalité, c'est-à-dire elle est conforme à la méthode manuelle (figure 4.24 a) ;
- l'ouverture : nous observons un léger écart des courbes entre les images de 0 à 8, l'écart est de 1 à 5 pixels. Cela est dû à l'emplacement approximatif de P6 et de P3 dans l'image initiale de la méthode manuelle (figure 4.23). A partir de l'image 8, les points ont été recalés dans notre algorithme, et nous constatons une parfaite superposition des deux courbes (figure 4. 24 b).
- la surface : le choix manuel des PC est approximatif, c'est pour cela que l'écart en pixels entre les deux courbes est remarquable, par contre à partir de l'image 8 de la séquence, un recalage a été établi. Nous constatons alors une réduction de l'écart, et une certaine conformité entre les deux courbes, (figure 4.25 a).
- le temps d'exécution : nous constatons que les allures des deux courbes sont proches l'une de l'autre, sauf pour l'image 7 de la séquence où le recalage de certains points a nécessité un temps supplémentaire (figure 4. 25 b).

Nous pouvons conclure que, pour une initialisation manuelle très précise, l'écart entre les deux courbes sera réduit, et cela pour tous les paramètres.

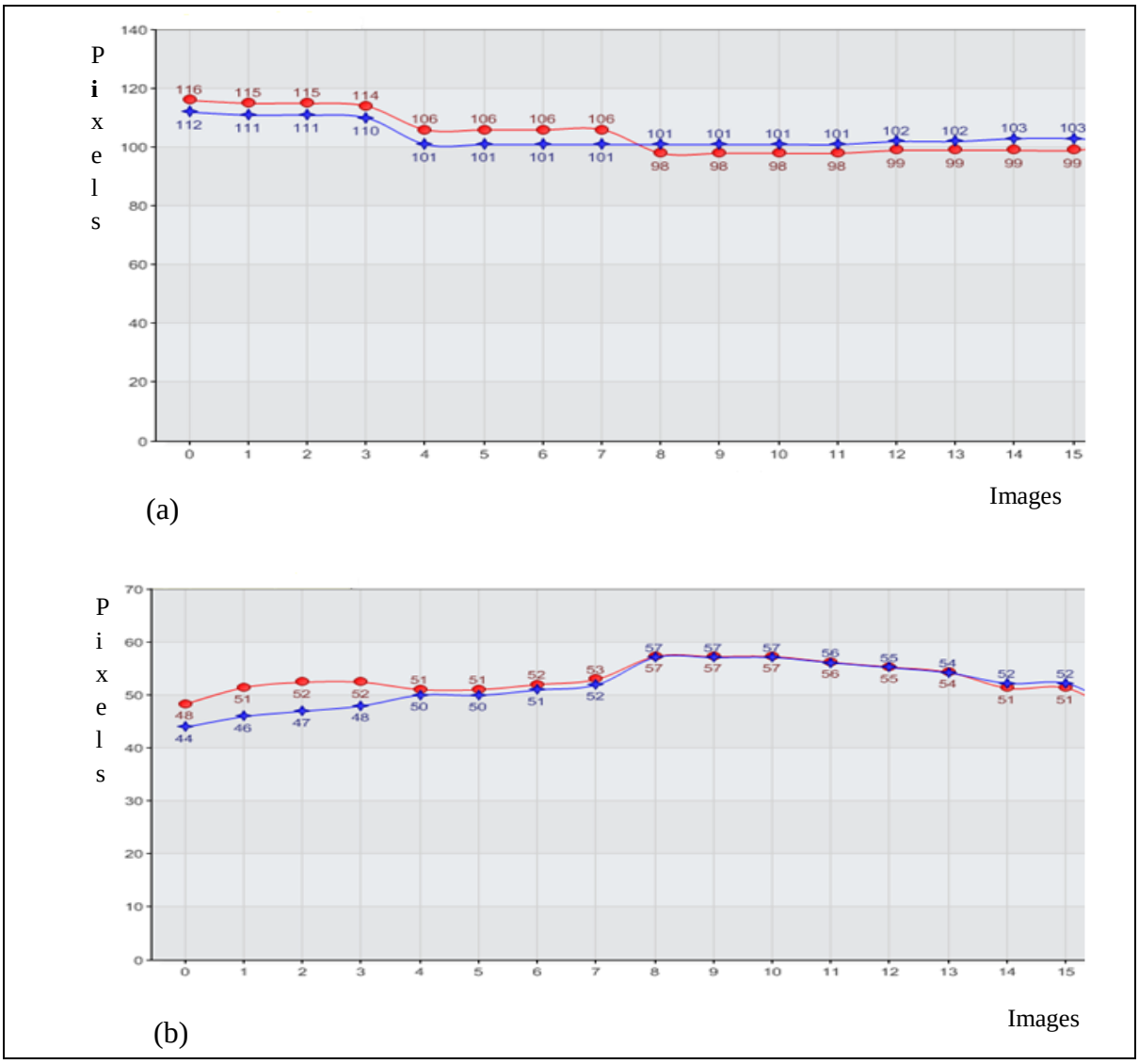


Figure 4. 24 : Allures à l'aide des méthodes manuelle et semi-automatique :
a) étirement b) ouverture.

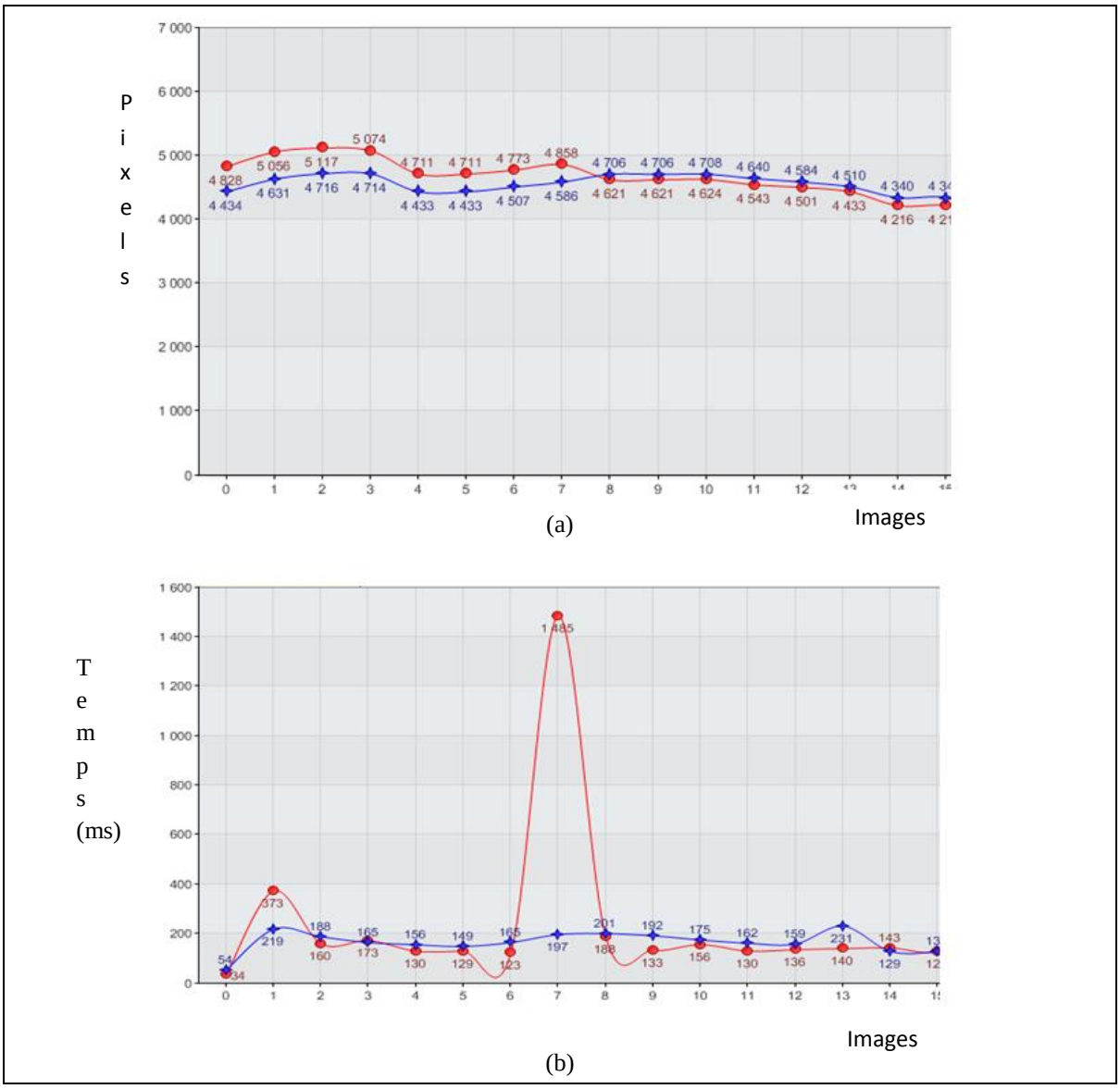
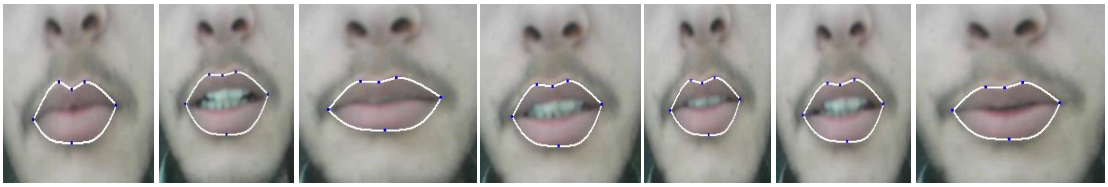


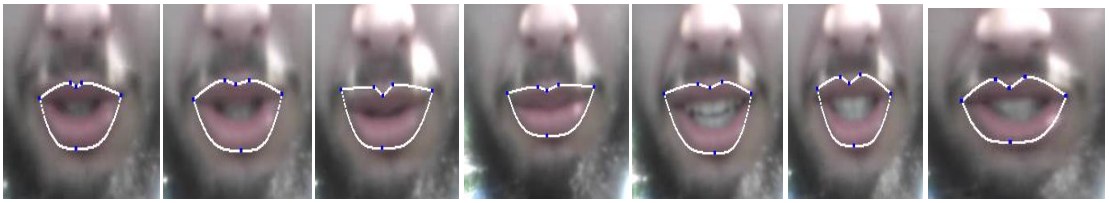
Figure 4.25 : Allures à l'aide des méthodes manuelle et semi- automatique :
a) surface b) temps d'exécution

4.6.4 Résultats sur des séquences vidéos variées

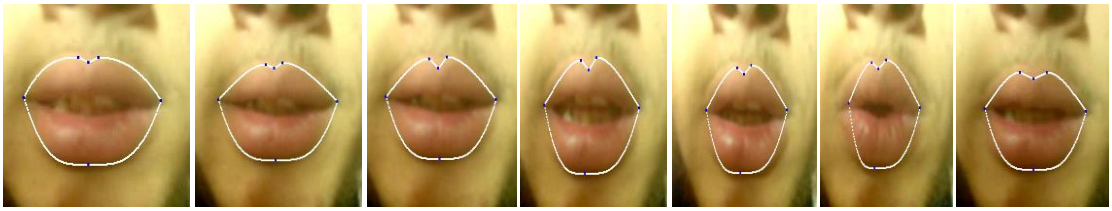
Nous présentons dans ce qui suit (figure 4. 26), les résultats obtenus à l'aide d'une variété de séquences vidéos traitées par notre application. Dans ce travail nous nous basons sur des faits courants, où la caméra n'est pas fixe, et le locuteur bouge doucement sa tête. Les 1ères images de chaque séquence sont obtenues par la segmentation statique (semi-automatique). Les autres images de chaque séquence sont obtenues par suivi des mouvements. La séquence ISL est constituée d'images bruitées. Les noms des images sont donnés par des abréviations, le numéro de l'image de la séquence est porté devant l'abréviation.



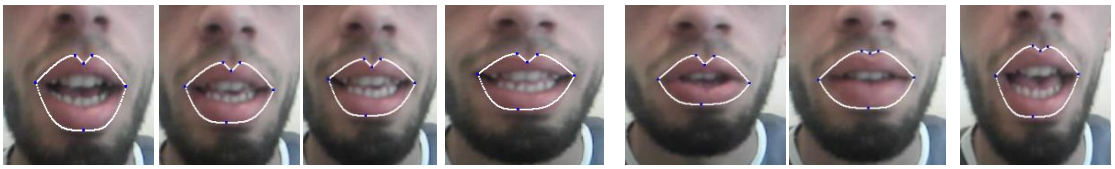
HAS1 HAS59 HAS64 HAS70 HAS75 HAS82 HAS100



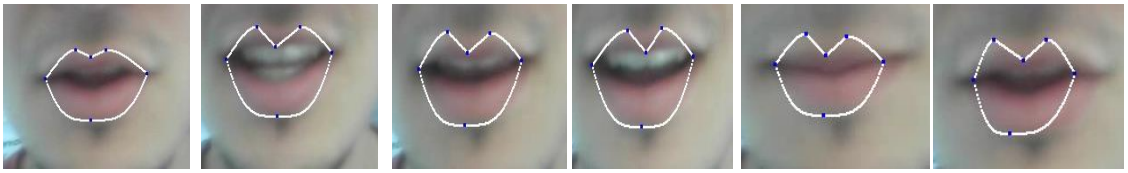
ABD1 ABD13 ABD20 ABD25 ABD33 ABD42 ABD50



BIS1 BIS6 BIS12 BIS18 BIS27 BIS32 BIS37



AL1 AL9 AL17 AL25 AL33 AL43 AL68



ISL1 ISL24 ISL63 ISL94 ISL112 ISL123

Figure 4.26 : Résultats de la segmentation et du suivi sur un cocktail de séquence d'images vidéos

Nous constatons à travers nos résultats, que notre application fait l'extraction et le suivi des contours labiaux pour des séquences d'images variées nettes ou bruitées. Nous voyons que les formes des lèvres obtenues sont très réalistes et convenables aux contours labiaux. La méthode est robuste même dans les cas suivants tel qu'un locuteur barbu, ou moustachu.

4.7 Conclusions

Au cours de ce chapitre, nous avons exposé les différentes étapes de la segmentation statique et dynamique, Le modèle élaboré est suffisamment flexible pour représenter presque toutes les formes de bouches. De plus, nous avons estimé la position des PC avec une précision moyenne comparable à une saisie manuelle rigoureuse, ce qui prouve la robustesse de nos algorithmes [10]. La prise en compte d'informations temporelles améliore significativement la qualité et la rapidité de la segmentation en vue de la lecture labiale.

La première étape des applications de la lecture labiale consiste à segmenter la région des lèvres et à estimer des paramètres labiaux représentés par différentes mesures calculées à partir des contours extérieurs.

Les paramètres labiaux ont été obtenus avec l'algorithme de suivi proposé dans cette thèse pour plus de 100 images par séquence et pour des sujets différents. Ces paramètres labiaux permettent d'obtenir une estimation de l'aire d'ouverture de la bouche.

Ainsi cette étude met en évidence l'apport de l'information visuelle qui est une information complémentaire intéressante pour aider à la perception de la parole dans un environnement bruité.

Dans cette thèse, nous nous sommes basés sur une variété des travaux antérieurs [3], [7], et [8]; et nous avons contribué dans les points suivants :

- l'emplacement manuel du point P3 d'une façon rigoureuse sur la lèvre, pour la détermination du cupidon supérieur;
- la localisation des commissures P1 et P5 en divisant notre rectangle de sélection en portions ;
- la détermination du point P6 sur la lèvre inférieure, en réduisant la zone de recherche à un rectangle R_0 centré par la ligne qui passe par P3 ;
- le tracé de la cubique par la méthode de pente optimale ;
- l'application des méthodes de mise en correspondance dans le cas du suivi des lèvres.

Nous constatons à travers nos résultats, que notre application fait l'extraction et le suivi des contours labiaux pour des séquences d'images variées nettes ou bruitées.

Conclusions Générales et Perspectives

Dans cette thèse nous avons présenté la perception visuelle de la parole à l'aide de la segmentation des lèvres, ensuite nous avons déterminé les paramètres labiaux afin d'établir ultérieurement un environnement logiciel permettant de lire la parole sur les lèvres même à une distance lointaine. Dans notre travail, nous nous sommes intéressés à l'extraction des contours labiaux externes sur des séquences vidéos acquises dans des conditions naturelles avec différents locuteurs. Les conditions de prise de vue relativement libres, et les impératifs de temps réel nous ont poussés à développer un algorithme robuste et précis. Par robustesse, nous voulions obtenir une méthode fiable ne nécessitant pas de réglage de paramètres. Par précision, nous voulions fournir une modélisation fidèle des contours de la bouche.

Tout d'abord, nous avons établi un état de l'art sur les connaissances essentielles qui décrivent les natures physiologiques et phonétiques de la parole, suivi par sa perception auditive, ensuite nous avons introduit certains concepts de la perception visuelle de la parole sur le plan physiologique, acoustique et phonétique.

Des techniques de traitement d'images ayant fait l'objet d'une utilisation dans le cadre de suivi des contours labiaux. Nous avons utilisé la pseudo-teinte qui permet d'effectuer une bonne séparation des lèvres et de la peau, de plus nous avons introduit un gradient hybride qui combine à la fois les informations de luminance et de chrominance, et qui facilite la localisation de la frontière supérieure de la bouche.

Pour détecter les PC : P2 et P4 dans la première image, nous avons introduit un nouveau type de contour actif : le jumping snack que nous avons initialisé par un seul point P3 situé au bas de l'arc du Cupidon. Le jumping snack permet de localiser le contour supérieur des lèvres ainsi que les PC : P2 et P4. Pour les points de commissures P1 et P5 sont obtenus en divisant notre rectangle de sélection, entourant la zone labiale, en portions. La détermination du point P6 sur la lèvre inférieure s'obtient en réduisant la zone de recherche à un rectangle R_0 centré par la ligne qui passe par P3. Le tracé des cubiques supérieures et inférieures se fait par la méthode proposée qui est la méthode de la pente optimale. Ainsi, nous avons exposé notre modèle composé de courbes cubiques et nous avons montré qu'il est suffisamment flexible pour reproduire la plupart des formes de bouches rencontrées lors de l'élocution [10]. Pour le suivi des lèvres nous avons proposé une méthode de mise en correspondance permettant de réduire les erreurs du suivi ; ainsi la méthode de recalage a été établie pour maintenir une parfaite segmentation tout au long de la séquence.

Nous avons montré que la compensation systématique des erreurs de suivi maintient le modèle près du contour des lèvres et permet d'envisager la segmentation des séquences très longues [11,12]. De plus, la précision de ce suivi est comparable à celle d'une saisie manuelle.

Nous constatons à travers nos résultats, que notre application fait l'extraction et le suivi des contours labiaux pour des séquences d'images variées nettes ou bruitées. Nous voyons que les formes des lèvres obtenues sont très réalistes et convenables aux contours labiaux. La méthode est robuste et précise même dans les cas suivants tel qu'un locuteur barbu, ou moustachu. A travers les graphes donnés dans le dernier chapitre, paragraphe 4.6.3, nous constatons aussi une conformité entre les résultats des deux méthodes : manuelle et semi-automatique, ce qui prouve la rigueur de nos algorithmes.

Enfin, nous avons déterminé les paramètres labiaux et le temps d'exécution [13] qui sont des éléments importants pour l'élaboration ultérieure d'une Base de Données (BD) permettant de reconnaître la parole.

De plus, la précision de la segmentation obtenue permet d'ores et déjà d'envisager une application aux domaines de l'identification par les lèvres.

Cette étude a aussi pour perspectives d'aider les sourds sévères ou profonds, oralistes c'est-à-dire éduqués essentiellement dans la langue orale à l'aide de la lecture labiale et soumis précocement aux compléments gestuels du LPC (Langage Parlé Complété). Pour cela, nous envisageons d'introduire une BD contenant les paramètres labiaux en mode long et tenu avec divers locuteurs prononçant le même visème, afin d'élaborer un système d'aide à la lecture labiale pour des sujets malentendants Algériens. Ce système nécessite un certain niveau de précision.

Références Bibliographiques

[1] L. Revéret et L. Le Chevalier, Un Modèle Géométrique de Lèvres 3D. 22èmes Journées d'Etudes sur la Parole, Martigny, pp. 213 – 216, 15-19, Juin 1998.

[2] C. Benoît, T. Mohamadi, S. Kandel, Effect of Phonetic Context on AudioVisual Intelligibility of French. J. Speech & Hearing Res, Vol. 37, pp. 1195-1203, 1994.

[3] T. Lallouache, Un Poste Visage-Parole. Acquisition et Traitement Automatique des Contours des Lèvres. Thèse de doctorat, Institut National Polytechnique de Grenoble, INPG France, 1991.

[4] P. Delmas, Extraction des contours des lèvres d'un visage parlant : Application à la communication multimodale. Thèse de Doctorat, Institut National Polytechnique de Grenoble (INPG), France, 2000.

[5] N. Eveno, A. Caplier and P.Y. Coulon, Automatic and Accurate lips traking, IEEE Transaction on Circuits and Systems of Video Technology, Vol 14, N°5, pp. 706-715, May 2004.

[6] P. Gacon, P.Y. Coulon and G. Bailly, Nonlinear active model for mouth inner and outer contours detection. European Signal Processing Conference (EUSIPCO-05), Antalya Turkey, 2005, http ://hal.archives-ouvertes.fr/hal-00378352/fr/.

[7] C. Bouvier, P.Y Coulon and X. Maldague, Unsupervised lips segmentation based on ROI optimisation and parametric Model. Proceeding of the IEEE International Conference on Image Processing, IEEE XPlore Press, San Antonio, pp. IV301 –IV 304, Sept 16-Oct 19, 2007.

[8] C. Bouvier, Segmentation Région-Contour des Contours des Lèvres. Thèse de doctorat Institut Polytechnique de Grenoble, France.

[9] La lecture labiale. ARDDS, www.ardds.org/content/la-lecture-labiale.

[10] M.L. Hamidatou, M. Guerti, and S. Ait-Aoudia, Static Segmentation of the Lips for Follow-up. Journal of Computer Science, Vol. 05, N° 12, pp. 991-997, 2009, ISSN : 1549-3636 http://www.scipub.org

[11] M.L Hamidatou, M.Guerti, F.Z.Hamidatou-Hamadi, and S.Ait Aoudia, Suivi des points caractéristiques des lèvres en mouvement. Colloque d'Informatique, automatique et électronique, CIAE 2011, Mundiapolis Université Casablanca, Maroc, 24 et 25 Mars 2011.

[12] M. L. Hamidatou, M. Guerti and S. Ait-aoudia. Follow-Up of the Lips for a Labial Reading. International Review on Computers and Software (*IRECOS),* Vol.7, N.1 January 2012, ISSN:1828-6003.

[13] M. L. Hamidatou, M. Guerti, S. Ait-aoudia, Determination Labial Parameters: Opening, Stretching and Labial Surface of Moving Lips. Wulfenia journal, No.5, volume 21, May 2014, ISSN: 1561-882X.

[14] J. L. Schwartz, La parole multisensorielle : Plaidoyer, problèmes, perspective. Journées d'Etude sur la Parole, Fès, Maroc, p. 11-18, 2004.

[15] M. Lavrut, A.Noiret, Facteurs Prédictifs Pour l'acquisition d'une Lecture Labiale Fonctionnelle Chez L'adulte Sourd. Mémoire Pour Le Certificat De Capacité D'orthophoniste, Université Paris, 2013.

[16] F. Le Huche, A Allali., La voix Tome 1, 3ème Edition. Paris, Masson, p 199, 1991.

[17] Phonétique et phonologie. www.linguistes.com/phonetique/phon.htm.

[18] Google, Images correspondant à Tableau des consonnes du Français.

[19] N. Eveno, Segmentation des lèvres par un modèle déformable analytique. Thèse de Doctorat Institut National Polytechnique de Grenoble, INPG, Novembre 2003.

[20] B.Virole, J. Cosnier, Psychologie de la surdité. 2e Edition augmentée, Editeur Paris, Bruxelles, De Book Université, 2000.

[21] A. Mokrane, Le fonctionnement du système auditif. Lobe Santé auditive et communication, Audiologiste Saint-Augustin-de-Desmaures, Québec Sainte-Foy, août 2013.

[22] A. de Cheveigné, Modèles de traitement auditif dans le domaine temps. Mémoire d'Habilitation à Diriger des Recherches Neurosciences, Université Paris 6, 2000.

[23] G. Ehret, R. Romand, The auditory midbrain, a shunting yard of acoustical information processing in The central auditory system. New York, Oxford University Press, pp. 259-316, 1997.

[24] www.ecoute.ch/Perte_Auditive_Presentation_Causes_Symptomes. Perte auditive, types, degrés, symptômes, conséquences.

[25] K. Strelnikov et al. Visual activity predicts auditory recovery from deafness after adult cochlear implantation. Brain, Vol. 36, pp. 82-95, 2013.

[26] D.C. Marr, Vision. freeman, Oxford, 1982.

[27] Y. Aloimonos, What I have learned. CVGIP,Image Understanding, Vol. 60(1), pp. 74-85. 1994.

[28] É. Godaux**,** Cent Milliards de Neurones. Collection, La science apprivoisée, 1990, ISBN : 27011-13784.

[29] L. Alquier, Analyse Et Représentation Des Scènes Complexes Par Groupement Perceptuel : Application A La Perception De Structures Curvilignes. Thèse de Doctorat, Montpellier, Septembre 1998.

[30] M. Trivedi, et A. Rosenfeld, On making computers see. IEEE Trans.Systems. Man.Cybern, Vol.19, N°6, pp. 1333-1335, 1989.

[31] J.T. Devlin, J. Aydelotte, Speech perception: Motoric contributions versus the motor theory, Current Biology, Vol.19 N°5, pp. 198-200, 2009.

[32] N. Deggouj, L'intégration audio-visuelle Connaissances sur surdités. Audition – Vision, UCL Saint Luc, Bruxelles, N°11, Mars 2005

[33] Y. Samson, P. Belin, L. Thivard, N. Boddaert, S. Crozier, M. Zilbovicius, Auditory perception and language: functional imaging of speech sensitive auditory cortex. Revue Neurol, Paris, Vol. 157, pp. 837-846, 8-9 Mars, 2001.

[34] G.Giraud, A. Lise, E. Truy, The contribution of visual areas to speech comprehension: a PET study in cochlear implants patients and normal-hearing subjects. Neuropsychologia, Vol.40, N°9, pp. 1562–1569, 2002.

[35] fr.wikipedia.org/wiki/Effet_McGurk

[36] H. McGurk et J. MacDonald, Hearing lips and seeing voices. Nature, Vol. 264, N° 5588, pp. 746–748, 1976.

[37] A. Geffray, Présentation D'un Nouveau Test En Audiométrie : De L'influence De La Perception Audiovisuelle De L'environnement Sur La Compréhension De La Parole. Mémoire du diplôme d'état d'Audioprothésiste Faculté de Médecine, Université de Rennes1, octobre 2012.

[38] D.W. Massaro, and D.G. Stork, Speech recognition and sensory integration. Am Science, Vol 86, N° 3, pp. 236-244, 1998.

[39] L. Revéret, Conception et Evaluation d'un Système de Suivi Automatique des Gestes Labiaux en Parole, Thèse de Doctorat, Institut National Polytechnique, Grenoble, 1999. tel-00207391

[40] Q. Summerfield, Some Preliminaries to a Comprehensive Account of Audio-Visual. Speech Perception. In Hearing by Eye: The Psychology of Lipreading, B. Dodd and R. Campbell editors, pp. 3-51, 1987.

[41] B. E. Walden, R. A. Prosek, A. A. Montgomery, C.K. Scherr, and C. j. jones, Effects of training on the visual recognition of consonants. Journal of Speech and Hearing Research, Vol.20, pp. 130-145, 1977.

[42] A. Summerfield, A. MacLeod, M. McGrath, et M. Brooke, Lips,Teeth, and the Benefits of Lipreading. In Handbook of Research on Face Processing, A.W. Young et H.D. Ellis, Editors, Elsevier Science Publishers, Amsterdam, pp. 223-233, 1989.

[43] G. Fant, Speech Sounds and Features. Cambridge, MA, USA The MIT Press, 1973.

[44] P.K. Kuhl, A.N. Meltzoff, The bimodal perception of speech in infancy. Science, Vol. 218, pp. 1138-1141, 1982.

[45] W.H. Sumby and I. Pollack, Visual contribution to speech intelligibility in noise. Journal of the Acoustical Society of America, Vol.26, pp. 212-215, 1954.

[46] C.A. Binnie, A.A. Montgomery,and P.L. Jackson, Auditory and visual contributions to the perception of consonants. Journal of Speech & Hearing Research, Vol.17, pp. 619-630, 1974.

[47] Q. Summerfield Use of visual information for phonetic perception. Phonetica, Vol.36 pp.314- 331, 1979.

[48] Q.Summerfield, A. MacLeod, and M.M.McGrath Brooke, Lips, teeth, and the benefits of lipreading. in Handbook of Research on Face Processing, A.W. Young and H.D. Ellis editors, Elsevier Science Publishers, Tulips, pp. 223-233, 1989.,

[49] C. Benoît, T. Guiard-Marigny, B. Le Goff, and A. Adjoudani, Which Components of the FaceDo Humans and Machines Best Speechread ? », in Speechreading by Humans and Machines, D. Stork and M. Hennecke, editors, Springer-Verlag, Berlin, pp. 351-372, 1996

[50] P. Gacon, Analyse d'images et modèles de formes pour la détection et la reconnaissance Application aux visages en multimédia. Thèse de doctorat, Institut National Polytechnique de Grenoble, juillet 2006.

[51] E. Petajan, Automatic lipreading to enhance speech recognition. Thèse de Doctorat, Univ. Illinois at Urbana-Champagne, 1984.

[52] H. Terrat, projet labiao, perspectives pour la scolarisation de jeunes sourds sévères et profonds. Master, Technologie et Handicap, septembre 2006.

[53] M. J. Lyons, M. Haehnel and N. Tetsutani, Designing, Playing, and Performing with a Vision-Based Mouth Interface. Conference on New Interfaces for Musical Expression (NIME-03), Montréal, Canada, pp. 116-121, 2003.

[54] X. Zhang and R.M. Mersereau, Lip Feature Extraction Towards an Automatic Speechreading System. In Proc. International Conference on Image Processing (ICIP'00), Vancouver, Canada, 2000.

[55] R.L. Hsu, M. Abdel-Mottaleb and A. K. Jain, Face Detection in Color Images. IEEE Trans. Pattern Analysis and Machine Intelligence, Volume 24, N°5, pp. 696-706, Mai 2002.

[56] W.C.A. Liew, K.L. Sum, S.H. Leung and W.H. Lau, Fuzzy segmentation of lip image using cluster analysis. European Conference on Speech Communication and Technology (EUROSPEECH'99), Hongrie, 1999.

[57] M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active Contour Models. International Journal of Computer Vision, pp. 321-331, 1987.

[58] S. Horbelt and J. L. Dugelay, Active Contours for Lipreading Combining Snakes with Templates. 15th GRETSI Symposium on Signal and Image Processing, Juan les Pins, France, 1995.

[59] T. F. Cootes, C. J.Taylor, D. Cooper and J. Graham, Training Models of Shape from Sets of Examples. In D. Hogg and R. Boyle, editors, 3rd British Machine Vision Conference, Springer-Verlag, pp. 9 –18, September 1992.

[60] T.F. Cootes, C.J. Taylor,and D.H. Cooper, Active Shape Models – Their Training and Application. Computer Vision and Image Understanding, Vol. 61, N°.1, pp. 38-59, January 1995.

[61] Y. Tian, T. Kanade and J. Cohn, Robust Lip Tracking by Combining Shape, Color and Motion. 4th Asian Conference on Computer Vision (ACCV'00), January, 2000.

[62] B. Lucas et T. Kanade, An Iterative Image Registration Technique with an Application in Stereo Vision. In The 7 th International Joint Conference on Artificial Intelligence, pp. 674–679, 1981.

[63] P. Daubias, Modèles A Posteriori de la Forme et de l'Apparence des Lèvres pour la Reconnaissance Automatique de la Parole Audiovisuelle. Thèse de Doctorat, Université du Maine, 2002

[64] J.J. Rousselle, Les contours actifs, une méthode de segmentation- Application à l'imagerie médicale. Thèse de doctorat, Université de Tours, 2003.

[65] N. Merzougui, Un algorithme évolutionnaire pour la segmentation d'image basé sur le diagramme de Voronoi. Thèse de magistère, Université Kasdi Merbah,Ouargla, Algérie 2012.

[66] C.C. Chu and J.K. Aggarwal, The integration of image segmentation maps using region and edge information. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 15 N°12, pp. 1241-1252, 1993.

[67] M. Kass, A.witkin, and D. Terzopoulos, Snakes: Active Contour Models. International Journal of Computer Vision, pp. 321-331, 1988.

[68] D.J. Burr, Elastic Matching of Line Drawings. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Vol. 3, N° 6, pp. 708-713, Novembre 1981.

[69] C. Xu, E. Segawa and S. Tsuji, A Robust active contours with insensitive parameters. In Proc of 4th International Conference on Computer Vision, pp. 562-566,1993. Et Pattern recognition, Vol. 27, N° 7, pp. 879-884, 1994.

[70] L.D. Cohen, On Active Contour Models and Balloons, Computer Vision. Graphics, and Image Processing: Image Understanding, Vol. 53, N° 2, p. 211-218,March 1991.

[71] A.A. Amini, S. Tehrani and T.E. Weymouth, Using Dynamic Programming for Minimizing the Energy of Active Contours in the Presence of Hard Constraints. 2nd Int. Conf. Comput. Vision, Tampa,FL, USA, pp. 95-99, décembre 1988.

[72] D.J. Williams et M. Shah. A fast algorithm for active contours and Curvature estimation. CVIGP Computer Vision Graphics Image Process : Image Understanding, Vol 55, N°1, pp. 14-25, Janvier 1992.

[73] M.O. Berger, Les contours actifs : modélisation, comportement et convergence », Thèse de doctorat de l'Institut National Polytechnique de Lorraine, spécialité: Informatique, 6 février 1991.

[74] W. Lu. Yang, and A. Waibel, Skin-color modeling and adaptation. Technical Report CMU-CS, School of computer Science, Carnegie Mellon University, pp. 97-146,1997.

[75] N. Eveno, A. Caplier and P.Y. Coulon. A Parametric Model for Realistic Lip Segmentation. 7th International Conference on Control, Automation, Robotics and Vision (ICARCV'02), Singapore, December 2002.

[76] E.D. Petajan, B. Bischoff, D. Bodoff and N.M. Brooke, An improved automatic lipreading system to enhance speech recognition. CHI 88, pp. 19-25, 1988.

[77] S. Stillittano, A. Caplier, Inner Lip Contour Segmentation by Combining Active Contours and Parametric Models. International Conference on Computer Vision Theory and Applications (VISAPP 2008), Madeira, Spain, 2008.

[78] Q. D. Nguyen, Détection et suivi automatique du mouvement des lèvres. Thèse de doctorat. Université Pierre et Marie Curie, 2010.

[79] Z. Hammal, N. Eveno, A. Caplier and P.Y. Coulon. Extraction réaliste des traits caractéristiques du visage à l'aide de modèles paramétriques adaptés. Colloque GRETSI sur le traitement du signal et d'image (GRETSI'03), Paris, France, 2003.

[80] X. Zhang, R.M. Mersereau, Lip Feature Extraction Towards an Automatic Speech reading System. Proc. International Conference on Image Processing, Vol.3, 2000.

[81] T. Coianiz, L. Torresani, and B. Caprile, 2D deformable models for visual speech analysis. D.G. Stork &M.E. Hennecke Editors, Speech reading by Humans and Machines: Models, Systems, and Applications. Springer-Verlag, pp. 391-398, 1996.

[82] J.C. Wojdel and L.J.M. Rothkrantz, Using Aerial and Geometric Features in Automatic Lip-Reading. In Proceedings 7th Eurospeech, Aalborg, Denmark, Vol. 4, pp. 2463-2466, 2001.

[83] E. K. Patterson, S. Gurbuz, Z. Tufekci and J. H. Gowdy, Moving-Talker, Speaker-Independent Feature Study and Baseline Results using the Cuave Multimodal Speech Corpus. EURASIP Journal on Applied Signal Processing, Issue 11, pp. 1189-1201, 2002.

[84] A.V. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao and K. Murphy, A Coupled HMM for Audio-Visual Speech Recognition. In Proceedings 2002 IEEE International Conference on Acoustics, Speech and Signal Processing, Orlando, Vol. 2, pp. 2013-2016, 2002.

[85] A. Hulbert and T. Poggio, Synthesizing a color algorithm from examples. Science, Vol. 239, pp.482 - 485, 1988.

[86] M. Hennecke, V. Prasad, et D. Stork. Using deformable templates to infer visual speech dynamics. 28 th Annual Asimolar Conference on Signals, Systems, and Computer, IEEE Computer, Pacific Grove, Volume 2, pp. 576-582, 1994.

[87] P. Radeva and E. Marti, Facial features segmentation by model-based snakes. International Conference on Computer Analysis of Images and Patterns, 1995.

[88] S. Werda, W. Mahdi and A. B. Hamadou, Automatic Hybrid Approach for Lip POI Localization: Application for Lip-Reading System, 1st International Conference on Information and Communication Technology and Accessibility (ICTA'07), Hammamet, Tunisia, April 2007.

[89] H. Seyedarabi, W. Lee, and A. Aghagolzadeh, Automatic Lip Tracking and Action Units Classification using Two-step Active Contours And Probabilistic Neural Networks, Canadian Conference On Electrical and Computer Engineering, (CCECE'2006), Ottawa, Canada, pp.2021-2024, 2006.

[90] S. Stillittano, V. Girondel and A. Caplier, Inner and Outer Lip Contour Tracking using Cubic Curve Parametric Models. IEEE International Conference on Image Processing (ICIP09), 2009.

[91] A. Botino, Real time head and facial features tracking from uncalibrated monocular views. In Proc. 5th Asian Conference on Computer Vision (ACCV'02), Melbourne 2002.

[92] A.W.C. Liew, S.H. Leung, and W.H. Lau, Lip contour extraction using a deformable model. Int. Conf. on Image Processing (ICIP'00), Vancouver, Canada, 2000.

[93] A. Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. Int. Journal of Computer Vision, Vol. 8, N°. 2, pp. 99-111, 1992.

[94] M.Pantic, M.Tome, and L.J.M. Rothkrantz, A hybrid Approach to Mouth Features Detection , Proceedings of IEEE Int' l Conf. Systems, Man and Cybernetics (SMC'01), pp.1188-1193,Tucson, USA.

[95] S. Lankton, D. Nain, A.Yezzi and A.Tannenbaum, "Hybrid geodesic region-based curve evolutions for image segmentation", *Proc. SPIE* 6510, Medical Imaging 2007: Physics of Medical Imaging, 65104U (March 16, 2007); doi:10.1117/12.709700; http://dx.doi.org/10.1117/12.709700 .

[96] T. Chan and L. Vese, Active contours without edges, Tech. Rep. 9853 Computational Applied Math Group, UCLA, 1998.

[97] B. Beaumesnil, Suivi labial couleur pour analyse-synthèse vidéo et communication temps-réel. Thèse de doctorat, Université de Pau et des Pays de l'Adour 2006.

[98] N. Eveno, A. Coplier and P.Y Coulon, "Jumping snakes and parametric model for lips segmentation", International Conference on Image Processing, Barcelona, Spain, pp. 867-870, Sept 2003.