



HAL
open science

Conception et utilisation d'atlas anatomiques pour la segmentation automatique : application à la radiothérapie des cancers ORL

Liliane Ramus

► **To cite this version:**

Liliane Ramus. Conception et utilisation d'atlas anatomiques pour la segmentation automatique : application à la radiothérapie des cancers ORL. Imagerie médicale. Université Nice Sophia Antipolis, 2011. Français. NNT : . tel-00845098

HAL Id: tel-00845098

<https://theses.hal.science/tel-00845098v1>

Submitted on 16 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE NICE - SOPHIA ANTIPOLIS
ÉCOLE DOCTORALE STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION
ET DE LA COMMUNICATION

THÈSE

pour obtenir le titre de

Docteur en Sciences

de l'Université de Nice - Sophia Antipolis

**Mention : AUTOMATIQUE, TRAITEMENT DU SIGNAL ET
DES IMAGES**

Présentée et soutenue par

Liliane RAMUS

Conception et utilisation d'atlas anatomiques pour la segmentation automatique : application à la radiothérapie des cancers ORL

Thèse dirigée par Grégoire MALANDAIN,
préparée à l'INRIA Sophia Antipolis, projet ASCLEPIOS,
soutenue le 19 juillet 2011

Jury :

<i>Rapporteurs :</i>	Elsa ANGELINI	-	Télécom ParisTech
	Jean-Philippe THIRAN	-	École Polytechnique Fédérale de Lausanne
<i>Directeur :</i>	Grégoire MALANDAIN	-	INRIA - Asclepios
<i>Président :</i>	Nicholas AYACHE	-	INRIA - Asclepios
<i>Examineurs :</i>	Vincent GRÉGOIRE	-	Université Catholique de Louvain
	David SARRUT	-	CNRS - Creatis
<i>Invités :</i>	Hanna KAFROUNI	-	DOSIsoft S.A.
	Juliette THARIAT	-	Centre Antoine Lacassagne

Remerciements

Je voudrais commencer par remercier les personnes qui m'ont donné l'opportunité de faire cette thèse : Grégoire Malandain et Hanna Kafrouni pour m'avoir proposé ce projet de thèse, et Nicholas Ayache pour m'avoir accueillie au sein de l'équipe Asclepios de l'INRIA.

Je souhaite vivement remercier Elsa Angelini et Jean-Philippe Thiran pour avoir accepté le rôle de rapporteurs sur ce travail de thèse, pour leur lecture attentive du manuscrit et pour leurs remarques constructives. Je remercie aussi David Sarrut et Vincent Grégoire pour l'intérêt qu'ils ont porté à mon travail en tant qu'examineurs. Merci enfin à Nicholas Ayache d'avoir accepté de présider le jury de soutenance.

Un travail de thèse ne s'entreprend pas seule. Je souhaite adresser ma profonde reconnaissance à Grégoire Malandain pour m'avoir guidé tout au long de ce projet. J'ai énormément appris en travaillant à ses côtés, sur les aspects théoriques évidemment mais aussi sur l'art et la manière de retourner un problème dans tous les sens jusqu'à en prendre le dessus. Je lui suis aussi reconnaissante pour ses critiques constructives qui m'ont beaucoup aidé à clarifier mes idées. Je le remercie enfin pour l'enthousiasme constant qu'il a eu par rapport à mes travaux et pour ses qualités humaines.

Je voudrais aussi remercier l'ensemble des personnes de Dosisoft avec lesquelles j'ai eu l'occasion de travailler. Tout d'abord merci à Hanna Kafrouni pour la confiance qu'il m'a accordée en me proposant cette thèse. Je souhaite aussi remercier les ingénieurs avec lesquels j'ai eu grand plaisir à travailler : Jean-Bernard Ruaud, Jérôme Durant, Souha Aouadi, Fabien Lebeaux et Mehdi Benkebil.

Je remercie également les radiothérapeutes avec lesquels j'ai eu la chance d'interagir, Vincent Grégoire et Juliette Thariat. J'ai beaucoup apprécié l'intérêt clinique qu'ils ont porté à mon travail. Je les remercie également d'avoir accepté de faire partie de mon jury de thèse. Enfin, j'adresse un merci particulier à Juliette Thariat pour nos nombreuses discussions et échanges fructueux, ainsi que pour son dynamisme, son enthousiasme et sa disponibilité. J'ai eu beaucoup de plaisir à travailler dans le cadre dans cette collaboration.

Je souhaite remercier sincèrement Olivier Commowick pour le temps qu'il a consacré à m'aider au début de ma thèse. Un grand merci pour sa patience et sa disponibilité. Sur le plan technique, je remercie aussi Nicolas Niclausse pour son soutien logistique sur le cluster.

Je remercie Daniel Barbeau, Florence Billet, Marine Breuilly, Aurélie Canale, François Chung, Florence Dru et Erik Pernod pour leur amitié et leur humour qui ont fait de ces 3-4 ans sur la Côte d'Azur une période formidable.

J'adresse des remerciements tout particuliers aux différentes personnes avec qui j'ai eu l'occasion de partager le bureau : Jimena Costa (merci pour ton accueil et ton amitié, et merci d'être revenue pour la soutenance), Florence Billet (un grand merci pour toutes nos discussions et pour ton soutien dans les moments difficiles), Marine Breuilly (merci pour ton soutien et tes encouragements pendant la rédaction,

sans oublier les amis de Flavigny :- (Vu!), Kristin McLeod (Door!), Barbara André (merci pour toutes nos discussions et fous rires, à Asclepios ou ailleurs), Vladlena Gorbunova (thank you for your infectious cheerfulness!). Merci à toutes d'avoir contribué à l'ambiance chaleureuse dudit *bureau des filles*!

Je remercie aussi l'ensemble des membres passés et présents de l'équipe Asclepios pour leur bonne humeur quotidienne et leur sympathie : Fatih Arslan, Benoît Bleuzé (malgré ton acharnement sur la pauvre baleine), Olivier Clatz, Jean Cousty, Hervé Delingette, Guillaume Dugas-Phocion, Stanley Durrleman, Romain Fernandez, Pierre Fillard, Vincent Garcia (merci pour ton humour pince-sans-rire et ton soutien pendant la rédaction), Ezequiel Geremia (merci pour nos discussions), Pascal Girard, Vikash Gupta, Tobias Heimann (thanks for the piscinema session!), Michael Knopke, Ender Konukoglu, Muriel Laffon-Pioger, Hans Lamecker, Damien Lepiller, Yonni Levy, Herve Lombaert, Marco Lorenzi, Tommasso Mansi, Stéphanie Marchesseau (merci pour le baptême de moto et pour nos discussions), Jan Margeta, Xavier Pennec, Jean-Marc Peyrat, Adityo Prakosa, Islem Rekik, Jatin Relan, Alexander Schmidt-Richberg, Stephan Schmitt, Maxime Sermesant, Jean-Christophe Souplet (merci pour ton accueil chaleureux), John Stark, Erin Stretton, Isabelle Strobant (merci pour ton soutien logistique sans faille et pour ton amitié), Andrew Sweet, Hugo Talbot, Nicolas Toussaint, Ken Wong. Merci à tous!

Je ne pourrai finir ces remerciements sans exprimer toute ma reconnaissance à mes amies de longue date, Sandra et Vérane, qui malgré la distance ont toujours été là quand j'en avais besoin. Un grand merci pour leurs conseils avisés. Je remercie aussi mes parents et mes sœurs pour leur soutien et leurs encouragements pendant la rédaction, ainsi que pour leur venue à ma soutenance.

Enfin, et non des moindres, je ne remercierai jamais assez Manu pour sa patience, son soutien et son amour qui m'ont permis d'arriver au bout de la rédaction. Puissent ces quelques lignes lui témoigner de ma reconnaissance la plus sincère.

Table des matières

1	Introduction	1
1.1	Contexte et objectifs	1
1.2	Organisation du manuscrit	3
2	Contexte médical	5
2.1	Cancers de la région ORL	5
2.1.1	Statistiques sur les cancers ORL	5
2.1.2	Le développement du cancer	8
2.1.3	Le diagnostic d'un cancer ORL	9
2.1.4	Les traitements possibles	10
2.1.5	Les structures critiques de la région ORL	12
2.2	Radiothérapie externe	12
2.2.1	Principe de base	12
2.2.2	Radiobiologie des rayonnements ionisants	13
2.2.3	Différents types de radiothérapies externes	15
2.2.4	Planification du traitement	23
2.2.5	Objectif de la thèse dans la planification du traitement	29
3	Segmentation par atlas	31
3.1	Segmentation d'images médicales	31
3.1.1	Principe et applications	31
3.1.2	Segmentation manuelle <i>versus</i> segmentation automatique	32
3.2	Bref état de l'art des méthodes de segmentation	33
3.2.1	Méthodes de bas niveau utilisant peu ou pas d' <i>a priori</i>	33
3.2.2	Méthodes utilisant des <i>a priori</i> de haut niveau	35
3.2.3	Notre application	38
3.3	Recalage	41
3.3.1	Principe	41
3.3.2	Utilisation du recalage en radiothérapie	41
3.3.3	Différentes approches	42
3.3.4	Classes de transformation	47
3.3.5	Mise en œuvre	48
3.4	État de l'art des stratégies de segmentation par atlas	49
3.4.1	Atlas unique utilisé pour tous les patients	50
3.4.2	Construction d'atlas adaptés à l'anatomie du patient	53
3.4.3	Conclusion	59

4	Calcul d'une segmentation moyenne à partir d'une base incomplète	61
4.1	Introduction	62
4.1.1	Contexte et problématique	62
4.1.2	Notations	63
4.1.3	État de l'art	64
4.1.4	Organisation du chapitre	66
4.2	Utilisation de l'algorithme STAPLE	66
4.2.1	Description de l'algorithme	66
4.2.2	Gestion des structures manquantes	68
4.2.3	Résultats	68
4.2.4	Analyse de la sur-segmentation obtenue	71
4.2.5	Correction de la sur-segmentation par post-traitement	75
4.2.6	Correction de la gestion des structures manquantes	76
4.3	Algorithme de gestion des structures manquantes dans un calcul de segmentation probabiliste moyenne	77
4.3.1	Principe	77
4.3.2	Mise en œuvre avec STAPLE	78
4.3.3	Mise en œuvre avec un vote par majorité	79
4.4	Calcul d'une segmentation moyenne à partir du degré d'accord local	79
4.4.1	Principe	79
4.4.2	Algorithme global	79
4.5	Évaluation	84
4.5.1	Correction <i>a posteriori</i> de la sur-segmentation dans l'atlas	84
4.5.2	Algorithmes comparés	85
4.5.3	Évaluation quantitative	87
4.5.4	Évaluation qualitative	91
4.5.5	Robustesse aux structures manquantes	91
4.6	Limites de l'utilisation d'un atlas moyen	98
4.7	Conclusion et synthèse des contributions	100
5	Algorithmes de segmentation multi-atlas adaptés au patient	101
5.1	Introduction	102
5.1.1	Motivations et vocabulaire	102
5.1.2	Notations	104
5.2	Méthodes de sélection des atlas les plus adaptés au patient	104
5.2.1	Revue des critères de sélection basés sur des méta-informations	104
5.2.2	Revue des critères de sélection basés intensité	104
5.2.3	Revue des critères de sélection basés déformation	106
5.2.4	Revue des critères de sélection basés sur des segmentations	107
5.2.5	Contributions relatives aux critères de sélection	108
5.2.6	Mise en œuvre des méthodes de sélection	109
5.3	Évaluation non biaisée des méthodes de sélection	113
5.3.1	Corrélation moyenne de classement entre méthodes de sélection	114
5.3.2	Algorithme de partitionnement par affinité-propagation	114

5.3.3	Algorithme d'évaluation non biaisée	115
5.4	Différentes stratégies pour la fusion des atlas sélectionnés	116
5.4.1	Préambule	116
5.4.2	Construction hors-ligne de sous-atlas	117
5.4.3	Création à la volée d'atlas adaptés au patient	119
5.4.4	Fusion des segmentations dans le référentiel du patient	122
5.5	Conclusion et synthèse des contributions	128
6	Évaluation des algorithmes multi-atlas adaptés au patient pour la segmentation des structures ORL	129
6.1	Évaluation des sous-atlas par localisation et stade de tumeur	130
6.1.1	Partitionnement de la base d'images	130
6.1.2	Évaluation clinique	131
6.1.3	Évaluation sur notre base d'images	133
6.1.4	Synthèse et discussion	135
6.2	Évaluation des méthodes de sélection d'atlas basées image	136
6.2.1	Méthodes de sélection d'atlas étudiées	136
6.2.2	Évaluation d'après les classements d'atlas	138
6.2.3	Évaluation d'après les performances rang par rang	146
6.2.4	Synthèse et discussion	147
6.3	Évaluation des approches de fusion des segmentations dans le réfé- rentiel du patient	149
6.3.1	Évaluation dans le cadre d'une sélection globale ou régionale	151
6.3.2	Évaluation dans le cadre d'une pondération locale	156
6.3.3	Comparaison quantitative	158
6.3.4	Comparaison qualitative	160
6.4	Création à la volée d'un atlas composite adapté au patient	170
6.4.1	Paramètres	170
6.4.2	Résultats	170
6.5	Synthèse des contributions	175
6.6	Conclusion : limites et perspectives	179
7	Segmentation automatique des dents pour l'évaluation des risques liés à une chirurgie dentaire post-irradiation	181
7.1	Introduction	182
7.1.1	Contexte clinique des soins dentaires post-irradiation	182
7.1.2	Comment estimer la dose reçue par chaque dent ?	182
7.1.3	Organisation du chapitre	184
7.1.4	Dénomination des dents	184
7.2	État de l'art sur la segmentation des dents	184
7.2.1	Difficultés principales	184
7.2.2	État de l'art	185
7.3	Description de l'algorithme	186
7.3.1	Base de données	186

7.3.2	Algorithme proposé	188
7.4	Résultats	190
7.4.1	Protocole d'évaluation	190
7.4.2	Évaluation en terme de qualité de segmentation	191
7.4.3	Évaluation en terme d'estimation de dose	195
7.4.4	Évaluation dent par dent	200
7.5	Discussion	202
7.5.1	Synthèse des résultats	202
7.5.2	Comparaison avec la dose sur la mandibule ou le maxillaire	203
7.5.3	Applications cliniques	204
7.5.4	Perspectives	204
7.6	Conclusion et synthèse des contributions	205
8	Conclusion	207
8.1	Synthèse des contributions	207
8.1.1	Construction d'une segmentation moyenne	208
8.1.2	Construction d'atlas anatomiques adaptés au patient	209
8.1.3	Segmentation automatique des dents	212
8.1.4	Intégration dans ISOgray et validation clinique	213
8.2	Perspectives	214
8.2.1	Perspectives liées aux données	214
8.2.2	Perspectives liées à l'évaluation	215
8.2.3	Perspectives liées à la méthodologie	215
A	Algorithme de recalage utilisé	221
A.1	Algorithme global	221
A.2	Algorithme utilisé pour le recalage rigide et affine	221
A.2.1	Estimation des appariements	222
A.2.2	Estimation de la correction de transformation	222
A.3	Algorithme utilisé pour le recalage non-linéaire	222
A.3.1	Estimation des appariements	223
A.3.2	Estimation de la correction de transformation	223
A.3.3	Rejet des appariements aberrants	223
B	Mesures d'évaluation de segmentation automatique	225
B.1	Mesures d'évaluation	225
B.1.1	Mesures volumiques	225
B.1.2	Mesures surfaciques	227
B.2	Mise en œuvre pour les structures de la région ORL	229
B.2.1	Particularités des structures considérées	229
B.2.2	Mise en œuvre proposée	230
C	Variabilité de contourage des structures ORL	233
	Bibliographie	239

Introduction

Sommaire

1.1	Contexte et objectifs	1
1.2	Organisation du manuscrit	3

1.1 Contexte et objectifs

Cette thèse se déroule dans le cadre d'une CIFRE (Convention Industrielle de Formation par la REcherche¹) entre DOSIsoft² et l'équipe-projet Asclepios³ de l'INRIA Sophia Antipolis. DOSIsoft est une entreprise qui commercialise des solutions logicielles de planification des traitements radiothérapeutiques pour des cancers de localisation variée (cou, cerveau, prostate, sein, ...). Parmi ces solutions, le logiciel ISOgray permet de calculer la dosimétrie optimale pour chaque patient ainsi que la configuration correspondante des faisceaux. Ce calcul nécessite d'une part de définir et de contourer les volumes cibles et les volumes sensibles à l'irradiation, et d'autre part d'y associer des contraintes de dose, l'objectif étant de délivrer une dose thérapeutique à la tumeur tout en épargnant au maximum les structures à risque environnantes. La phase de contourage est généralement faite manuellement par le radiothérapeute.

Les travaux effectués lors de cette thèse ont été réalisés dans le contexte des cancers des voies aéro-digestives supérieures (VADS), aussi appelés cancers de la région ORL (Oto-Rhino-Laryngologie). Ils représentent une part relativement importante des cancers, en particulier chez les hommes. Pour ces cancers, le radiothérapeute doit contourer trois types de volumes :

- le volume tumoral macroscopique visible sur l'image ainsi que ses extensions microscopiques ;
- les aires ganglionnaires envahies (qui seront traitées quoi qu'il en soit) et certaines des aires ganglionnaires saines (qui seront traitées par prévention) ;
- les structures à risque environnantes : suivant la localisation de la tumeur, il peut s'agir de la moelle épinière, du tronc cérébral, des glandes salivaires (parotides et sous-maxillaires) et/ou de la mandibule.

¹<http://www.anrt.asso.fr>

²<http://www.dosisoft.com>

³<http://www-sop.inria.fr/asclepios>

Le contournage du volume tumoral est nécessairement fait manuellement par le radiothérapeute car il est par nature complètement patient-dépendant et nécessite une forte expertise médicale. En revanche, le contournage des aires ganglionnaires et des structures à risque peut être automatisé car ces volumes correspondent à des structures définies de façon identique pour tous les patients sur les images scanner. Les contours automatiques doivent ensuite être validés (et corrigés si besoin) par le radiothérapeute. Le contournage automatique permet d'une part de faire gagner du temps au radiothérapeute puisque le contournage manuel complet d'emblée est une tâche fastidieuse et longue (plusieurs heures). D'autre part, cela permet également de réduire la variabilité de contournage inter-expert puisque les contours obtenus sont reproductibles. Du point de vue applicatif, l'objectif de cette thèse est de fournir des outils de contournage automatique des aires ganglionnaires et des structures à risque de la région ORL.

Une première thèse CIFRE entre DOSIsoft et l'INRIA avait déjà été menée par Olivier Commowick. Cette thèse avait permis de définir des algorithmes de recalage non-linéaires adaptés à la sphère ORL et de construire un atlas anatomique moyen de cette région. L'utilisation de cet atlas moyen pour la segmentation automatique des structures ORL avait alors fourni des résultats prometteurs mais une légère sur-segmentation avait été constatée.

Cette thèse s'inscrit dans la suite de ces travaux. Nous avons tout d'abord cherché à optimiser la construction de l'atlas moyen, en particulier à réduire la sur-segmentation. Afin de dépasser les limites inhérentes à l'atlas moyen face à la grande variabilité anatomique inter-individuelle de la région ORL, nos recherches se sont ensuite dirigées vers la conception d'atlas anatomiques adaptés à l'anatomie du patient. Le dernier objectif de la thèse se positionne dans le contexte des soins dentaires post-irradiation et vise à fournir un algorithme de segmentation automatique des dents pour prédire *a posteriori* la dose reçue par chacune d'entre elles. Cette estimation de dose permet ensuite aux cliniciens d'estimer le niveau de risque associé à un soin dentaire donné. Enfin, cette thèse a également une finalité industrielle à travers l'intégration d'une partie des algorithmes de segmentation automatique proposés dans le logiciel ISOgray commercialisé par DOSIsoft.

Par ailleurs, les travaux réalisés dans cette thèse se sont inscrits dans le cadre de deux projets (le projet européen MAESTRO et le projet MINIARA). Ils ont aussi fait l'objet d'étroites collaborations avec plusieurs partenaires cliniques, notamment l'Université Catholique de Louvain (UCL) à Bruxelles et le Centre Antoine Lacassagne (CAL) à Nice pour la collecte d'un nombre conséquent de données (105 images manuellement segmentées pour les structures ORL, et 21 images manuellement segmentées pour les dents). De plus, nous avons collaboré avec le CHU de Tours et l'Institut Gustave Roussy (IGR) pour l'évaluation clinique de certains des algorithmes proposés.

1.2 Organisation du manuscrit

La suite du manuscrit est organisée selon les chapitres suivants.

Chapitre 2 : Contexte médical

Ce chapitre expose le contexte médical de nos recherches, à savoir l'anatomie de la région ORL, les cancers de cette région et leur traitement par radiothérapie externe.

Chapitre 3 : Segmentation par atlas

Ce chapitre présente tout d'abord un bref état de l'art de la segmentation et du recalage d'images médicales. Après avoir motivé le choix de la segmentation par atlas, nous explicitons ensuite brièvement les différentes stratégies proposées dans la littérature pour mettre en œuvre cette approche.

Chapitre 4 : Calcul d'une segmentation moyenne à partir d'une base incomplète

Nous abordons dans ce chapitre la construction d'un atlas moyen et son utilisation pour la segmentation automatique. En particulier, ce chapitre traite le problème de la construction d'une segmentation multilabel moyenne à partir d'une base d'images manuellement segmentées comportant des contours manquants. Nous présentons d'abord la sur-segmentation obtenue lorsque cette phase est faite en utilisant l'algorithme STAPLE [Warfield 2004], nous détaillons ensuite une analyse théorique de cette sur-segmentation puis nous proposons différents algorithmes pour s'en affranchir. La fin du chapitre décrit les limites de l'utilisation d'un atlas moyen.

Chapitre 5 : Algorithmes de segmentation multi-atlas adaptés au patient

Afin de dépasser les limites de l'atlas moyen, plusieurs approches de segmentation par atlas avancées ont été proposées dans la littérature. Étant donnée une base d'atlas, ces approches consistent en général à sélectionner et combiner les atlas qui sont les plus adaptés à l'anatomie du patient à segmenter. Elles peuvent aussi être basées sur une stratification de la base d'atlas en plusieurs sous-bases, et la construction d'un atlas moyen par sous-base. Ce chapitre présente un état de l'art détaillé de ces approches de segmentation par atlas avancées et plusieurs contributions sont proposées : en particulier, des nouveaux critères de sélection des atlas les plus similaires au patient et des méthodes de fusion régionale et locale des atlas sélectionnés. Nous proposons également un protocole d'évaluation non biaisée des méthodes de sélection d'atlas.

Chapitre 6 : Évaluation des algorithmes de segmentation multi-atlas adaptés au patient pour la segmentation des structures ORL

Ce chapitre propose une évaluation approfondie des algorithmes de segmentation par atlas avancés présentés dans le chapitre 5 (état de l'art et contributions). Il permet de conclure sur le choix optimal pour la segmentation des structures ORL.

Chapitre 7 : Segmentation automatique des dents pour l'évaluation des risques liés à une chirurgie dentaire post-irradiation

Ce chapitre propose un algorithme de segmentation automatique des dents dans le contexte des soins dentaires post-irradiations. Le schéma utilisé est une stratégie multi-atlas présentée dans le chapitre 5. Il fournit des contours automatiques qui peuvent ensuite être utilisés pour prédire *a posteriori* la dose reçue par chaque dent et le risque associé pour le soin dentaire envisagé. Nous évaluons notre méthode en comparant les contours automatiques aux contours manuels, et également en menant une analyse en terme d'estimation de dose. Nous évaluons aussi l'amélioration apportée par rapport à l'utilisation d'un atlas moyen.

Chapitre 8 : Conclusion

Le chapitre de conclusion présente une synthèse des contributions et les perspectives ouvertes par ces travaux de recherche.

Annexes

Enfin, trois annexes détaillent les points suivants :

- l'algorithme de recalage utilisé dans la thèse (Annexe A),
- les mesures d'évaluation de segmentation automatique utilisées dans la thèse (Annexe B),
- la variabilité de contournage manuel des structures ORL (Annexe C).

Contexte médical

Sommaire

2.1	Cancers de la région ORL	5
2.1.1	Statistiques sur les cancers ORL	5
2.1.2	Le développement du cancer	8
2.1.3	Le diagnostic d'un cancer ORL	9
2.1.4	Les traitements possibles	10
2.1.5	Les structures critiques de la région ORL	12
2.2	Radiothérapie externe	12
2.2.1	Principe de base	12
2.2.2	Radiobiologie des rayonnements ionisants	13
2.2.3	Différents types de radiothérapies externes	15
2.2.4	Planification du traitement	23
2.2.5	Objectif de la thèse dans la planification du traitement	29

Ce chapitre présente le contexte médical de la thèse. Dans un premier temps, nous présentons les cancers de la région ORL (Oto-Rhino-Laryngologie) et les traitements possibles. Ensuite, nous développons plus particulièrement les traitements par radiothérapie, qui sont à l'origine des problématiques traitées dans la thèse.

2.1 Cancers de la région ORL

Les cancers ORL regroupent une grande diversité de cancers dont la tumeur primitive se situe dans la sphère ORL, c'est-à-dire dans la zone englobant le cou et les voies aéro-digestives supérieures (VADS) (cf. figure 2.1). Chacun de ces cancers peut avoir des caractéristiques (causes, symptômes, évolution) qui sont propres à sa localisation.

2.1.1 Statistiques sur les cancers ORL

Incidence L'incidence désigne le nombre de nouveaux cas diagnostiqués dans une population donnée et sur une période donnée. Les principaux facteurs influençant l'incidence des cancers ORL sont les suivants :

- **Le sexe** : au niveau mondial, les cancers ORL représentent 7% des nouveaux cancers diagnostiqués chez les hommes, contre 3% chez les femmes [Ferlay 2010], comme le met en évidence la figure 2.2.

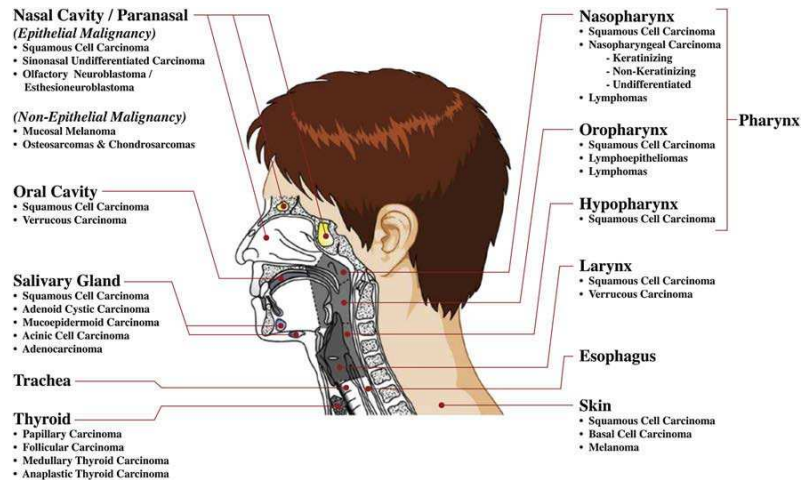


FIG. 2.1 – Diversité des cancers ORL. Source : [Stadler 2008].

- **L'âge** : le diagnostic d'un cancer ORL survient souvent entre 50 et 80 ans.
- **La consommation régulière d'alcool et de tabac.**
- **La catégorie socio-économique** : Johnson et al. ont mis en évidence l'influence de facteurs tels que le célibat, le faible niveau d'éducation et le faible revenu sur l'incidence des cancers ORL chez les hommes [Johnson 2008], une telle influence n'ayant par ailleurs pas été observée pour les cancers en général.
- **Certains papillomavirus (HPV)** : une influence a été montrée pour certains cancers de l'oropharynx.

L'incidence relative des cancers ORL par rapport aux autres cancers est différente suivant les zones géographiques, comme l'illustre la figure 2.3. En effet, les cancers ORL représentent jusqu'à plus de 8% des cancers diagnostiqués chez les hommes en Afrique et en Asie contre 4,6% en Amérique du Nord. Cette figure montre également que certains types de cancers ORL sont plus fréquents dans certaines zones géographiques que dans d'autres. L'exemple le plus frappant est celui des cancers du nasopharynx qui sont beaucoup plus fréquents dans les populations d'Asie et d'Afrique (où ils représentent 1,4% et 2% respectivement des cancers chez les hommes) que dans le reste du monde (où ils sont quasiment inexistant, avec moins de 0,2% des cancers).

Mortalité Le taux de survie des patients atteints de cancer ORL dépend fortement de deux paramètres : le stade auquel est diagnostiqué le cancer et la localisation tumorale. La figure 2.4 illustre cette double influence. Elle présente pour différentes localisations de cancers le temps médian de survie à partir du diagnostic en fonction du stade auquel le cancer a été diagnostiqué : *in situ* et *localisé* correspondent à une tumeur confinée dans l'organe concerné, *régional* caractérise un envahissement externe à l'organe, ou un envahissement des aires ganglionnaires, et *distant* désigne

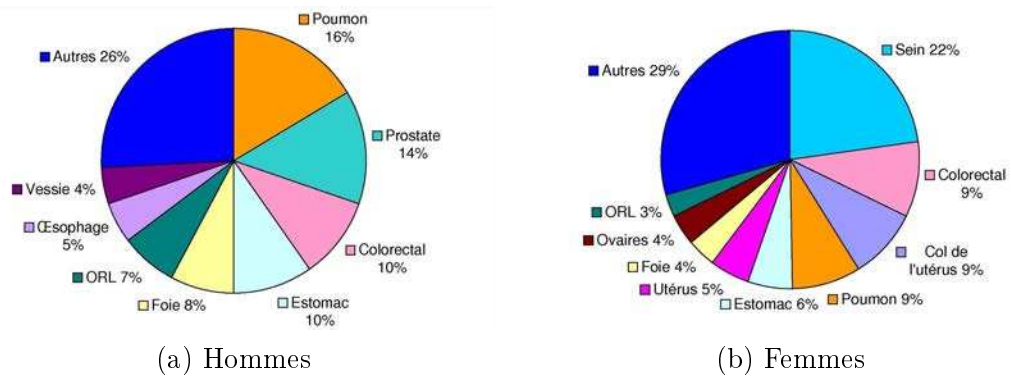


FIG. 2.2 – Incidence mondiale des principaux cancers chez l'homme (a) et chez la femme (b) (les cancers ORL incluent ici les cancers des lèvres, de la cavité orale, du pharynx et du larynx) [Ferlay 2010].

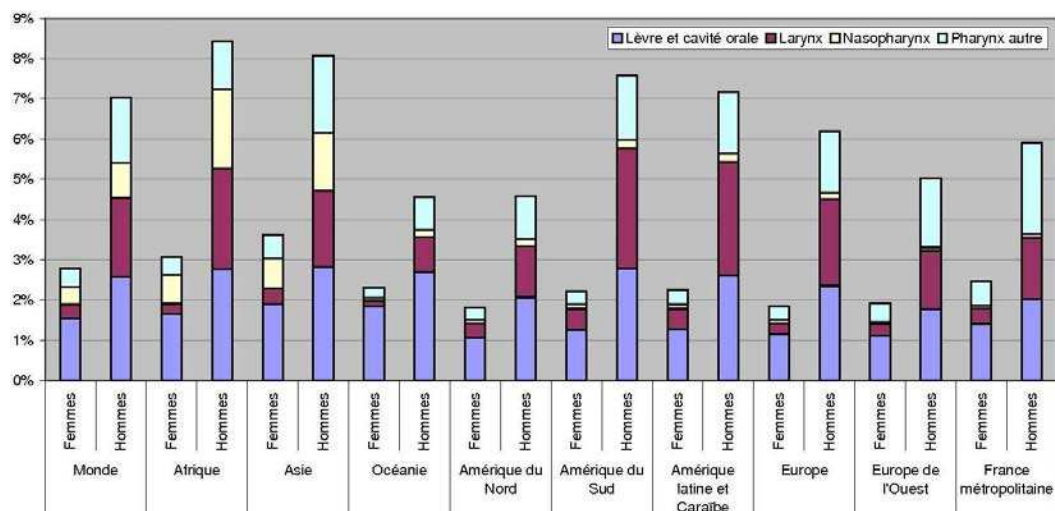


FIG. 2.3 – Incidence relative des cancers ORL par rapport à l'ensemble des cancers par région et par sexe en 2008 (en % de l'ensemble des cancers) [Ferlay 2010].

la présence de métastases à distance. Certaines études ont également montré l'influence de certains facteurs socio-économiques sur le pronostic des cancers ORL [de Graeff 2001].

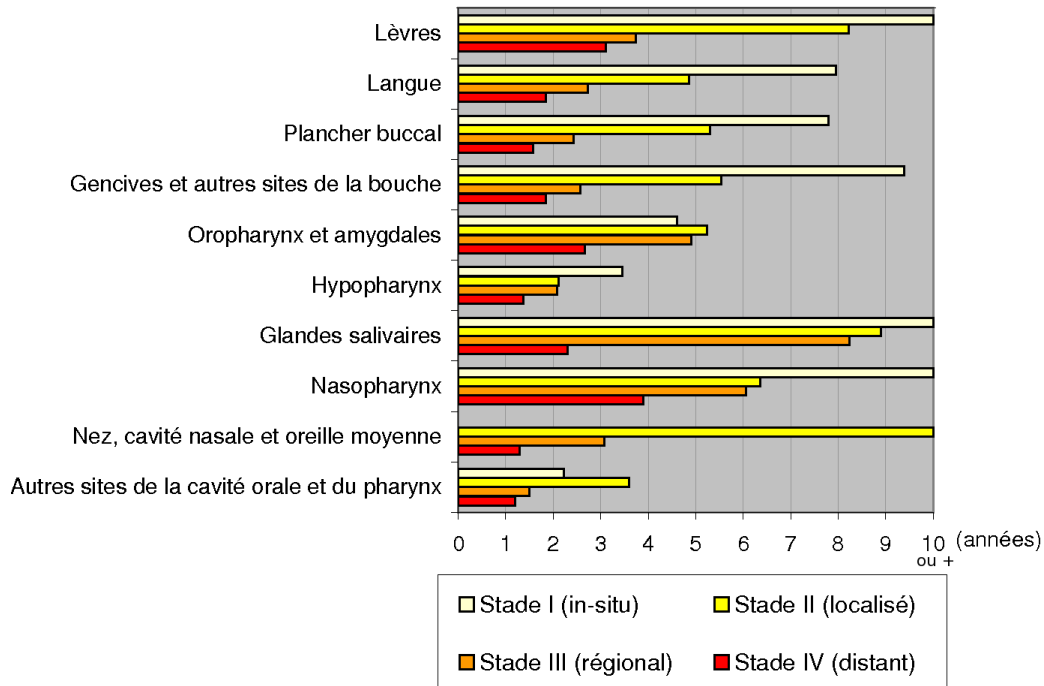


FIG. 2.4 – Temps médian de survie en années à partir du diagnostic suivant la localisation tumorale et le stade auquel a été diagnostiqué le cancer (voir texte pour les détails sur les stades). Données issues de [Piccirillo 2007].

2.1.2 Le développement du cancer

Nous présentons ici les modes de développement d'un cancer en général, quelle que soit sa localisation. A la base, les cellules cancéreuses sont issues d'une seule cellule normale qui a subi successivement plusieurs mutations spécifiques non corrigées par le système de réparation habituel. Les différentes étapes menant de la cellule normale initiale à la tumeur cancéreuse et aux éventuelles métastases sont les suivantes.

De la mutation d'une cellule normale au groupe de cellules cancéreuses

Le cycle cellulaire comprend un ensemble de points de contrôle censés vérifier l'état normal de la cellule avant sa division. Si un état anormal est détecté pour une cellule, le cycle cellulaire met en place le processus de réparation de l'anomalie. En cas d'échec de la réparation, le cycle cellulaire programme la mort de la cellule (phénomène d'apoptose). Dans certains cas, l'anomalie peut toucher des gènes qui sont

impliqués dans le fonctionnement même du cycle cellulaire, et il peut alors arriver que la cellule survive et continue à se multiplier même si la réparation a échoué. En se multipliant, cette cellule crée alors un ensemble de cellules dites *pré-cancéreuses*, qui vont à leur tour se multiplier. En se multipliant, les cellules pré-cancéreuses sont plus propices aux mutations que les cellules normales. Les mutations successives des cellules pré-cancéreuses lors de leur multiplication créent des clones de cellules pré-cancéreuses de plus en plus anormaux, et de plus en plus propices aux mutations. On considère qu'un minimum de 10 mutations en cascades est nécessaire pour transformer le groupe initial de cellules pré-cancéreuses en cellules cancéreuses (ce qui peut prendre de 5 à 20 ans).

Multiplication anarchique des cellules cancéreuses Une fois que les cellules mutées ont atteint le stade de cellules cancéreuses, elles ont acquis un certain nombre de propriétés caractéristiques. Parmi ces propriétés, elles ont acquis une indépendance vis-à-vis des signaux du cycle cellulaire (réparation et apoptose), ce qui conduit à leur multiplication anarchique et non contrôlée et permet à la tumeur de grossir jusqu'à une certaine taille (typiquement 1 millimètre). Sans vascularisation, la tumeur ne peut pas grossir davantage car elle a besoin de nutriments et d'oxygène.

Angiogénèse Au delà d'une certaine taille de tumeur (typiquement 1 millimètre), la tumeur va commencer à se vasculariser.

Croissance et propagation La tumeur ainsi vascularisée peut alors grossir davantage et peut potentiellement migrer à d'autres parties du corps via le système lymphatique et/ou le système sanguin créant ainsi des métastases.

2.1.3 Le diagnostic d'un cancer ORL

Divers symptômes peuvent indiquer un cancer ORL, notamment un changement de voix, des troubles de la déglutition, une inflammation des ganglions, une lésion buccale ou encore une toux. Devant de tels symptômes, le patient doit être envoyé vers un radiothérapeute qui prescrira des examens supplémentaires, notamment des examens d'imagerie.

A partir de ces éléments, le radiothérapeute est amené à faire son diagnostic sur la présence ou non d'un cancer ORL, et le cas échéant, il doit également diagnostiquer la localisation et le degré d'avancement de ce cancer selon la classification dite *TNM* associée à la localisation tumorale. Le système de classification *TNM* des cancers a été proposé par Denoix en 1946 et consiste à caractériser chaque cancer selon trois critères :

- la taille de la tumeur primitive (T),
- l'envahissement des ganglions lymphatiques (N) (nombre de ganglions atteints et taille des métastases),
- la présence ou non de métastases à distance (M).

	Oropharynx	Hypopharynx	Nasopharynx
T	T0 Pas de signe de tumeur primitive		
	T1 Tumeur ≤ 2 cm	T1 Tumeur ≤ 2 cm limitée à une seule sous-localisation	T1 Tumeur limitée au cavum
	T2 $2\text{cm} < \text{Tumeur} \leq 4\text{cm}$	T2 $2\text{cm} < \text{Tumeur} \leq 4\text{cm}$ ou s'étendant à plus d'une sous-localisation ou région adjacente sans fixation de l'hémi-larynx	T2 Tumeur étendue aux fosses nasales ou à l'oropharynx avec (T2b) ou sans (T2a) extension parapharyngée
	T3 Tumeur $> 4\text{cm}$	T3 Tumeur $> 4\text{cm}$ ou avec fixation de l'hémi-larynx	T3 Tumeur avec atteinte osseuse ou des sinus de la face
	T4 Tumeur envahissant les structures du voisinage (os, peau, muscle)		
N	N0 Pas d'adénopathie		
	N1 Adénopathie unique homo-latérale $\leq 3\text{cm}$		N1 Adénopathie(s) unilatérale(s) au dessus du creux sus-claviculaire et $\leq 6\text{cm}$
	N2a Adénopathie unique homo-latérale $> 3\text{cm}$ mais $\leq 6\text{cm}$		
	N2b Adénopathies homo-latérales toutes $\leq 6\text{cm}$		
	N2c Adénopathies bilatérales ou contralatérale(s) toutes $\leq 6\text{cm}$		N2 Adénopathies bilatérales au dessus des creux sus-claviculaires et $\leq 6\text{cm}$
N3 Adénopathie(s) $> 6\text{cm}$			
M	M0 Pas de métastase à distance		
	M1 Présence d'au moins une métastase à distance		

TAB. 2.1 – Classification TNM pour les tumeurs du pharynx. Le terme *adénopathie* utilisé pour la classification N signifie présence d'une métastase dans un ganglion lymphatique.

Ce système de classification existe pour la plupart des cancers, mais les échelles de classification sont différentes suivant les localisations tumorales. Pour les cancers ORL, on distingue des échelles de classification TNM différentes suivant la localisation tumorale : lèvres et cavité orale, nasopharynx, oropharynx, hypopharynx, larynx, cavité nasale et sinus, ou glandes salivaires. A titre d'exemple, le tableau 2.1 montre la classification TNM pour les tumeurs du pharynx (nasopharynx, oropharynx, hypopharynx).

2.1.4 Les traitements possibles

Différents traitements sont possibles pour soigner les cancers (traitement curatif) ou atténuer les symptômes et la douleur (traitement palliatif) :

- **Chirurgie** : ce traitement consiste à enlever tout ou partie de la tumeur (ou l'organe envahi dans certains cas) ainsi que les tissus sains environnants quand cela est possible pour être sûr d'enlever les extensions microscopiques de la tumeur. Suivant les cas, les ganglions lymphatiques voisins peuvent aussi être enlevés.
- **Radiothérapie** : ce traitement est basé sur l'utilisation thérapeutique de rayonnements ionisants au niveau de la tumeur.
- **Chimiothérapie** : ce traitement consiste à administrer au patient des substances chimiques (par voie intraveineuse dans la majorité des cas, ou par voie orale). Contrairement à la chirurgie et à la radiothérapie qui ont une action locale, la chimiothérapie permet d'agir sur les cellules cancéreuses présentes dans le corps entier.

Chacun de ces traitements peut être utilisé seul ou en association avec un autre. Par exemple, l'ablation de la tumeur par chirurgie peut être suivie d'un traitement par radiothérapie qui va détruire les extensions non visibles de la tumeur ainsi que les zones de tumeur qui n'ont pas pu être enlevées lors de la chirurgie. L'inverse est également possible (radiothérapie puis chirurgie) pour réduire la taille de la tumeur à enlever, et ainsi permettre une chirurgie conservatrice (par exemple pour le sein). Enfin, dans certains cas, la chimiothérapie et la radiothérapie peuvent être envisagées simultanément ou successivement. Le choix d'un de ces traitements ou d'une combinaison dépend de la localisation du cancer, de son état d'avancement (métastases ou pas) et de son type.

Nous nous intéressons à la radiothérapie. En radiothérapie, la quantité d'irradiation absorbée par un tissu est appelée la dose et se mesure en Gray. Elle permet de quantifier l'énergie absorbée par le tissu (en Joule) par unité de masse (en kg). L'objectif de la radiothérapie est d'administrer une dose thérapeutique sur la tumeur tout en minimisant la dose reçue par les structures à risque environnantes.

Il existe trois types différents de radiothérapie :

- **Radiothérapie externe** : c'est la plus courante. Le rayonnement ionisant est alors administré sous forme d'un faisceau externe qui va traverser le corps humain au niveau de la tumeur. Ce faisceau externe peut être généré de différentes façons suivant la nature du rayonnement. Toute la difficulté de la mise en œuvre du traitement consiste à administrer une dose thérapeutique à la tumeur, tout en épargnant autant que possible les tissus sains environnants et en particulier les structures critiques (détaillées dans le paragraphe 2.1.5).
- **Curiethérapie** : elle consiste à placer pendant une durée limitée des substances radioactives (comme l'iridium 192 ou le césium 137) soit directement à l'intérieur de la tumeur, soit au contact de la tumeur quand celle-ci se situe dans une cavité naturelle comme le vagin ou l'utérus. Elle est adaptée pour les tumeurs de petite taille bien localisées et facilement accessibles.
- **Radiothérapie métabolique** (auss appelée radiothérapie systémique) : peu utilisée, elle consiste à injecter par voie intraveineuse des médicaments radioactifs qui vont cibler les cellules tumorales.

Nous nous intéressons plus particulièrement aux techniques de radiothérapie externe et les décrivons plus en détail dans la section 2.2.

2.1.5 Les structures critiques de la région ORL

Lors des traitements par radiothérapie externe des cancers ORL, il convient d'épargner au maximum les tissus sains environnants la tumeur. Parmi ces tissus sains, certaines structures (appelées *structures à risque* ou encore *organes à risque* par abus de langage) sont particulièrement sensibles à l'irradiation :

- **la moelle épinière et le tronc cérébral**,
- **les glandes salivaires**, c'est-à-dire les glandes produisant la salive (les plus grosses étant les glandes parotides et sous-maxillaires) : elles sont très sensibles à la radiothérapie. Leur irradiation peut causer un assèchement de la bouche (xérostomie) à plus ou moins long terme, qui peut se répercuter par des complications dentaires ;
- **la mandibule et les dents** : leur irradiation entraîne un risque accru d'ostéoradionécrose (nécrose osseuse). Nous détaillons ces risques dans le chapitre 7 sur la segmentation des dents.

Le risque pour ces organes étant un effet secondaire à long terme, ils sont donc pris en compte dans la planification du traitement de façon à être épargnés autant que possible. Leur sensibilité à l'irradiation est différente : pour certains organes (dits organes *en parallèle*), la fonction de l'organe est préservée si un volume suffisant de l'organe est épargné (par exemple, les parotides) ; pour d'autres (dits organes *en série*), la perte de fonction de l'organe survient dès lors qu'une partie de l'organe a reçu une dose supérieure à la dose critique (par exemple la moelle, avec le risque de tétraplégie). D'autres structures sont sujettes à des effets secondaires à court terme difficilement évitables, comme les muqueuses de la cavité orale et du tube digestif qui sont très souvent inflammées pendant le traitement et les jours/semaines qui suivent.

2.2 Radiothérapie externe

2.2.1 Principe de base

La radiothérapie externe consiste à irradier la tumeur avec des faisceaux externes de rayonnements ionisants pour bloquer la multiplication anarchique des cellules cancéreuses et pour les détruire, tout en épargnant au maximum les cellules saines environnantes. L'utilisation d'un seul faisceau ne permet pas de délivrer une dose thérapeutique à la tumeur tout en épargnant les tissus sains puisqu'en général la tumeur se situe à une certaine profondeur. Toutes les techniques de radiothérapie externe utilisent donc plusieurs faisceaux d'orientations différentes et de plus faible intensité qui s'intersectent tous au niveau de la tumeur. Ainsi, la tumeur reçoit les doses déposées par tous les faisceaux, alors que les tissus sains ne sont *a priori* traversés que par un seul faisceau et donc reçoivent une dose beaucoup plus faible.

Ceci est d'autant plus vrai que le nombre de faisceaux utilisés est élevé car la dose par faisceau est alors d'autant plus faible.

Le paragraphe 2.2.2 décrit brièvement l'effet des rayonnements ionisants sur les cellules cancéreuses et saines. Le paragraphe 2.2.3 détaille les différentes techniques de radiothérapie externe. Enfin, le paragraphe 2.2.4 décrit les différentes étapes de la planification d'un traitement de radiothérapie externe.

2.2.2 Radiobiologie des rayonnements ionisants

2.2.2.1 Action des rayonnements ionisants sur la matière vivante

Le terme de *rayonnement ionisant* désigne tout type de rayonnement capable de produire des ionisations durant son absorption par les tissus, c'est-à-dire capable d'arracher des électrons aux atomes des tissus. Lorsqu'un faisceau de rayonnements ionisants traverse la matière, il va transférer son énergie aux cellules sur son parcours par différents processus selon sa nature, son énergie et les atomes rencontrés. Les étapes reliant le dépôt primaire d'énergie à l'échelle atomique à ses conséquences biologiques et médicales à l'échelle cellulaire et tissulaire sont les suivantes :

- **Étape physique** : le rayonnement primaire ionise la matière de façon directe (s'il est constitué de particules chargées) ou indirecte (s'il est constitué de particules neutres). La durée de cette étape est inférieure à une picoseconde, on l'appelle l'étape *pico*.
- **Étape physico-chimique** : les ionisations créés dans l'étape physique vont induire des cassures de liaisons chimiques, et donc la création de nouvelles espèces chimiques. Certaines des espèces créées vont interagir avec l'eau pour créer des radicaux libres (espèce chimique instable ayant un électron non apparié sur son orbite externe). Cette étape dure moins d'une nanoseconde, c'est l'étape *nano*.
- **Étape biochimique** : les ionisations (effet direct) ainsi que les radicaux libres (effet indirect) vont créer des lésions de l'ADN de la cellule, notamment des cassures mono-brin (concerne un seul brin de l'ADN, représente plus de 99% des lésions) et double-brin (concerne simultanément les deux brins de l'ADN au même endroit). La durée de cette étape est de l'ordre de la seconde.
- **Étape biologique** : la cellule va mettre en place un processus de réparation des lésions. La réussite de la réparation est dépendante de la phase du cycle cellulaire dans laquelle se trouve la cellule et de la présence des enzymes de réparation. Si les conditions sont réunies, la réparation des lésions mono-brin est fidèle (l'ADN après réparation est exactement le même qu'avant la lésion) et rapide (moins d'une demi-heure) car la cellule est habituée à réparer ce genre de lésions puisqu'elles peuvent être induites naturellement. En revanche, les lésions double-brin sont plus difficilement réparables. En effet, le processus de réparation est plus long (environ 12 heures) et il n'est pas infaillible : dans tous les cas, la réparation n'est pas fidèle (l'ADN est différent après réparation), mais elle peut également être incomplète (auquel cas la cellule

survit mais avec des mutations) ou carrément échouer (auquel cas la cellule meurt de façon immédiate ou programmée). Les lésions létales pour la cellule sont donc les lésions double-brin, les lésions mono-brin étant qualifiées de lésions sublétales (c'est-à-dire pouvant provoquer la mort cellulaire en cas d'accumulation). Cette étape dure plusieurs heures.

- **Étape biomédicale** (au niveau tissulaire) : l'accumulation des effets au niveau cellulaire contribue à créer un effet au niveau tissulaire. Cette étape est beaucoup plus longue (de l'ordre de plusieurs semaines).

2.2.2.2 Radiosensibilité relative des cellules saines et tumorales

Le paragraphe précédent a mis en évidence les mécanismes par lesquels les rayonnements ionisants peuvent induire la mort cellulaire, que ce soit sur les cellules saines ou cancéreuses. Cela dit, cet effet des radiations n'a pas la même ampleur sur toutes les cellules. Chaque cellule a une radiosensibilité intrinsèque qui dépend de son type.

D'une manière générale, les cellules tumorales sont plus radiosensibles que les cellules saines, et c'est ce qui contribue à l'efficacité des traitements radiothérapeutiques. Les principales raisons expliquant la radiosensibilité accrue des cellules cancéreuses par rapport aux cellules saines sont les suivantes :

- De façon générale, les cellules cancéreuses se multiplient plus vite que les cellules saines. Elle se trouvent donc plus souvent en phase de mitose, et les cellules saines plus souvent en phase de repos cellulaire (G0). La différence de radiosensibilité entre les cellules saines et tumorales s'explique alors par le fait qu'une cellule est généralement peu radiosensible en phase de repos cellulaire mais très radiosensible en phase de mitose. Ceci rejoint en fait la loi de Bergonié et Tribondeau (1906) selon laquelle les cellules qui se divisent plus rapidement sont plus sensibles aux radiations.
- Les cellules tumorales arrivent moins bien que les cellules saines à réparer les lésions d'ADN car cette faculté a été altérée lors des mutations en cascade subies par les clones de cellules tumorales (cf. section 2.1.2), le dysfonctionnement pouvant par exemple empêcher ou perturber la production d'enzymes. Ceci est valable pour les lésions mono-brin et double-brin.

Cela dit, cet effet différentiel de radiosensibilité entre cellules saines et tumorales est à nuancer. D'une part, certains types de cellules tumorales sont intrinsèquement radorésistantes. C'est le cas en particulier des cellules hypoxiques, c'est-à-dire peu ou pas oxygénées à cause d'une vascularisation insuffisante de la tumeur. D'autre part, les tissus sains à renouvellement rapide (typiquement la peau ou les muqueuses) sont relativement radiosensibles (mais contrairement aux cellules tumorales, leur système de réparation des lésions ADN est parfaitement opérationnel).

Ce paragraphe a décrit les mécanismes contribuant à créer un effet différentiel entre les cellules saines et tumorales au niveau cellulaire. Nous verrons ensuite comment les développements technologiques en radiothérapie permettent d'accentuer cet effet différentiel au niveau tissulaire, notamment en conformant l'irradiation à la tumeur (cf. 2.2.3) et en fractionnant le traitement en plusieurs séances (cf. 2.2.4.4).

2.2.3 Différents types de radiothérapies externes

Les traitements par radiothérapie externe englobent différents types de rayonnements auxquels correspondent différentes technologies pour les administrer :

- **Les photons X (et gamma)** : les photons gamma sont créés par la désintégration d'une source radioactive, comme le cobalt. Pour les photons X, des accélérateurs linéaires (cf. figure 2.5) permettent d'accélérer des électrons qui sont ensuite freinés dans une couche fine de tungstène, ce qui libère des photons X d'énergie comprise entre 4 et 25 MeV. C'est dans ce contexte que se situe le travail de thèse. Ces traitements, qui sont de loin les plus courants, sont donc expliqués plus en détail dans la suite.
- **Les électrons** : ils sont utilisés pour les tumeurs superficielles car du fait de leur forte absorption dans la matière, ils transfèrent toute leur énergie dans les premiers centimètres (jusqu'à 2 cm pour des électrons de 6 MeV, et 5 cm pour 15 MeV). Ils sont produits par un accélérateur linéaire.
- **Les protons et les ions carbone** (hadronthérapie) : ils sont utilisés pour les tumeurs très superficielles ou pour les tumeurs très proches d'organes à risque sensibles à l'irradiation (par exemple tumeur de l'œil ou de la base du crâne). Généralement produits par un cyclotron, ils ont la particularité de déposer leur énergie en fin de parcours sous forme du très étroit pic de Bragg, ce qui permet de très bien épargner les tissus sains avant et après la tumeur. Pour cette raison, ils sont aussi utilisés chez les enfants pour limiter les cancers induits.
- **Les neutrons** : ils ont l'avantage d'avoir une efficacité biologique relative trois fois supérieure à celles des photons (c'est-à-dire 15 Gray de neutrons produisent le même effet biologique que 45 Gray de photons), mais ils ont une toxicité supérieure à celle des photons. Ils sont issus de cyclotrons. En pratique, ils ne sont quasiment plus utilisés à cause des effets secondaires importants.

Le choix du type de rayonnement et de son énergie dépend généralement de l'organe impliqué et de la profondeur de la tumeur.

Nous allons maintenant détailler successivement les différents types de radiothérapie externe utilisant les photons X (ou gamma dans certains cas) présentés sur la figure 2.6.

2.2.3.1 Radiothérapie conventionnelle

Les premiers accélérateurs linéaires datent des années 1960. A cette époque, le nombre de faisceaux utilisés est limité (quatre au maximum) et des caches en plomb permettent de modifier la forme des faisceaux, mais dans tous les cas les champs d'irradiation restent carrés ou rectangulaires. C'est la radiothérapie dite *conventionnelle*. Si les faisceaux sont coplanaires, on parle de radiothérapie conventionnelle 2D ; sinon, on parle de radiothérapie conventionnelle 3D.



FIG. 2.5 – Accélérateur linéaire. Modèle ONCOR de Siemens. Source : <http://www.medical.siemens.com>.

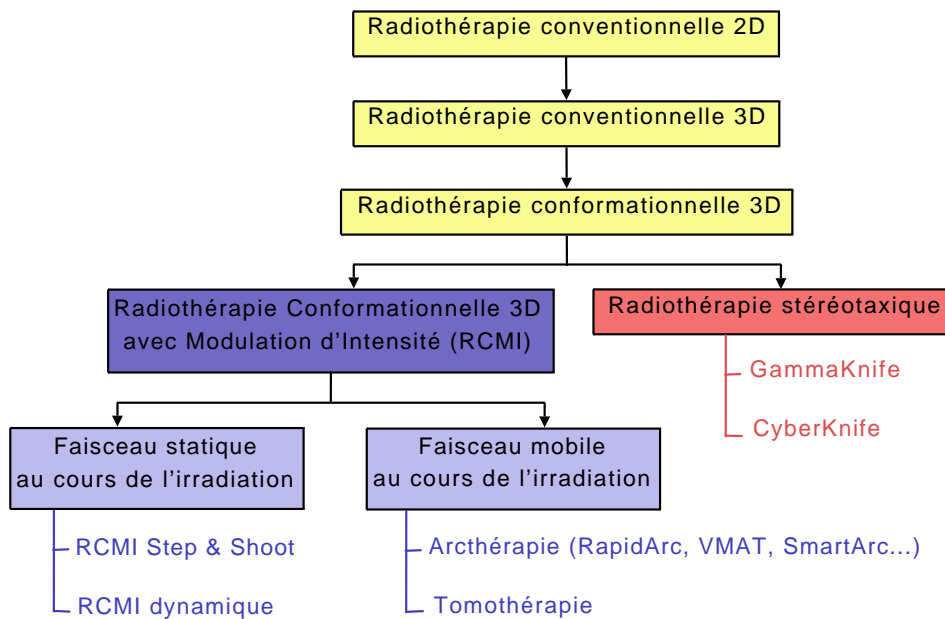


FIG. 2.6 – Différents types de radiothérapie externe utilisant les photons X ou gamma.

2.2.3.2 Radiothérapie conformationnelle

Le principal inconvénient de la radiothérapie conventionnelle est qu'il est difficile d'épargner les organes à risque de l'irradiation en se limitant à des champs carrés ou rectangulaires. L'objectif des développements technologiques en radiothérapie a donc été de focaliser au maximum la dose sur la tumeur en affinant la restriction anatomique de l'irradiation. Cela est doublement intéressant : cela permet de diminuer la toxicité pour les organes à risque, et par conséquent cela permet aussi de pouvoir augmenter la dose administrée à la tumeur (c'est l'escalade de dose).

La radiothérapie conformationnelle consiste à donner des formes non rectangulaires/carrées aux faisceaux pour qu'ils se conforment mieux à la forme de la tumeur, et à augmenter le nombre de faisceaux (typiquement 5 à 9). Les principes de la radiothérapie conformationnelle (épargner les structures à risque en utilisant en modulant la forme des faisceaux) étaient connus dès le milieu du XX^{ème} siècle mais les caches en plomb alors utilisés pour tronquer les champs d'irradiation ne permettaient pas d'obtenir des champs de forme complexe autre que rectangulaire ou carrée (et en plus ils devaient être fabriqués sur mesure). La radiothérapie conformationnelle apparaît dans les années 1990, avec l'avènement des accélérateurs linéaires avec collimateurs multi-lames (MLC) (cf. figure 2.7). Le collimateur multi-lames est constitué de fines lames de tungstène de 1 à 1.5 mm d'épaisseur qui peuvent bouger les unes par rapport aux autres pour créer des trous de formes complexes. Un collimateur multi-lames est placé sur le chemin de chacun des faisceaux de rayons X pour moduler sa forme.

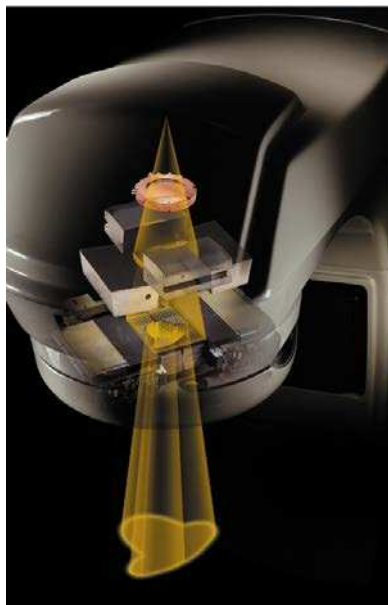


FIG. 2.7 – Le collimateur multi-lames (MLC, pour MultiLeaf Collimator) permet de moduler la forme du faisceau de façon complexe. Source : <http://varian.mediaroom.com>.

2.2.3.3 Radiothérapie conformationnelle avec modulation d'intensité (abrégée RCMI en français ou IMRT en anglais)

Principe En radiothérapie conformationnelle 3D standard, les collimateurs multi-lames ne servent qu'à fixer la forme de chaque faisceau, et cette forme reste la même pour toute la durée de l'irradiation. En radiothérapie conformationnelle avec modulation d'intensité (abrégée RCMI ou IMRT pour Intensity-Modulated Radiation Therapy), les lames des collimateurs bougent dynamiquement et permettent de faire varier la forme de chaque faisceau en fonction du temps. En fait, le nom de la technique est trompeur car ce n'est pas l'intensité (énergie par unité de surface et de temps, $\frac{\text{Joule}}{\text{sec.m}^2}$) qui est modulée (comme le nom le suggère) mais la fluence (énergie par unité de surface, $\frac{\text{Joule}}{\text{m}^2}$). La modulation de la fluence de chaque faisceau a deux avantages principaux :

- D'une part, elle permet de sculpter des isodoses de forme complexe qui peuvent se conformer à la tumeur quelle que soit sa forme (contrairement à la radiothérapie conformationnelle standard qui ne pouvait pas, par principe, se conformer aux tumeurs concaves). Cela permet d'épargner encore mieux les organes à risque, et du coup de délivrer une dose encore supérieure à la tumeur vu qu'elle est mieux ciblée. Le rapport de la dose délivrée à la tumeur sur celle délivrée aux organes à risque est amélioré.
- De plus, la modulation de la fluence permet de délivrer une dose homogène à la tumeur. En effet, supposons qu'une partie du faisceau doive traverser une zone dense (os) avant d'atteindre la tumeur. Sans modulation de fluence, la dose délivrée à la tumeur par cette partie du faisceau sera plus faible qu'aux autres endroits à cause de l'absorption de l'os. Ce contraste de dose au niveau de la tumeur disparaît avec la RCMI car alors la fluence peut être augmentée localement pour compenser l'absorption de l'os par cette partie du faisceau.

Historique Avec la radiothérapie conformationnelle standard, le mode de planification était dit *direct* (forward planning) : on fixe tous les paramètres de configuration des faisceaux (par expertise du radiophysicien) et on calcule la carte 3D de dose résultante en cumulant les contributions de tous les faisceaux. Si la carte de dose ne satisfait pas les contraintes de dose de la prescription, alors on réitère la procédure en changeant légèrement les paramètres de la balistique, et cela jusqu'à avoir une dosimétrie satisfaisante. Ce mode de planification directe est difficile à mettre en œuvre avec la RCMI car le nombre de paramètres à optimiser est trop important pour arriver à un traitement acceptable en un nombre raisonnable d'itérations. A la fin des années 1990, la planification *inverse* (inverse planning) est intégrée dans les principaux logiciels de planification de traitements, ce qui permet également les débuts de la RCMI (principalement sur fantôme dans un premier temps). La planification inverse en RCMI consiste à partir des positions des faisceaux et des objectifs cliniques (les contraintes de dose) et à résoudre un problème inverse pour déterminer les modulations dynamiques de forme de chaque faisceau qui permettent de respecter les contraintes de dose, ainsi que les mouvements des lames correspondants.

Mise en œuvre de la RCMI La façon classique de délivrer la RCMI consiste à utiliser plusieurs (5 à 9) faisceaux statiques d'orientations différentes (non-coplanaires si besoin), et à moduler la fluence de chaque faisceau selon l'une des deux méthodes suivantes :

- **La technique dite de *Step and Shoot***, pour laquelle l'irradiation et le mouvement des lames des collimateurs se font de façon séquentielle. Pour chaque faisceau, les lames des collimateurs bougent quand le faisceau est éteint. Ensuite, le faisceau est rallumé et l'irradiation a lieu quand les lames sont immobiles. Puis le faisceau est de nouveau éteint et les lames changent à nouveau de configuration, etc.
- **La RCMI dynamique** pour laquelle les lames des collimateurs bougent en continu pendant l'irradiation pour chaque faisceau.

Pour ces deux techniques, l'accélérateur linéaire se place successivement en position pour chacun des 5 à 9 faisceaux, mais il reste statique pendant la durée de l'irradiation pour chaque position de faisceau. Au contraire, d'autres types de traitement par RCMI combinent simultanément une rotation continue et une modulation de forme du faisceau d'irradiation. L'objectif de ces techniques est d'augmenter le nombre d'angles d'irradiation pour obtenir une meilleure conformation de dose à la forme de la tumeur. Parmi ces techniques, on distingue deux cas :

- **La technique *IMAT* (Intensity-Modulated Arc Therapy)** : introduite par Yu et al. en 1995 [Yu 1995, Yu 2002], l'arc-thérapie est en fait une RCMI dynamique combinée avec une rotation continue du bras de l'accélérateur linéaire délivrant le faisceau. La forme du faisceau 3D est modulée en continu grâce à des collimateurs multi-lames classiques tout au long de l'irradiation et de la rotation du bras, et la vitesse de rotation du bras est variable au cours de l'irradiation. Le principal avantage de cette technique est qu'elle réduit considérablement la durée du traitement (4 minutes au lieu de 20 minutes pour une RCMI standard). De plus, elle peut être délivrée avec la même machine que les traitements de RCMI standards (à savoir un accélérateur linéaire équipé de collimateurs multi-lames), et ne représente donc pas un coût supplémentaire pour les hôpitaux. Elle est commercialisée par Varian (RapidArc[®]) depuis 2007, par Elekta (VMAT[®]) et par Philips (SmartArc[®]).
- **La tomothérapie** : au moyen d'un accélérateur linéaire monté sur un anneau de scanner CT, un faisceau très fin tourne en continu autour du patient et délivre l'irradiation coupe par coupe. Pour que la totalité du volume cible soit irradiée, il y a deux solutions : (1) Tomothérapie coupe par coupe : le faisceau décrit une rotation simple de 360° autour du patient (appelée *arc*) permettant d'irradier une seule coupe, et le patient est déplacé longitudinalement entre chaque arc de façon à traiter successivement chaque coupe ; (2) Tomothérapie hélicoïdale : le mouvement longitudinal du patient se fait de façon simultanée à la rotation du faisceau, le faisceau décrivant ainsi un mouvement hélicoïdal par rapport au patient. Dans les deux cas, des collimateurs multi-lames binaires (chaque lame est soit ouverte soit fermée) ou standards sont utilisés pour moduler la forme du faisceau à chaque angle d'irradiation.

Avantages et inconvénients de la RCMI La RCMI a l'avantage de permettre une meilleure conformation de la dose à la tumeur par rapport à la radiothérapie conformationnelle standard. Cela permet de mieux épargner les organes à risque de l'irradiation, et donc aussi d'augmenter la dose délivrée à la tumeur (c'est le phénomène d'escalade de dose). La figure 2.8 montre pour un volume cible donné (en jaune), et un organe à risque proche (tronc cérébral, en rose) les isodoses obtenues avec la RCMI en comparaison de celles obtenues avec la radiothérapie conventionnelle, et avec la radiothérapie conformationnelle sans modulation d'intensité.

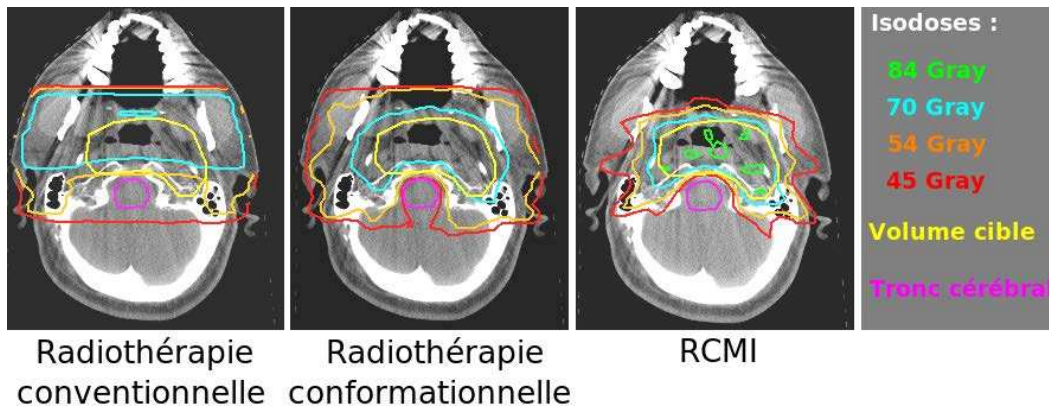


FIG. 2.8 – Exemple d'isodoses obtenues par un traitement RCMI (droite) en comparaison d'un traitement par radiothérapie conventionnelle (gauche) et d'un traitement par radiothérapie conformationnelle sans modulation d'intensité (milieu). Les isodoses se conforment bien à la forme du volume cible (en jaune), et épargnent le tronc cérébral (en rose).

De nombreuses études ont montré que la RCMI donne de meilleurs résultats que la radiothérapie conformationnelle standard, en particulier quand il y a des organes à risque très proches de la tumeur. C'est le cas notamment des cancers de l'oropharynx [Turaka 2010, Terhaard 2005] et du nasopharynx [Wu 2009, Hsiung 2006, Kam 2003] pour lesquels la RCMI permet de réduire considérablement l'irradiation des parotides, et de diminuer le phénomène de xérostomie (bouche sèche).

Cependant, les gradients de dose abrupts permis par la RCMI représentent également un inconvénient dans le sens où ils rendent la technique plus sensible aux erreurs géométriques que les traitements 2D ou 3D conventionnels. Pour cette raison, la RCMI requiert donc une précision accrue, notamment par rapport :

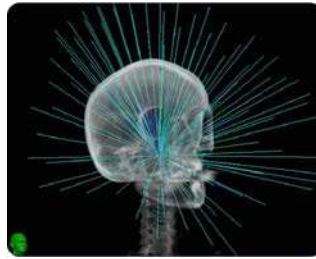
- au positionnement et à l'immobilisation du patient, en particulier pour les localisations anatomiques qui sont amenées à bouger avec la respiration (poumon, foie) ;
- au contournage des volumes cibles et des organes à risque (l'utilisation d'autres modalités d'imagerie (IRM, TEP) et de la fusion d'images peut aider [Emami 2003]).

2.2.3.4 La radiothérapie stéréotaxique

La radiothérapie stéréotaxique consiste à irradier la tumeur avec un très grand nombre de mini-faisceaux convergents dans le but d'obtenir une nécrose localisée, comme illustré sur la figure 2.9(a). Elle peut être vue comme une forme particulière de RCMI puisqu'elle permet d'atteindre une très bonne conformation de la dose à la tumeur, mais le degré de précision atteint est encore meilleur (de l'ordre du millimètre). Elle est recommandée pour les tumeurs de petite taille non superficielles de forme relativement sphérique. Cependant, ces technologies sont assez peu implantées dans les hôpitaux. Les deux principales techniques de radiothérapie stéréotaxique sont le Gamma Knife et le CyberKnife[®]. Elles sont illustrées sur la figure 2.9 et détaillées dans la suite.

Le Gamma Knife Le Gamma Knife est une machine de radiochirurgie réservée aux tumeurs intracrâniennes, et son invention date des années 1980. Un casque avec des trous est placé sur la tête du patient, et le patient muni de ce casque est positionné sur la table de traitement. En position de traitement, les trous du casque laissent passer 201 faisceaux très fins de rayons gamma qui s'intersectent au niveau de la tumeur. Ces rayons gamma proviennent de la désintégration de sources de cobalt 60, comme le montre la figure 2.9(c). En général, la dose délivrée est de l'ordre de 12 à 25 Gy, mais peut atteindre 60-70 Gy pour certaines localisations (thalamus). Une seule séance de traitement est nécessaire. Seuls deux centres sont équipés de Gamma Knife en France (Marseille et Lille).

Le CyberKnife[®] Contrairement au Gamma Knife qui est spécifique au crâne, le CyberKnife[®] permet de traiter le corps entier. Il est constitué d'un accélérateur linéaire miniaturisé fonctionnant à une fréquence environ trois fois supérieure à celle d'un accélérateur linéaire standard. L'accélérateur miniaturisé est monté sur un bras robotisé afin d'atteindre tous les angles possibles. Un système d'imagerie fluoroscopique est également inclus pour localiser et suivre en continu le volume cible à irradier (et notamment prendre en compte son éventuel mouvement), et renvoyer au bras motorisé les informations correspondantes. Le CyberKnife[®] fait donc partie des systèmes de radiothérapie guidés par l'image (IGRT pour Image-Guided Radiation Therapy), contrairement aux autres systèmes de radiothérapie standard pour lesquels l'imagerie ne sert qu'à positionner le patient au début de la séance, et à vérifier le positionnement à la fin de la séance. Le CyberKnife[®] est donc particulièrement approprié pour les tumeurs mobiles (poumons, foie). Plus récent que le Gamma Knife, il reste peu implanté dans les hôpitaux, avec 100 à 150 implantations dans le monde, et quatre en France (Nancy, Nice, Lille, Tours). La figure 2.9(b) montre un appareil de CyberKnife[®].

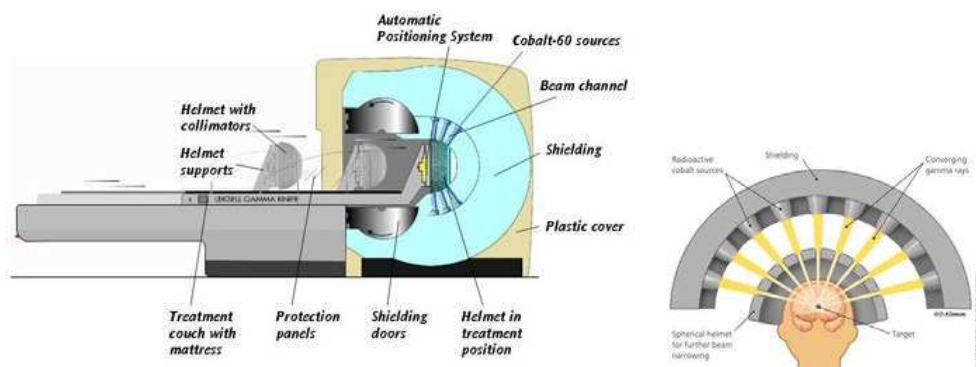


(a) Configuration des faisceaux en radiothérapie stéréotaxique



1. Sources d'imagerie fluoroscopique
2. Détecteurs associés
3. Accélérateur linéaire miniaturisé
4. Table

(b) CyberKnife®



(c) Gamma Knife

FIG. 2.9 – Radiothérapie stéréotaxique. Sources : (a) <http://www.cyberknife.com>, (b) <http://www accuray.com>, (c) <http://www.carolinaradiation.com> et <http://www.aafp.org>.

2.2.4 Planification du traitement

La planification d'un traitement en radiothérapie conformationnelle 3D standard ou RCMi commence par la réalisation d'un examen d'imagerie scanner CT (Computed Tomography), appelé le scanner dosimétrique car il servira à calculer la dosimétrie du plan de traitement. A partir de ce scanner dosimétrique, les étapes sont les suivantes :

1. Contourage du volume cible (section 2.2.4.1) et des organes à risque (section 2.2.4.2) sur le scanner dosimétrique.
2. Définition des contraintes de dose pour les volumes contourés sur le scanner dosimétrique (section 2.2.4.3).
3. Prescription de fractionnement et d'étalement du traitement (section 2.2.4.4).
4. Choix de la balistique et calcul de la dosimétrie (section 2.2.4.5).

2.2.4.1 Contourage du volume cible sur le scanner dosimétrique

Le radiothérapeute doit tout d'abord contourer le volume cible (c'est-à-dire l'ensemble du volume à irradier) sur l'examen de scanner dosimétrique. Le volume cible ne se limite pas seulement aux tissus cancéreux visibles sur l'image scanner (GTV), mais inclut également les extensions microscopiques de la tumeur et les zones de dissémination possible (CTV) ainsi que des « marges de sécurité » (PTV). La nomenclature détaillée des différents volumes cibles est la suivante :

Le GTV (Gross Tumor Volume) Il désigne la tumeur macroscopique visible sur l'image scanner et/ou palpable à l'examen (GTV tumoral) ainsi que les aires ganglionnaires atteintes, c'est-à-dire celles dont les ganglions présentent des métastases (GTV ganglionnaire ou métastatique).

Le CTV (Clinical Target Volume) Il inclut :

- le GTV (tumoral et ganglionnaire),
- les possibles extensions microscopiques de la tumeur non visibles à l'imagerie scanner mais connues par le radiothérapeute,
- certaines aires ganglionnaires non atteintes (par prévention).

Selon le site tumoral et le stade, il peut en effet être nécessaire de traiter les aires ganglionnaires même si elles ne sont pas atteintes car elles correspondent aux zones de dissémination préférentielle des tumeurs ORL. Grégoire et al. ont proposé des indications sur les aires ganglionnaires à irradier suivant la localisation et le stade de la tumeur primitive pour le cas des cancers N0 (c'est-à-dire quand aucun ganglion n'est atteint) [Grégoire 2000]. Historiquement, l'anatomie du système lymphatique cervical a été décrite par Rouvière [Rouvière 1948]. Les aires ganglionnaires sont divisées en plusieurs niveaux de chaque côté (droite/gauche) : I, II, III, IV, V et VI. Certaines d'entre elles sont encore divisées en sous-niveaux (par exemple IA et IB). Les aires ganglionnaires ne correspondent pas à des structures

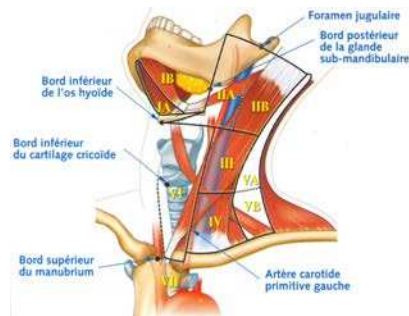


FIG. 2.10 – Les niveaux ganglionnaires et leurs limites anatomiques. Source : [Benoudiba 2007].

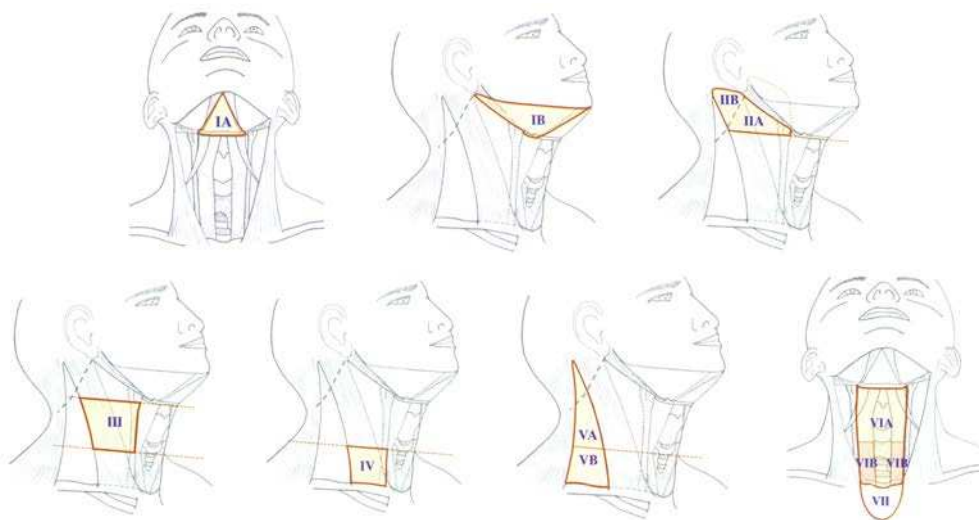


FIG. 2.11 – Les niveaux ganglionnaires et leurs limites anatomiques, niveau par niveau. Par exemple, la coupe axiale séparant les aires III et IV est définie comme étant la limite crâniale du cartilage cricoïde. Source : [Benoudiba 2007].

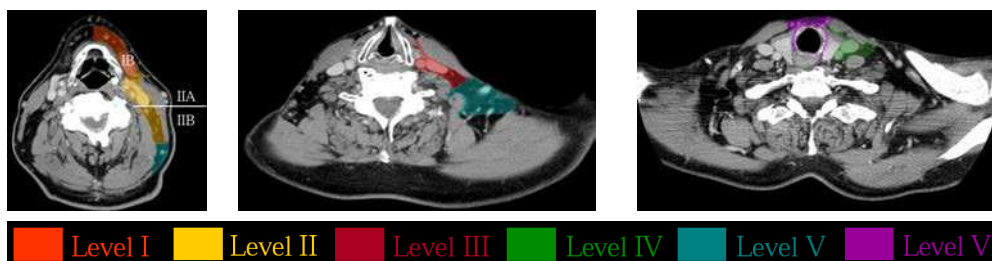


FIG. 2.12 – Coupes axiales montrant les contourages des aires ganglionnaires sur une image scanner. Source : <http://www.rtog.org>.

anatomiques à proprement parler mais plutôt à des zones et sont donc difficilement discernables sur l'image pour un novice. Leur contourage est basé sur le repérage de points anatomiques précis tels que des muscles ou des cartilages. Suite à diverses études et propositions sur le contourage des aires ganglionnaires pour la radiothérapie, un consensus a été proposé [Grégoire 2003, Levendag 2004] afin d'uniformiser les différents protocoles. Ce consensus présente les repères anatomiques définissant les limites crâniennes/caudales, latérales/médianes, ventrales/dorsales nécessaires au contourage des aires ganglionnaires IA, IB, IIA, IIB, III, IV, V et VI. Les figures 2.10 et 2.11 mettent en évidence certains de ces repères anatomiques, et la figure 2.12 montre des exemples d'images scanner sur lesquelles les aires ganglionnaires ont été contourées.

Le PTV (Planning Target Volume) Il inclut le CTV ainsi que des marges de sécurité liées aux imprécisions géométriques. Ces imprécisions géométriques peuvent être dues au positionnement du patient ou à son mouvement pendant l'irradiation (par exemple à cause de la respiration). En pratique, le radiothérapeute prend également soin dans la mesure du possible de ne pas inclure les organes à risque environnants dans le contourage du CTV et du PTV, surtout ceux qui sont très sensibles à l'irradiation comme la moelle épinière. Cela peut conduire localement à l'absence de marge de sécurité dans la définition du PTV.

2.2.4.2 Contourage des organes à risque sur le scanner dosimétrique

Afin de prendre en compte la position des organes à risque pour les préserver au maximum de l'irradiation, le radiothérapeute doit les contourer sur le scanner dosimétrique. Pour les cancers de la région ORL, les organes à risque généralement pris en compte sont la moelle épinière, le tronc cérébral, les glandes salivaires (parotides et sous-maxillaires) et la mandibule. Dans la plupart des cas, d'autres structures à préserver sont également contourées (cavité buccale, lèvres, palais, ...) dans le but d'imposer des contraintes supplémentaires pour forcer le logiciel à conformer la dose.

2.2.4.3 Définition des contraintes de dose pour les volumes contourés sur le scanner dosimétrique

Le radiothérapeute doit ensuite définir la dose minimale que doit recevoir chacun des volumes cibles, et la dose maximale acceptable que peut recevoir chacun des organes à risque environnants. Ces doses limites se fixent au cas par cas suivant les caractéristiques, la localisation et l'avancement de la tumeur. Des recommandations ont cependant été proposées pour guider les prescriptions de doses, notamment par le RTOG (Radiation Therapy Oncology Group) sur le site www.rtog.org. En général, le radiothérapeute doit également fixer un ordre de priorité ou un système de pondération des différentes prescriptions de dose (par exemple : (1) en priorité respecter les doses limites de la moelle épinière et du tronc cérébral, (2) ensuite, respecter les prescriptions de dose sur les volumes cibles, (3) ensuite, respecter les

limites de dose de l'oropharynx, puis de la parotide controlatérale à la tumeur, puis de l'œsophage, puis des lèvres, etc).

Prescription de dose pour les volumes cibles Les doses prescrites sur les volumes cibles sont variables suivant le volume cible considéré, mais sont généralement comprises entre 50 à 70 Gray pour un traitement de radiothérapie à visée curative. La prescription pour le GTV ainsi que les aires ganglionnaires métastatiques est souvent de l'ordre de 70 Gy (ou comprise entre 60 et 70 Gy), alors que la prescription pour les aires ganglionnaires non métastatiques à risque est plus proche des 50-60 Gy (typiquement 54 ou 60 Gy) [RTOG 2004, RTOG 2005, RTOG 2010]. Si la radiothérapie est à visée palliative, ou si elle est combinée avec une chirurgie (avant ou après), alors la dose prescrite pour les volumes cibles est souvent plus faible (45 à 50 Gray).

Contraintes de dose pour les organes à risque Les contraintes de dose pour les organes à risque dépendent là encore de la configuration de chaque tumeur, mais les valeurs suivantes sont courantes comme doses maximales acceptables [RTOG 2004, RTOG 2005, RTOG 2010] : 45 Gy pour la moelle épinière, 54 Gy pour le tronc cérébral, les nerfs optiques et le chiasma optique, 70 Gy pour la mandibule, 35 Gy pour les yeux. Certaines de ces doses limites doivent être impérativement respectées (organes en série comme la moelle ou les nerfs optiques), d'autres sont à respecter sur une certaine proportion du volume (organes en parallèle). Par exemple pour les parotides, comme il est souvent plus facile d'épargner une parotide que l'autre, une recommandation courante est de satisfaire à l'une des trois limites suivantes : (1) pour au moins une des deux parotides, la dose moyenne doit être inférieure à 26 Gy, (2) au moins 50% d'au moins une des deux parotides doit recevoir moins de 30 Gy, (3) au moins 20 ml du volume combiné des deux parotides doit recevoir moins de 20 Gy.

2.2.4.4 Prescription de fractionnement et d'étalement du traitement

Les prescriptions de dose sur les volumes cibles détaillées ci-dessus concernent les doses totales reçues lors du traitement. Cependant, l'irradiation ne se fait pas en une seule séance mais elle est fractionnée en plusieurs séances selon la prescription du radiothérapeute (dose par séance, intervalle de temps entre les séances et durée totale du traitement). En pratique, le traitement est souvent fractionné en 2 Gray par jour à raison de 5 jours par semaine, soit 10 Gray par semaine, et cela pendant 5 à 7 semaines.

La grande problématique liée à la radiothérapie consiste à préserver les tissus sains tout en détruisant les tissus cancéreux. Il existe différents moyens pour satisfaire au mieux cet objectif. Le premier moyen consiste à limiter au maximum la zone irradiée à la tumeur. Cela est rendu possible avec les techniques de conformation présentées dans la section 2.2.3. Cependant, ces techniques ont leur limites (notamment liées à l'imprécision des contourages et au mouvement de la cible), et il reste difficile

d'épargner complètement les tissus sains environnant la tumeur. Le second moyen d'action, qui est complémentaire du premier, est le fractionnement du traitement en plusieurs séances et son étalement sur plusieurs semaines. Cela permet d'amplifier au niveau tissulaire l'effet différentiel de l'irradiation créé au niveau cellulaire entre les cellules saines et tumorales qui est décrit dans le paragraphe 2.2.2.

Intérêt du fractionnement du traitement La motivation principale derrière l'idée du fractionnement est de laisser le temps aux tissus sains de réparer les lésions de leur ADN entre deux séances, et de profiter du fait que le processus de réparation est plus long et défaillant pour les cellules tumorales. En pratique, cela est un peu plus compliqué : la dose par séance, l'intervalle entre deux séances et la durée totale du traitement sont autant de paramètres qui ont un impact sur l'efficacité du traitement ainsi que sur la survie des tissus sains à renouvellement lent et à renouvellement rapide (et donc sur les effets secondaires).

Les tissus sains à renouvellement lent sont peu radiosensibles, donc ils ont peu de lésions ADN à réparer entre deux séances. L'intervalle entre deux séances doit être suffisant pour leur permettre de réparer la majorité des lésions (au moins 6-8 heures), mais s'il est trop long (supérieur à 24 heures), l'effet de l'irradiation sur les cellules tumorales est perdu car elles vont réparer une partie des lésions et surtout proliférer par repopulation cellulaire. La dose par séance a elle aussi une grande influence sur la survie des tissus sains à renouvellement lent : si elle est trop importante, le mécanisme de réparation chez ces tissus risque d'être saturé. Dans les deux cas (intervalle trop court entre les séances ou dose par séance trop importante), cela va créer des effets secondaires tardifs sur les tissus à renouvellement lent, car la régénération cellulaire à partir des cellules survivantes est faible après l'irradiation pour ces tissus. Il s'agit typiquement du système nerveux central ou des os.

Les tissus à renouvellement rapide sont radiosensibles, et ont donc beaucoup de lésions ADN à réparer entre deux séances. En pratique, le système de réparation sature, et une proportion non négligeable des lésions n'est pas réparée ou est mal réparée, induisant la mort cellulaire. Contrairement aux tissus sains à renouvellement lent, la survie à long terme des tissus sains à renouvellement rapide n'est donc que peu due à la réparation des lésions entre deux séances, mais davantage à la prolifération des cellules survivantes après le traitement qui va combler les pertes cellulaires causées par l'irradiation. Par conséquent, l'intervalle de temps entre deux séances et la dose par séance ont une influence limitée sur la survie à long terme de ces tissus. En revanche, des effets secondaires aigus vont être observés pour ces tissus (typiquement une inflammation provisoire des muqueuses des parois digestives pour l'ORL).

D'autre part, le fractionnement permet aussi de ré-oxygéner les cellules tumorales hypoxiques et donc de les rendre radiosensibles selon le processus suivant : parmi les cellules tumorales, celles qui sont radiosensibles et qui sont amenées à mourir sont les cellules bien oxygénées, donc généralement les cellules superficielles de la tumeur. La mort progressive de ces cellules va permettre à la vascularisation de

s'étendre progressivement aux cellules tumorales internes (hypoxiques), ce qui va les ré-oxygéner et les rendre radiosensibles, etc.

La durée totale du traitement (étalement) ne doit pas être trop longue car sinon la repopulation cellulaire de la tumeur diminue l'efficacité du traitement et la dose totale doit alors être augmentée pour compenser, ce qui accentue la toxicité et les effets secondaires à court et long terme.

Enfin, le fractionnement utilisé dans la plupart des cas (2 Gy par jour, 5 jours par semaine pendant 5-7 semaines) peut être modifié dans certains cas : par exemple un hypofractionnement (dose par séance plus élevée, étalement moins long) peut être préconisé pour raccourcir le traitement en radiothérapie palliative. L'hyperfractionnement consiste quant à lui à répartir la dose quotidienne en deux séances au lieu d'une seule pour limiter la repopulation cellulaire tumorale.

2.2.4.5 Choix de la balistique et calcul de la dosimétrie

Une fois que les contraintes de dose pour les volumes cibles et les organes à risque ainsi que le fractionnement ont été définis par le radiothérapeute, c'est le radiophysicien dosimétriste qui prend le relais. Il est alors chargé de proposer le plan de traitement le plus adapté pour satisfaire la prescription médicale du radiothérapeute. Pour cela, il doit tout d'abord définir le nombre de faisceaux et leurs positions et orientations. Ensuite, un logiciel de planification de traitement radiothérapeutique (Treatment Planning System, abrégé TPS) permet de déterminer la modulation de forme de chaque faisceau (et les mouvements correspondants des lames du collimateur en IMRT) et la carte de dose associée en résolvant un problème inverse à partir des données suivantes :

- les positions des faisceaux (correspondant aux meilleurs angles d'attaque pour la tumeur) et les caractéristiques de l'accélérateur linéaire,
- le scanner dosimétrique qui représente l'absorption des rayons X par les tissus (unités Hounsfield),
- les contraintes de doses sur les volumes cibles et sur les organes à risque, ainsi que l'ordre de priorité de ces différentes contraintes.

Le radiophysicien est amené à comparer différents plans de traitements pour voir lequel satisfait le mieux les prescriptions de dose. En effet, dans certains cas, le TPS peut avoir des difficultés à fournir un plan de traitement optimal du premier coup, par exemple si les contraintes de dose sont complexes à cause d'un nombre élevé d'organes à risque. Pour comparer deux plans de traitements, le radiophysicien va regarder si les isodoses se conforment bien aux prescriptions et limites de doses. De façon plus quantitative, il va également regarder les distributions de dose dans chaque volume contouré en s'aidant des histogrammes dose-volume (HDV) tracés automatiquement par le TPS (cf. figure 2.13). Enfin, le radiothérapeute doit valider le plan de traitement proposé par le radiophysicien.

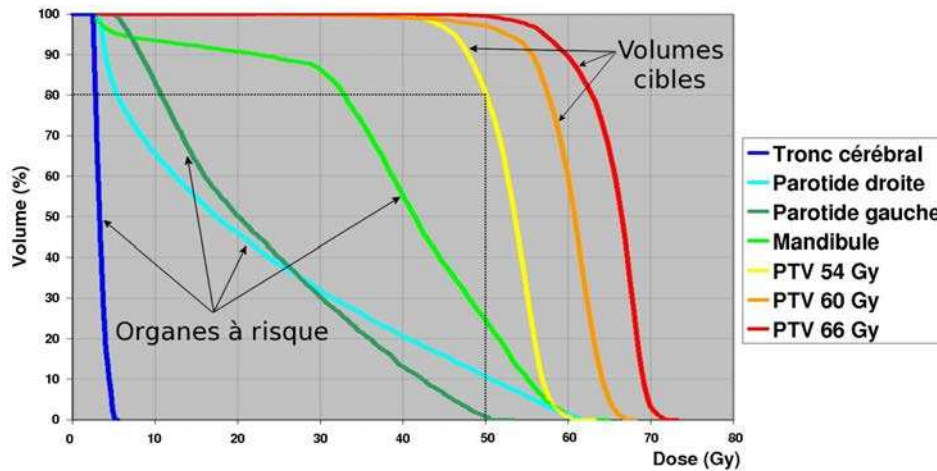


FIG. 2.13 – Exemple d’histogrammes dose-volume (HDV). En abscisse est représentée la dose D en Gray. Pour chaque volume contouré sur le scanner dosimétrique et pour chaque dose D , la courbe correspondante permet de visualiser quel pourcentage du volume reçoit une dose supérieure à D . Par exemple (lignes en pointillé), 80% du volume cible $PTV\ 54\ Gy$ reçoit une dose supérieure à 50 Gy.

2.2.5 Objectif de la thèse dans la planification du traitement

L’objectif de la thèse est relatif à l’étape de contournage des volumes cibles et des organes à risque par le radiothérapeute sur le scanner dosimétrique présentée dans les parties 2.2.4.1 et 2.2.4.2. En routine clinique, ces contourages sont faits manuellement par le radiothérapeute mais cela comporte plusieurs inconvénients. L’inconvénient majeur est que cette tâche est longue (de l’ordre de 1 à 3 heures par patient) et fastidieuse car elle nécessite le contournage coupe par coupe de chacun des volumes à considérer (volumes cibles et organes à risque). Pour gagner du temps, le radiothérapeute aura donc tendance à ne contourer qu’un nombre limité d’organes à risque (ceux qui lui paraissent primordiaux pour le traitement) et à ignorer les autres. Le second inconvénient est que les contourages manuels ne sont pas ou peu reproductibles du fait de la variabilité inter-expert (variabilité de contournage entre deux experts différents) et intra-expert (variabilité de contournage d’un même expert d’un jour à l’autre). Cette variabilité est décrite dans l’annexe C.

L’objectif de cette thèse est de fournir des outils de segmentation automatique des organes à risque ainsi que des aires ganglionnaires pour les cancers ORL. Le contournage de la tumeur (GTV) et de ses extensions microscopiques (CTV proche de la tumeur) restent manuels car ils ne dépendent pas seulement de l’image mais aussi de caractéristiques histologiques de la tumeur. Notre objectif est de fournir des contours automatiques les plus exacts possibles (de façon à minimiser le temps de correction par le médecin) tout en assurant un temps de calcul raisonnable.

Les outils de segmentation automatique comportent plusieurs avantages :

- Ils permettent un gain de temps important pour le radiothérapeute par rapport à un contourage manuel. Celui-ci doit tout de même vérifier la validité des contours automatiques et les corriger si besoin, mais plusieurs études ont montré que le temps de correction est moindre par rapport à un contourage manuel d'emblée [Chao 2007, Hu 2008, Stapleford 2010, Teguh 2010]. Notre objectif est de minimiser ce temps de correction. Le gain de temps peut potentiellement permettre de traiter davantage de patients puisque la procédure de planification est alors accélérée.
- Ils permettent de prendre en compte dans la planification du traitement certains organes à risque qui ne seraient pas contourés par manque de temps si le contourage était manuel.
- Ils fournissent des segmentations reproductibles. Plusieurs auteurs ont montré que la variabilité de contourage inter-expert entre deux experts corrigeant manuellement les mêmes contours automatiques est moindre en comparaison du cas où les deux experts contourent manuellement d'emblée [Chao 2007, Stapleford 2010].

Cette thèse s'est déroulée dans le cadre d'une convention CIFRE entre l'INRIA et DOSIsoft. La suite logicielle commercialisée par DOSIsoft (ISOgray) permet de planifier les traitements radiothérapeutiques selon les étapes décrites dans les sections 2.2.4.1 à 2.2.4.5. Mon travail s'est incorporé dans le module particulier de ISOgray dédié aux outils de contourage automatique. Une capture d'écran du module de segmentation utilisant l'atlas moyen est présentée sur la figure 2.14.

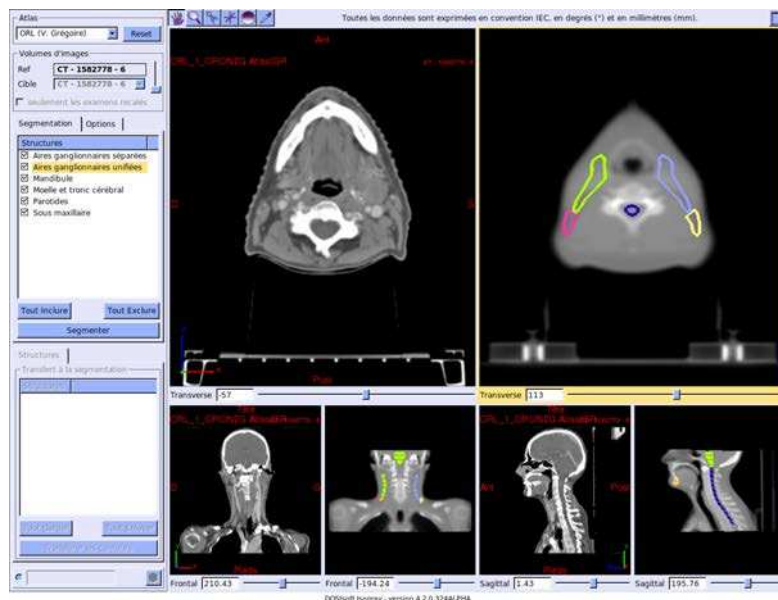


FIG. 2.14 – Capture d'écran du logiciel ISOgray commercialisé par DOSIsoft.

Segmentation par atlas

Sommaire

3.1	Segmentation d'images médicales	31
3.1.1	Principe et applications	31
3.1.2	Segmentation manuelle <i>versus</i> segmentation automatique	32
3.2	Bref état de l'art des méthodes de segmentation	33
3.2.1	Méthodes de bas niveau utilisant peu ou pas d' <i>a priori</i>	33
3.2.2	Méthodes utilisant des <i>a priori</i> de haut niveau	35
3.2.3	Notre application	38
3.3	Recalage	41
3.3.1	Principe	41
3.3.2	Utilisation du recalage en radiothérapie	41
3.3.3	Différentes approches	42
3.3.4	Classes de transformation	47
3.3.5	Mise en œuvre	48
3.4	État de l'art des stratégies de segmentation par atlas	49
3.4.1	Atlas unique utilisé pour tous les patients	50
3.4.2	Construction d'atlas adaptés à l'anatomie du patient	53
3.4.3	Conclusion	59

Ce chapitre décrit les principes de la segmentation et du recalage d'images, ainsi que les principales approches de la littérature. Il présente également un résumé de l'état de l'art des approches de segmentation par atlas.

Publications associées Un état de l'art sur l'utilisation des différentes stratégies de segmentation par atlas pour la radiothérapie a été publié dans le journal *Cancer/Radiothérapie* [Ramus 2010b].

3.1 Segmentation d'images médicales

3.1.1 Principe et applications

La segmentation d'une image médicale P est la tâche consistant à contourer les structures anatomiques d'intérêt sur cette image. Ces structures d'intérêt peuvent être des organes (par exemple poumons, cœur, foie, ...) ou des structures liées à une pathologie (par exemple lésions de sclérose en plaques ou tumeur). Le contourage

de ces structures peut être vu comme l'étiquetage des voxels de l'image P , chacun d'entre eux se voyant assigné un label $s \in [0 \dots L]$. Le label peut représenter une structure d'intérêt ($s \in [1 \dots L]$) ou le fond de l'image ($s = 0$). La segmentation d'une image P est notée $\{S_P(x)\}_{x \in P}$.

La segmentation d'images médicales a principalement deux applications cliniques : (1) l'aide au diagnostic (par exemple, le contourage d'une tumeur permet de mesurer sa taille et son volume, informations qui sont indispensables pour le diagnostic) et (2) la planification d'un traitement (radiothérapeutique ou chirurgical).

3.1.2 Segmentation manuelle *versus* segmentation automatique

Avantages et inconvénients La segmentation manuelle est la procédure utilisée traditionnellement pour les images médicales, mais elle a plusieurs inconvénients :

- d'une part, c'est une tâche longue et fastidieuse car elle nécessite le contourage coupe par coupe de chaque structure d'intérêt ;
- d'autre part, elle introduit un biais lié à la variabilité de contourage intramais surtout inter-expert, et fournit donc des contourages non reproductibles. Cette variabilité est décrite dans l'annexe C.

Les méthodes de segmentation automatique permettent de s'affranchir de ces deux inconvénients : elles permettent un gain de temps pour le médecin, et fournissent des segmentations reproductibles.

Intérêt des outils de segmentation automatique en radiothérapie malgré la nécessité de corrections manuelles Pour le contourage des volumes cibles et organes à risque de la sphère ORL, les contours automatiques sont cependant rarement parfaits et le médecin doit alors passer un peu de temps pour vérifier leur validité, et les corriger manuellement si besoin. Cette étape amène à nuancer d'une certaine manière les intérêts de la segmentation automatique puisqu'elle rallonge le temps-médecin nécessaire d'une part, et qu'elle ré-introduit une certaine variabilité de contourage via les corrections manuelles d'autre part. Cependant, de nombreuses études ont mis en évidence l'intérêt du contourage automatique malgré ces nuances.

Tout d'abord, le temps nécessaire pour la validation et la correction manuelle des contours automatiques a été comparé au temps nécessaire pour le contourage manuel d'emblée. L'ensemble des études mettent en évidence un gain de temps en faveur du contourage automatique corrigé, qui dépend essentiellement de deux facteurs : la localisation de la tumeur primaire [Chao 2007, Hu 2008] et le degré d'expérience en contourage du médecin [Chao 2007]. Le gain de temps moyen a été évalué à 35 % dans [Stapleford 2010], et 63 % dans [Teguh 2010]. Chao et al. ont montré que le gain de temps était moindre pour les médecins ayant une bonne expérience du contourage (gain de temps moyen sur 4 patients de 26% à 29% selon le type de cancer) par rapport aux médecins moins expérimentés dans le contourage (gain de temps moyen sur 4 patients de 38% à 47% selon le type de cancer) [Chao 2007]. Hu et al. ont reporté une différence non significative entre le temps nécessaire pour

corriger les contours automatiques, et le temps nécessaire pour corriger les contours manuels faits par un interne [Hu 2008].

Enfin, plusieurs auteurs ont montré que la variabilité de contourage inter-expert entre deux experts corrigeant manuellement les mêmes contours automatiques est moindre en comparaison du cas où les deux experts contourent manuellement d'emblée [Chao 2007, Stapleford 2010].

3.2 Bref état de l'art des méthodes de segmentation

Les algorithmes de segmentation automatique d'images peuvent se répartir selon les deux catégories suivantes :

- méthodes de bas niveau utilisant peu ou pas d'*a priori* sur l'objet à segmenter ;
- méthodes utilisant des *a priori* de haut niveau sur l'objet à segmenter sous forme d'un modèle.

3.2.1 Méthodes de bas niveau utilisant peu ou pas d'*a priori*

Cette première catégorie regroupe les méthodes qui n'utilisent que des informations extraites de l'image pour la segmenter. Elle englobe typiquement les algorithmes de traitement d'images traditionnels (parmi lesquels les approches basées régions et celles basées contours), les modèles déformables ainsi que les algorithmes de partitionnement ou clustering (classification non supervisée). Les informations exploitées sont ici réduites aux informations image (intensités des voxels et leurs dérivées (gradient par exemple)), ainsi qu'au nombre de structures à segmenter pour les algorithmes de classification non supervisée.

Approches basées régions Les approches basées régions définissent l'objet à segmenter comme l'ensemble des voxels satisfaisant une certaine condition. La plus intuitive de ces méthodes est la segmentation par seuillage : un voxel appartient à l'objet à segmenter si son intensité est supérieure ou inférieure à une valeur fixée, ou comprise entre deux valeurs fixées. Cette méthode est peu robuste au bruit dans l'image et nécessite des post-traitements dans la plupart des cas. Elle peut être utilisée pour segmenter des structures homogènes et très contrastées comme les os sur des images scanner, mais est peu utilisée en tant que telle pour des structures plus difficilement visibles sur les images. Des variantes de cette méthode, comme le seuillage par hystérésis ou encore le seuillage adaptatif (par exemple celui utilisé dans [Teng 2006]), permettent d'améliorer la robustesse. Un autre exemple de méthode basée région est la croissance de région, utilisée dans [Mazonakis 2001] pour segmenter les organes à risque du pelvis (vessie, rectum, prostate). Cette technique consiste à partir d'une région constituée d'une graine (souvent définie manuellement par l'utilisateur) et à faire croître cette région en y agrégeant itérativement les voxels voisins qui satisfont une condition d'homogénéité.

Approches basées contours Contrairement aux approches basées région, les approches basées contours consistent à détecter les contours de l'objet plutôt que l'objet lui-même. Ces approches se basent typiquement sur les gradients d'intensité de l'image. Le cas d'école consiste à extraire les maxima du gradient, qui correspondent aux contours de l'objet pour un objet bien contrasté. Ici encore, la méthode est très sensible au bruit dans les images et peut par exemple conduire à des contours non fermés qui doivent être corrigés par des post-traitements adéquats (typiquement des opérations de morphologie mathématique). Un exemple plus complexe de détection de contours est la ligne de partage des eaux [Vincent 1991], qui considère l'image en niveaux de gris comme une carte topographique et cherche à déterminer ses lignes de crête par des opérations de morphologie mathématique. En général, la ligne de partage des eaux est appliquée sur le gradient de l'image. Un inconvénient courant lié à cette méthode est la sur-segmentation (plus de régions obtenues que souhaité).

Modèles déformables Introduite par [Kass 1987, Terzopoulos 1988], la segmentation par modèles déformables consiste à faire évoluer itérativement un contour initial jusqu'à convergence de façon à maximiser ou minimiser un certain critère à définir. La forme de l'objet que l'on fait évoluer peut être représentée de façon explicite (typiquement par un maillage) ou implicite (typiquement par l'ensemble des zéros d'une fonction à valeurs dans \mathbb{R} , c'est la technique dite des « level set »). Dans les deux cas, le critère à optimiser fait intervenir un terme d'attache aux données (énergie ou force externe, visant à faire coller le contour à l'image) et un terme de régularisation (énergie ou force interne, visant à assurer un contour lisse). La bonne convergence des modèles déformables requiert en général une bonne initialisation. Pour cette raison, ils nécessitent souvent l'introduction d'un *a priori* sur la forme de l'objet à segmenter. Le contour initial peut être une forme proche de celle attendue (par exemple une sphère pour la prostate), ou bien le résultat de segmentation obtenu avec une autre méthode (par exemple par une segmentation par atlas comme dans [Han 2010, Hollensen 2010, Martin 2010]). Dans ce dernier cas, le modèle déformable est utilisé pour affiner le résultat. Un état de l'art détaillé sur l'utilisation des modèles déformables pour la segmentation d'images médicales est présenté dans [Montagnat 1999, Montagnat 2001].

Les principaux inconvénients des modèles déformables sont qu'ils sont sensibles au bruit dans les images et aux artefacts, et qu'ils nécessitent une bonne initialisation. De plus, bien que déjà utilisés pour la segmentation simultanée de plusieurs structures voisines [Ciofalo 2005, Chupin 2009], ils ne sont pas très adaptés pour cette application. En effet, d'une part chaque structure requiert une optimisation indépendante (définition des forces utilisées pour guider la segmentation pour cette structure), et d'autre part une contrainte spécifique doit alors être ajoutée pour gérer chaque frontière entre deux structures voisines.

Algorithmes de partitionnement ou classification non supervisée Les algorithmes de partitionnement (*clustering* en anglais) répartissent les voxels de

l'image en un nombre fixé de *classes* (aussi appelées *clusters*) en se basant sur leurs intensités, et souvent de façon itérative. Ce partitionnement peut être exclusif (c'est-à-dire chaque voxel est associé à une seule classe exclusivement) ou flou (c'est-à-dire chaque voxel a une probabilité d'appartenance à chacune des classes). Les deux algorithmes de classification floue les plus connus sont l'algorithme des fuzzy c-means [Bezdek 1981] et l'algorithme EM (Expectation-Maximisation) [Dempster 1977, Leemput 1999]. Ces deux algorithmes alternent itérativement entre l'estimation des paramètres du modèle (intensité moyenne de chaque classe pour les fuzzy c-means, à laquelle on ajoute l'écart-type de chaque classe pour l'EM) et l'estimation des probabilités d'appartenance à chacune des classes de chaque voxel étant donnés les paramètres courants du modèle. La différence principale entre les deux est que la distribution d'intensités de chaque classe est supposée gaussienne dans l'algorithme EM, ce qui n'est pas le cas pour les fuzzy c-means. La segmentation de l'image à proprement parler se déduit ensuite en assignant chaque voxel à la classe ayant la probabilité maximale. En pratique, l'algorithme EM utilise dans la plupart des cas un *a priori* spatial obtenu avec une autre méthode, à la fois comme initialisation et comme contrainte à chaque itération.

3.2.2 Méthodes utilisant des *a priori* de haut niveau

Dans les techniques précédentes, l'introduction d'*a priori* sur l'objet à segmenter se limite à la détermination empirique des valeurs optimales à donner à certains paramètres de l'algorithme (par exemple, la valeur du seuil pour le seuillage). On connaît pourtant dans la plupart des cas une certaine partie des caractéristiques sur l'objet à segmenter (parmi lesquelles sa forme, son intensité, sa localisation ou encore son environnement dans l'image) puisqu'on est en général capable de le repérer visuellement sur l'image. La seconde catégorie de méthodes consiste donc à tirer profit de ces caractéristiques connues pour modéliser l'objet à segmenter, et à introduire dans l'algorithme ces connaissances sous forme d'*a priori* pour guider la segmentation.

L'objectif n'étant pas de faire un état de l'art exhaustif des méthodes de segmentation, nous nous contentons ici de décrire les principales approches suivant le type d'*a priori* qu'elles utilisent.

3.2.2.1 Méthodes de classification supervisée

Les algorithmes de classification supervisée supposent qu'on connaît le nombre de classes dans lesquelles répartir les voxels de l'image, et ont pour objectif de regrouper entre eux les voxels qui se ressemblent selon un critère à définir en s'aidant d'une base d'apprentissage d'images manuellement segmentées. Chaque voxel est décrit par un vecteur de caractéristiques (intensité, gradient, texture par exemple) puis projeté dans l'espace des caractéristiques dans lequel ont été préalablement projetées les données d'apprentissage. L'algorithme des K plus proches voisins consiste alors à classer le voxel dans la classe majoritaire parmi les K données d'apprentissage qui

lui sont le plus proche dans l'espace des caractéristiques. Cela nécessite de définir une distance dans l'espace des caractéristiques. Les *a priori* introduits ici sont donc (1) le nombre d'objets à segmenter, et (2) des *a priori* basés intensités (ou autres informations image) au travers des vecteurs de caractéristiques décrivant les voxels. D'autres exemples de classification supervisée sont par exemple les random forests ou encore les SVM (Support Vector Machine).

3.2.2.2 Méthodes utilisant des *a priori* sur la forme de l'objet

Introduction de contraintes de forme dans les modèles déformables Dans le cadre des modèles déformables, un *a priori* sur la forme de la structure à segmenter peut tout d'abord être introduit à travers l'initialisation du contour (par exemple contour initial sphérique pour segmenter la prostate). De plus, des contraintes spécifiques à la forme de l'objet à segmenter peuvent facilement être prises en compte tout au long de l'évolution du modèle déformable en rajoutant des termes de rappel dans le critère à optimiser (par exemple, contraintes tubulaires ou contraintes de non-chevauchement entre deux structures).

Utilisation des modèles statistiques de forme Si l'on dispose d'un ensemble de segmentations de la structure à segmenter provenant de différents patients, il est possible d'exploiter ces données pour pousser à l'extrême l'introduction de contraintes de formes dans l'évolution d'un modèle déformable. En effet, la base de segmentations peut être utilisée pour construire un modèle statistique de forme (Statistical Shape Models, SSM) qui contient la forme moyenne ainsi que les principaux modes de variations de la forme en question dans la base d'apprentissage. La construction d'un tel modèle se fait selon les étapes suivantes : (1) définition de points de correspondance entre les formes, (2) alignement des formes entre elles, et (3) PCA (Principal Component Analysis). Une fois construit, ce modèle statistique peut être utilisé dans un schéma de type modèles déformables pour contraindre la forme à évoluer selon les modes de variations observés dans la base d'apprentissage. Ces modèles déformables contraints sont appelés Active Shape Models (ASM) [Cootes 1995]. Une extension de ces modèles consiste à inclure dans le modèle statistique non seulement la variabilité de forme mais aussi la variabilité des informations image associées (Active Appearance Models, AAM [Cootes 2001]).

Le principal obstacle à l'utilisation des modèles statistiques de forme (ASM et AAM) est que leur construction nécessite de définir des correspondances de points entre les différentes formes de la base, ce qui peut être compliqué voire impossible pour des structures anatomiques complexes. Frangi et al. ont proposé une méthode [Frangi 2001] pour estimer automatiquement ces correspondances en faisant une PCA sur les champs de déformation non-linéaires, mais cette technique est tributaire du recalage et peut donc aboutir à des correspondances fausses si celui-ci échoue (ce qui peut arriver si les anatomies sont trop différentes).

3.2.2.3 Méthodes utilisant des *a priori* sur l'environnement de l'objet

Dans certains cas, les informations présentes dans le voisinage proche de l'objet à segmenter peuvent être avantageusement utilisées pour guider la segmentation de l'objet. Dans le cadre de la segmentation d'images médicales, l'idée consiste à introduire des connaissances *a priori* sur la position relative de la structure à segmenter par rapport à d'autres structures ou repères anatomiques de référence qui sont plus facilement segmentables [Kobashi 1995]. Bloch et al. se sont appuyés sur ce principe et ont proposé une méthode de segmentation successive des structures du cerveau (de la plus facile à la plus difficile à segmenter) [Bloch 2005]. Ce principe a aussi été utilisé pour segmenter successivement les repères anatomiques définissant les aires ganglionnaires de la région ORL pour la radiothérapie [Teng 2006].

3.2.2.4 Méthodes utilisant des *a priori* sur la localisation de l'objet

La segmentation dite *par atlas* consiste à regrouper dans un *atlas* des *a priori* de forme et de localisation sur les objets à segmenter, et à utiliser cet *atlas* pour guider la segmentation. Concrètement, un atlas peut être défini comme une anatomie préalablement étiquetée, et se présente généralement sous la forme d'une image en niveaux de gris et de sa segmentation associée. La segmentation peut être représentée par une image de labels ou encore par un ensemble de cartes de probabilités (une carte de probabilité par structure). Étant donné un atlas, la segmentation par atlas consiste alors simplement à mettre en correspondance l'image à segmenter avec l'image d'intensités de l'atlas (étape de recalage), puis à déformer la segmentation de l'atlas sur l'image (cf. figure 3.1). Les deux ingrédients principaux de la segmentation par atlas sont donc :

- l'atlas en lui-même : la section 3.4 décrit les différentes possibilités pour définir un atlas ;
- la méthode de recalage utilisée pour recaler l'atlas sur le patient : la section 3.3 présente les problématiques du recalage ainsi qu'un bref état de l'art des différentes classes de méthodes.

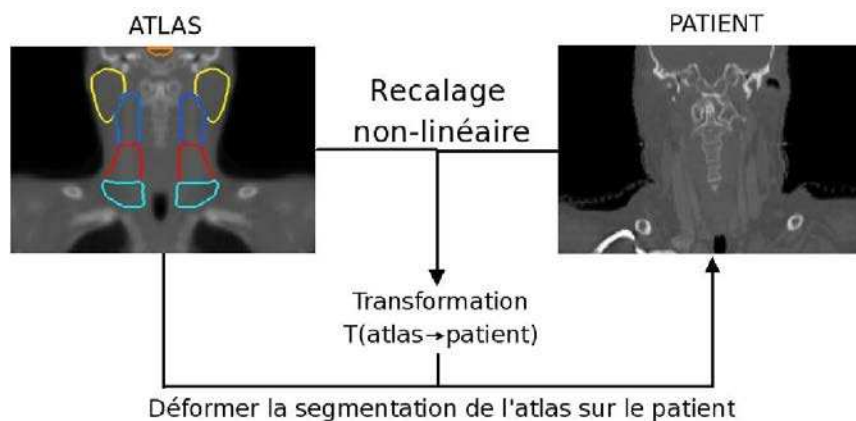


FIG. 3.1 – Principe de la segmentation par atlas.

Le principal avantage de la segmentation par atlas est qu'elle permet de segmenter simultanément un grand nombre de structures, ce qui n'est pas le cas des modèles déformables par exemple. De plus, étant basée sur le recalage d'un atlas sur l'image, elle exploite les informations de l'image entière et pas seulement des informations locales, ce qui en fait une méthode robuste au bruit dans les images. Son principal inconvénient est qu'elle peut avoir des difficultés à segmenter de façon précise des formes complexes.

3.2.3 Notre application

Parmi les structures d'intérêt de la sphère ORL (aires ganglionnaires et structures à risque détaillées dans la section 2.1.5), seule la mandibule, qui est une structure osseuse, présente un contraste suffisant avec les tissus environnants pour être segmentée efficacement avec une méthode bas niveau basée intensité : 60% de ses voxels se distinguent bien des tissus mous avec une intensité supérieure à 300 HU, et les autres voxels sont des voxels internes à la structure qui peuvent être récupérés par une opération de morphologie mathématique.

Pour les autres structures en revanche, deux problèmes se posent : (1) elles ne sont pas beaucoup contrastées par rapport aux graisses et tissus mous qui les entourent, et (2) certaines (les aires ganglionnaires) sont fondamentalement inhomogènes en terme d'intensité. En effet, la figure 3.2 montre que les distributions d'intensité des aires ganglionnaires II, III et IV couvrent toute la plage des tissus mous $[-300HU; 300HU]$ (ce qui rend difficile l'estimation des frontières de ces aires avec les tissus mous), et se chevauchent les unes par rapport aux autres (ce qui rend difficile l'estimation des frontières II/III et III/IV). Les autres structures (parotides, sous-maxillaires, tronc cérébral et moelle épinière) sont davantage homogènes, comme le montrent leurs distributions d'intensités plus étroites. Cependant, dû à la proximité de la mâchoire, les parotides et sous-maxillaires sont sujettes au biais d'intensité introduit par les artefacts d'acquisition liés aux plombages dentaires. Les méthodes de bas niveau ne sont pas suffisamment robustes à ces artefacts, et la segmentation de l'ensemble des structures de la sphère ORL requière donc l'introduction d'*a priori* de haut niveau pour guider la segmentation. La figure 3.3 montre l'image manuellement contourée correspondant aux distributions d'intensité de la figure 3.2 et permet de visualiser les difficultés mentionnées précédemment.

Les modèles déformables ont été utilisés pour segmenter tout ou partie des organes à risque du pelvis (prostate, vessie, rectum) [Costa 2007, Zimmer 2005] mais ils sont peu adaptés à la segmentation simultanée d'un grand nombre de structures (une quinzaine dans notre cas). De plus, ces modèles risquent de manquer de robustesse par rapport aux artefacts dentaires, au faible contraste et à l'inhomogénéité de certaines de nos structures d'intérêt. L'utilisation de modèles statistiques de forme pour contraindre les modèles déformables n'est pas envisageable dans notre cas car la grande variabilité anatomique ainsi que la variabilité de contourage rendent difficile la définition de points de correspondance valables entre les différentes segmentations de notre base. Éventuellement, les modèles déformables pourraient être utilisés dans

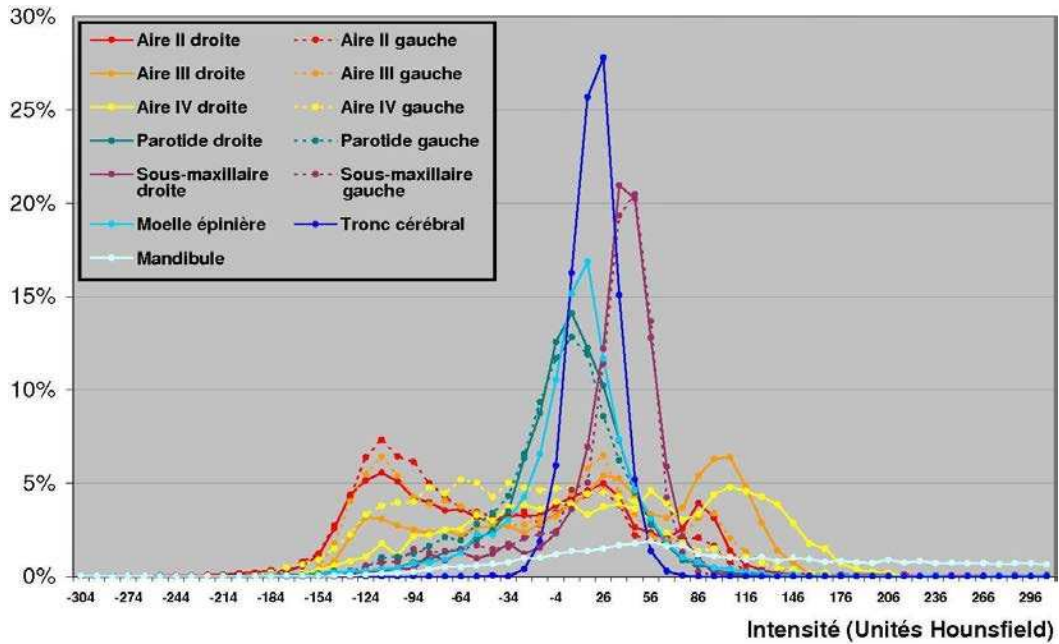


FIG. 3.2 – Histogrammes d'intensité structure par structure tracés à partir d'une image contourée manuellement (en pourcentage du nombre de voxels de la structure et avec une largeur de bande de 10 HU). Toutes les structures occupent la même plage d'intensité $[-300HU; 300HU]$, sauf la mandibule pour laquelle plus de 60% des voxels ont une intensité supérieure à 300 HU.

notre cas pour affiner une segmentation obtenue avec une autre méthode (comme c'est le cas dans [Han 2010, Hollensen 2010] pour la sphère ORL également), mais semblent en conclusion peu appropriés pour être utilisés d'emblée.

La segmentation par atlas permet de dépasser les différentes difficultés mentionnées. En effet, elle a souvent été utilisée pour segmenter simultanément un grand nombre de structures dans le cadre de la planification en radiothérapie, par exemple les organes à risques du cerveau [Bondiau 2005, Dawant 1999b] et les organes à risque et aires ganglionnaires de la sphère ORL [Han 2008, Commowick 2007b]. De plus, en tirant profit de l'information de toute l'image dans sa globalité et pas seulement au voisinage proche des structures, elle est relativement robuste aux artefacts d'acquisition, au faible contraste et à l'inhomogénéité de certaines de nos structures. Nous avons donc choisi cette approche pour la segmentation des structures d'intérêt de la sphère ORL. Les sections 3.3 et 3.4 abordent les éléments clés de la segmentation par atlas, à savoir, le recalage et la construction de l'atlas respectivement.

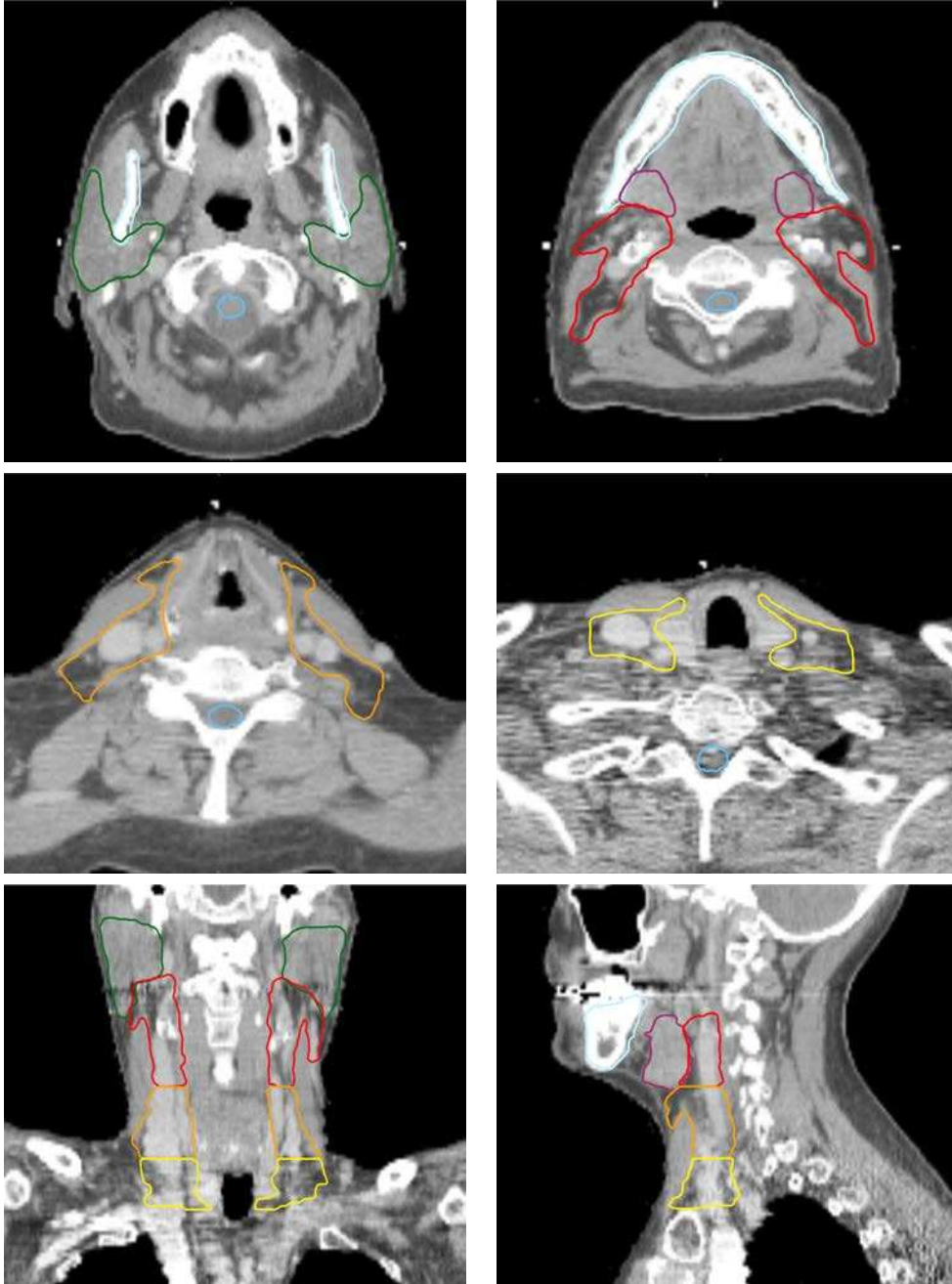


FIG. 3.3 – Contourages manuels correspondant aux distributions d'intensité représentées sur la figure 3.2 (même code de couleur).

3.3 Recalage

3.3.1 Principe

Le recalage d'une image flottante F sur une image de référence R consiste à rechercher la transformation optimale $\hat{T}_{F \leftarrow R}$ qui permet de ré-échantillonner l'image flottante en la rendant le plus similaire possible, selon un certain critère de similarité S , à l'image de référence R . La formalisation de ce problème selon [Brown 1992] est la suivante :

$$\hat{T}_{F \leftarrow R} = \arg \max_{T \in \mathfrak{T}} S(R, F \circ T) \quad (3.1)$$

Cette équation montre les points clés qui caractérisent un algorithme de recalage : la classe de transformations \mathfrak{T} dans laquelle on va chercher la transformation $\hat{T}_{F \leftarrow R}$, le critère de similarité S , et l'algorithme d'optimisation. Nous détaillons brièvement certains de ces éléments dans les paragraphes 3.3.3 et 3.3.4, mais avant cela nous allons mettre en évidence l'intérêt du recalage pour la radiothérapie.

3.3.2 Utilisation du recalage en radiothérapie

3.3.2.1 Applications

Le recalage de deux images est indispensable pour pouvoir comparer les informations qu'elles contiennent, les fusionner ou bien transférer l'information de l'une sur l'autre. Plusieurs articles ont proposé des résumés sur l'état de l'art de l'utilisation du recalage en imagerie médicale [Betrouni 2009] et en radiothérapie [Bonniaud 2006]. Les applications principales du recalage pour la radiothérapie sont les suivantes :

- **Recaler deux images de même modalité d'un même patient :** la radiothérapie adaptative [Castadot 2010, Castadot 2008, Zhang 2007, Peroni 2009a] consiste à adapter la planification faite sur l'image scanner initiale en fonction des changements morphologiques du patient au cours de traitement (diminution des volumes). Pour cela, les contours des structures sur l'image scanner en cours de traitement sont nécessaires, et ils peuvent être obtenus par recalage avec l'image scanner initiale et déformation de ses contours.
- **Recaler deux images de modalités différentes d'un même patient :** le recalage d'images d'autres modalités (IRM, PET) sur l'image scanner permet d'améliorer la qualité du contourage manuel des volumes cibles et organes à risque sur l'image CT [Boisserie 2001, Emami 2003]. En radiothérapie, le recalage intra-patient inter-modalité est aussi utilisé pour positionner le patient avant l'irradiation (recalage entre l'image scanner et le Cone Beam CT).
- **Recaler deux images de même modalité de deux patients différents :** ce type de recalage est utilisé pour recaler l'image du patient et un atlas anatomique de même modalité, ainsi que pour construire un atlas anatomique à partir de plusieurs images.
- **Recaler deux images de patients différents et de modalités différentes :** ce type de recalage est utilisé pour recaler un atlas d'une cer-

taine modalité sur l'image d'un patient d'une autre modalité. Par exemple [Peroni 2009b] recalent un atlas IRM sur des images de scanner CT en radiothérapie stéréotaxique, [Houhou 2005] recalent de façon hiérarchique des masques binaires sur une image CT, et [D'Agostino 2004] recalent une image de labels d'un atlas sur une image IRM.

3.3.2.2 Prise en compte de la tumeur dans le recalage

De nombreuses études se sont intéressées à la prise en compte de la tumeur dans le recalage, en particulier dans le cadre du recalage d'un atlas représentant une anatomie saine (sans tumeur) sur une image pathologique comportant une tumeur. La variabilité liée à la tumeur vient s'ajouter à la variabilité anatomique « normale », et aboutit à des erreurs de recalage qui peuvent être importantes quand la tumeur atteint une certaine taille.

Tout d'abord, certains auteurs ont proposé de capturer les grandes déformations introduites par la tumeur en utilisant une régularisation non uniforme dans le recalage non-linéaire [Stefanescu 2004].

Une autre approche présente dans la littérature consiste à modéliser les déformations causées par la tumeur en introduisant un modèle *a priori* de croissance de tumeur. Un tel modèle peut tout d'abord être utilisé pour générer une tumeur artificielle dans l'image de l'atlas sain à partir d'une graine spécifiée selon la localisation de la tumeur sur l'image du patient [Dawant 1999a, Bach Cuadra 2004, Pollo 2005]. Cette approche est appelée *seeded atlas deformation* en anglais. L'approche inverse a également été proposée : Kyriacou et Davatzikos ont présenté une méthode permettant de modéliser les déformations des tissus mous environnants la tumeur pour ensuite estimer une image de patient sans tumeur avec laquelle recaler l'atlas sain [Kyriacou 1999]. La déformation modélisant la croissance de tumeur est ensuite appliquée à l'atlas sain ainsi recalé.

A noter que dans nos travaux, nous n'avons pas pris en compte la tumeur dans le recalage non-linéaire. Le principal argument justifiant ce choix est la faible proportion d'images de notre base possédant une tumeur de grande taille.

3.3.3 Différentes approches

Il existe un grand nombre et une grande variété de méthodes de recalage. Un état de l'art détaillé est disponible dans [Zitová 2003]. Ce paragraphe présente les deux principales approches algorithmiques du recalage (géométrique ou iconique) ainsi que les approches hybrides combinant des aspects géométriques et iconiques.

3.3.3.1 Recalage géométrique

L'approche de recalage géométrique est basée sur les trois étapes suivantes :

- extraction de primitives pertinentes communes aux deux images R et F ,
- appariement des primitives,
- estimation de la transformation d'après les appariements.

Extraction des primitives Les primitives géométriques sont des points, courbes, surfaces ou volumes identifiables sur les images. Il peut s'agir de repères anatomiques particuliers (comme par exemple les lignes de crête [Subsol 1998]) ou de marqueurs extrinsèques fixés sur le patient. L'extraction peut être manuelle pour des primitives simples (points dans les images par exemple) ou automatique dans le cas contraire.

Appariement des primitives extraites Tout comme l'extraction, l'appariement peut être manuel ou automatique. Les critères d'appariement automatique peuvent être basés sur les intensités des images au voisinage des primitives, sur des caractéristiques géométriques propres aux primitives, ou plus simplement sur des distances entre primitives. L'étape d'appariement fournit un ensemble de couples $(P_i(R), P_i(F))$ où la primitive $P_i(R)$ de l'image de référence est appariée avec la primitive $P_i(F)$ de l'image flottante.

Estimation de la transformation d'après les appariements Le principe est ici un peu différent de celui décrit par l'équation 3.1 : les deux images R et F étant remplacées par deux ensembles de primitives $\{P_i(R)\}_i$ et $\{P_i(F)\}_i$, le critère optimisé n'est plus directement la similarité entre les deux images (critère à maximiser) comme suggéré dans l'équation 3.1, mais la distance entre les primitives appariées (critère à minimiser). Sous réserve que les primitives extraites soient pertinentes par rapport aux images et communes aux deux images, l'hypothèse sous-jacente aux approches géométriques est que la transformation estimée devrait alors aussi maximiser la similarité entre les images recalées.

Le critère de distance entre les primitives appariées va se formuler différemment suivant la nature des primitives extraites (car on ne définit pas de la même façon la distance entre deux points et la distance entre deux surfaces par exemple). Pour les primitives complexes (autres que des points), le critère peut être basé sur des calculs de cartes de distance par rapport aux primitives de l'image de référence comme c'est le cas pour le recalage de chanfrein [Borgefors 1988, van Herk 1994]. Pour les cas simples où les primitives sont des points, la méthode la plus courante pour réaliser l'optimisation est la méthode des moindres carrés formulée ainsi :

$$\hat{T}_{F \leftarrow R} = \arg \min_{T \in \mathfrak{T}} \sum_i \|T(P_i(R)) - P_i(F)\|^2. \quad (3.2)$$

Une variante de cette méthode est la méthode des moindres carrés tamisés. Elle consiste à rejeter de la somme les appariements aberrants pour lesquels la distance $\|T(P_i(R)) - P_i(F)\|$ est la plus grande. En pratique, un schéma itératif bouclant entre estimation des résidus et estimation de la transformation à partir des plus petits résidus permet d'être plus robuste aux erreurs d'appariement. Par ailleurs, il est également possible d'introduire une notion de confiance dans les appariements en pondérant les résidus dans la somme (moindres carrés tamisés pondérés).

Approches itératives couplant appariement et estimation de la transformation A l'exception des appariements manuels, des algorithmes d'appariement

automatique doivent être utilisés et introduisent potentiellement des erreurs dans les appariements. Afin d'être robuste par rapport à ces erreurs potentielles, une approche classique consiste à alterner de façon itérative entre estimation des appariements et estimation de la transformation. Un exemple classique de l'approche itérative est l'algorithme ICP (Iterative Closest Point) décrit par [Besl 1992]. Cet algorithme prend en entrée une initialisation T_0 de la transformation, et répète ensuite à chaque itération l jusqu'à convergence les étapes suivantes : (1) appariement de chaque primitive de R à la primitive la plus proche dans $F \circ T_{l-1}$ au sens de la distance euclidienne, (2) estimation de la correction de transformation δT_l expliquant au mieux ces nouveaux appariements, et (3) $T_l = T_{l-1} \circ \delta T_l$.

Avantages et inconvénients du recalage géométrique Toute la difficulté des approches géométriques consiste à extraire des primitives pertinentes et communes aux deux images. L'automatisation de cette étape d'extraction peut être difficile pour certaines images. Cela dit, une fois les primitives extraites, toute l'information présente dans les images est alors réduite en un nombre fini de primitives et cette représentation compacte de l'information assure une certaine robustesse par rapport au bruit dans les images. Un paramètre important pour la robustesse est le nombre de primitives extraites et appariées dans les images. D'une manière générale, la robustesse augmente avec le nombre de primitives dans les images.

L'approche géométrique est surtout utilisée pour estimer des transformations linéaires (et en particulier rigides). Estimer correctement une transformation non-linéaire avec une approche géométrique nécessite l'extraction d'un nombre élevé de primitives, ce qui est difficile à mettre en œuvre de façon automatique pour la plupart des images.

3.3.3.2 Recalage iconique

Pour les approches géométriques, l'étape critique est l'extraction des primitives mais le critère à optimiser est relativement simple puisqu'il est basé sur les distances euclidiennes entre les primitives appariées. Les approches iconiques quant à elles sont basées sur l'utilisation directe des intensités des voxels des deux images à recaler : elles ne nécessitent donc pas d'extraction préalable de primitives, mais la difficulté est reportée sur la définition du critère à optimiser. En effet, il est alors nécessaire de définir une mesure de similarité qui permettra de caractériser correctement la ressemblance entre les deux images en comparant leurs intensités. Cette mesure de similarité doit être adaptée à la relation supposée entre les intensités des deux images. Cette relation peut être plus ou moins complexe suivant la nature des images à recaler, allant de la conservation des intensités pour un recalage monomodal intrapatient, à une relation statistique pour un recalage multimodal. Plusieurs mesures ont été proposées dans la littérature pour estimer la similarité entre les deux images d'après la relation *a priori* attendue entre leurs intensités.

Relation d'égalité entre les intensités C'est typiquement le cas du recalage monomodal intra-patient. La mesure de similarité couramment utilisée est alors la somme des différences au carré entre les intensités (SSD), ou encore la somme des valeurs absolues des différences (SAD). L'équation 3.3 explicite le problème à résoudre en utilisant la SSD :

$$\hat{T}_{F \leftarrow R} = \arg \min_{T \in \mathcal{T}} \sum_{(x,y,z) \in R} \|R(x,y,z) - F \circ T(x,y,z)\|^2. \quad (3.3)$$

Relation linéaire ou affine entre les intensités Dans le cadre d'un recalage monomodal inter-patient, la SSD et la SAD peuvent éventuellement être utilisées, mais des pré-traitements sont alors nécessaires pour corriger les différences de contrastes entre les images. Pour être plus robuste à ces différences de contraste, on préférera maximiser le coefficient de corrélation (éventuellement au carré).

Relation fonctionnelle entre les intensités Pour un recalage multimodal, la relation supposée entre les intensités des deux images est plus complexe qu'une simple relation affine. Par exemple, certains tissus vont être en hypersignal dans une modalité, et en hyposignal dans l'autre, et/ou *vice versa*. La similarité entre les deux images doit donc être caractérisée par une relation fonctionnelle, ce qui est possible en utilisant le rapport de corrélation introduit par [Roche 1998].

Relation statistique entre les intensités De façon encore plus générale, la similarité entre deux images de modalités différentes peut être basée sur une relation statistique entre leurs intensités. Des mesures telles que l'entropie, l'information mutuelle [Viola 1997], ou encore l'information mutuelle normalisée [Studholme 1999] sont alors utilisées. Par exemple, Houhou et al. utilisent l'information mutuelle pour recalibrer une image CT 2D de la région ORL avec des masques binaires 2D de la même région [Houhou 2005].

Quelle que soit la mesure de similarité utilisée, elle peut *a priori* être calculée de façon globale entre les deux images (on parle alors de recalage iconique global) ou locale (recalage iconique local). Cependant, supposer une relation globale (c'est-à-dire valable sur toute l'image) entre les intensités des images est une hypothèse forte et rarement vérifiée. Le recalage iconique local au contraire restreint au niveau local la validité de la relation supposée entre les intensités. Le critère optimisé est alors une somme de critères locaux. Le voisinage sur lequel est calculée la mesure de similarité en chaque voxel doit être suffisamment grand pour que les variances ou histogrammes conjoints estimés soient statistiquement significatifs.

3.3.3.3 Recalage hybride

Les algorithmes dits *hybrides* combinent certains aspects des recalages géométriques et iconiques. Martin et al. alternent entre la minimisation d'une énergie de type iconique basée sur une mesure de similarité entre les images, et la minimisation

d'une énergie de type géométrique basée sur des distances entre des repères anatomiques fournis manuellement et des points d'un maillage de référence [Martin 2008]. Cachier et al. minimisent aussi une énergie comprenant un terme iconique et un terme géométrique (basé sur les lignes sulcales du cerveau), mais l'estimation de la transformation se fait ici de façon conjointe (et non pas alternée) à partir des deux types d'appariements [Cachier 2001]. Teng et al. cherchent à optimiser un critère géométrique au voisinage des maillages disponibles et un critère iconique sur le reste de l'image. Enfin, Sotiras et al. proposent de coupler les approches iconique et géométrique de façon simultanée en utilisant des champs de Markov [Sotiras 2010].

Nous présentons ici l'algorithme de recalage par blocs (ou block matching) [Ourselin 2000], qui peut être qualifié d'algorithme hybride géométrique/iconique pour les raisons suivantes :

- il emprunte aux approches géométriques un système d'appariement de primitives et une minimisation aux moindres carrés (tamisés) de la distance entre primitives appariées ;
- contrairement aux approches géométriques, les primitives ne sont pas extraites préalablement mais elles sont définies par des blocs régulièrement répartis sur l'image de référence. Chacun de ces blocs est apparié au bloc de l'image flottante qui lui est le plus similaire selon un critère iconique local du type de ceux présentés dans le paragraphe 3.3.3.2, (par exemple, selon une SSD calculée entre les blocs).

L'algorithme est encapsulé dans un schéma itératif proche de celui de l'algorithme ICP décrit en section 3.3.3.1. A chaque itération l , les étapes de l'algorithme sont les suivantes :

1. Chaque bloc défini sur l'image de référence est apparié au bloc de son voisinage dans l'image flottante avec lequel la similarité selon le critère iconique local choisi est maximale. Une variante permettant d'améliorer la robustesse consiste à ne pas traiter les blocs de l'image de référence de trop faible variance (blocs uniformes), considérant qu'ils pourraient être appariés avec n'importe quel bloc uniforme de l'image flottante.
2. La correction de transformation δT_l expliquant au mieux ces nouveaux appariements est estimée par moindres carrés tamisés (éventuellement pondérés en pondérant dans le critère à minimiser chaque appariement de blocs par la similarité atteinte pour cet appariement).
3. La transformation totale est obtenue par composition : $T_l = T_{l-1} \circ \delta T_l$.

Si les blocs utilisés sont suffisamment petits (typiquement $3 \times 3 \times 3$ ou $5 \times 5 \times 5$), ils ne contiennent en général que deux tissus différents au maximum. L'hypothèse de la relation affine entre les intensités des deux blocs paraît donc raisonnable, ce qui justifie l'utilisation du coefficient de corrélation comme mesure de similarité entre les blocs.

3.3.4 Classes de transformation

Le recalage de l'image flottante F sur l'image de référence R consiste à estimer la transformation $\hat{T}_{F \leftarrow R}$ permettant de ré-échantillonner au mieux F dans la géométrie de R . Nous appelons $F \circ \hat{T}_{F \leftarrow R}$ l'image flottante ré-échantillonnée sur R . L'intensité en chaque voxel (x, y, z) de $F \circ \hat{T}_{F \leftarrow R}$ est alors : $F \circ \hat{T}_{F \leftarrow R}(x, y, z) = F(\hat{T}_{F \leftarrow R}(x, y, z)) = F(x', y', z')$, si l'on appelle $(x', y', z') = \hat{T}_{F \leftarrow R}(x, y, z)$ le transformé d'un point (x, y, z) par la transformation $\hat{T}_{F \leftarrow R}$.

3.3.4.1 Transformations linéaires

On appelle *transformation linéaire* toute transformation T pour laquelle $(x', y', z') = T(x, y, z)$ peut s'exprimer comme fonction linéaire de (x, y, z) :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \vec{t} \quad (3.4)$$

où \mathbf{M} est une matrice de taille 3×3 et \vec{t} est un vecteur de translation dans \mathbb{R}^3 . Ceci est équivalent à un produit matriciel simple en coordonnées homogènes :

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = T \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{M} & \vec{t} \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3.5)$$

Les transformations linéaires englobent différents types de transformation ayant des nombres de degrés de liberté (abrégié DOF pour Degree Of Freedom) différents :

- **La transformation rigide** : c'est la composition d'une rotation (3 DOF) et d'une translation (3 DOF). Elle permet de bouger un objet dans l'espace sans le déformer (conservation des angles et des distances).
- **La similitude** : c'est une transformation rigide (6 DOF) avec en plus un facteur d'échelle isotrope (soit 7 DOF au total).
- **La transformation affine** : c'est une transformation rigide (6 DOF) avec en plus des facteurs d'échelles (+ 3 DOF) et des cisaillements (+ 3 DOF) différents selon les 3 directions de l'espace (soit 12 DOF au total).

Le recalage linéaire peut fournir des résultats satisfaisants si les deux images à recaler proviennent d'un même patient. Pour la problématique du recalage inter-patient qui est la nôtre, le recalage linéaire constitue une première approximation mais ne permet pas de compenser les déformations locales liées aux différences anatomiques et morphologiques entre deux patients différents. Nous l'utiliserons donc essentiellement comme initialisation pour l'estimation de la transformation non-linéaire.

3.3.4.2 Transformations non-linéaires

Les transformations linéaires permettent d'amener deux images en correspondance globale mais elles ne permettent pas de compenser les déformations locales

entre les deux images. Pour cela, il est nécessaire de raffiner localement la transformation obtenue en recherchant une transformation moins contrainte c'est-à-dire possédant davantage de degrés de liberté : c'est la classe des transformations non-linéaires, qui englobe en théorie toutes les transformations possibles. Certaines transformations non-linéaires sont dites paramétriques car elles peuvent être paramétrées par un nombre de degrés de liberté qui reste raisonnable. Par exemple, certaines transformations sont définies comme combinaison linéaire de B-Splines [Rueckert 1999] ou de fonctions à base radiale [Rohde 2003]. Un autre exemple est celui des transformations dites *localement affines* [Little 1997, Commowick 2008a], qui sont basées sur des transformations affines définies de façon locale sur des sous-régions de l'image. A l'extrême, il est également possible de définir une transformation non-linéaire par un champ dense de vecteurs de déplacement indépendants les uns des autres (un vecteur par voxel, soit 3 degrés de liberté par voxel). Dans ce dernier cas (ainsi que pour les transformations non-linéaires paramétriques ayant un grand nombre de degrés de liberté), il est alors nécessaire de régulariser la transformation de sorte qu'elle soit suffisamment lisse et reste réaliste par rapport à l'anatomie sous-jacente (de la même façon qu'un terme de régularisation des contours est ajouté dans les modèles déformables). Cette régularisation peut s'appliquer à la transformation elle-même (régularisation élastique) ou bien à la correction de transformation dans le cadre d'algorithmes itératifs (régularisation fluide). Dans certains cas, une contrainte d'inversibilité est également imposée à la transformation [Vercauteren 2009, Leow 2005].

Les transformations non-linéaires sont très largement utilisées en imagerie médicale car elles permettent de représenter des déformations locales entre deux images, ceci grâce à un nombre de degrés de liberté plus élevé que les transformations linéaires. Cependant, l'estimation d'une transformation avec beaucoup de degrés de liberté tend à diminuer la robustesse en augmentant les risques de n'atteindre qu'un optimum local lors de l'optimisation.

3.3.5 Mise en œuvre

Comme nous l'avons mentionné précédemment, plus le nombre de degrés de liberté de la transformation recherchée est élevé, plus le risque de rester bloqué dans un optimum local est grand. Pour limiter ce risque et rendre le recalage plus robuste au bruit dans les images, deux astuces algorithmiques sont fréquemment utilisées.

3.3.5.1 Approche hiérarchique

La première astuce consiste à adopter une approche hiérarchique c'est-à-dire rechercher d'abord une transformation très contrainte (typiquement une transformation rigide), puis relâcher progressivement les contraintes en augmentant les degrés de liberté (typiquement une transformation affine, puis non-linéaire en dernier). A chaque étape, le recalage précédent permet de fournir une bonne initialisation de la transformation recherchée, ce qui limite le risque de tomber dans un optimum local par rapport au cas où l'on rechercherait d'emblée une transformation non-linéaire.

Dans notre cas, nous effectuons les recalages inter-patient en trois étapes successives : recalage rigide, recalage affine, puis recalage non-linéaire. La transformation totale est alors obtenue en composant les transformations affine et non-linéaire : $\hat{T}_{F \leftarrow R} = \hat{T}_{F \leftarrow R}^{aff} \circ \hat{T}_{F \leftarrow R}^{nonlin}$. L'image flottante ré-échantillonnée sur l'image de référence est alors notée :

$$F' = F \circ \hat{T}_{F \leftarrow R} = F \circ \hat{T}_{F \leftarrow R}^{aff} \circ \hat{T}_{F \leftarrow R}^{nonlin} \quad (3.6)$$

Pour les trois types de recalage (rigide, affine, non-linéaire), nous utilisons des algorithmes de recalage par blocs ([Ourselin 2000] pour les recalages rigide et affine, et [Commowick 2007b, Garcia 2010] pour le recalage non-linéaire). Ces algorithmes sont décrits brièvement dans l'annexe A.

3.3.5.2 Approche multi-résolution

Une deuxième astuce algorithmique qui contribue à rendre le recalage plus robuste est l'approche multi-résolution. Cette approche consiste à estimer successivement la transformation sur des représentations sous-échantillonnées à des résolutions croissantes (de la plus grossière à la plus fine) des images flottante et de référence jusqu'à atteindre la résolution initiale des images. A chaque niveau de la pyramide, la résolution est en général doublée par rapport au niveau précédent, et la transformation estimée au niveau précédent est sur-échantillonnée pour servir d'initialisation au niveau courant.

3.4 État de l'art des stratégies de segmentation par atlas

Notations Dans tout le manuscrit, nous appelons P l'image du patient à contourner, et A l'image de l'atlas. Le recalage entre A et P fournit la transformation $T_{A \leftarrow P}$ permettant de déformer la segmentation S_A de l'atlas sur le patient : $S_P^{auto} = S_A \circ T_{A \leftarrow P}$.

Le recalage de l'atlas sur l'image du patient fait l'objet d'un compromis entre régularité et similarité. En effet, les algorithmes de recalage cherchent à optimiser simultanément (ou successivement) ces deux critères. D'un côté, la transformation recherchée doit permettre de rendre l'image flottante suffisamment bien superposée à l'image de référence. D'un autre côté, la transformation recherchée doit être cohérente au niveau anatomique, notamment, elle doit être suffisamment lisse et ne pas présenter de gradients trop forts. Ainsi, plus l'image de l'atlas a une anatomie « proche » de celle de l'image du patient, plus l'optimisation conjointe de ces deux critères sera facile, plus le résultat du recalage sera satisfaisant, et plus la segmentation résultat sera cohérente anatomiquement.

Nous détaillons dans cette partie les différentes stratégies possibles pour définir un atlas, avec leurs avantages et inconvénients. Un état de l'art détaillé est également disponible dans [Rohlfing 2005a].

3.4.1 Atlas unique utilisé pour tous les patients

La stratégie la plus simple à mettre en œuvre est d'utiliser un seul et même atlas pour tous les patients. Dans ce cas, l'atlas choisi doit être le plus « proche » possible du maximum d'individus de la population d'intérêt.

3.4.1.1 Atlas défini par un patient unique

L'atlas unique peut être simplement l'image particulière d'un patient manuellement contourée par un médecin [Gee 1993, Dawant 1999b] ou encore une image artificielle également contourée par un médecin [Bondiau 2005]. Cette solution a été la première introduite historiquement, avec l'atlas de Talairach [Talairach 1988], car elle ne nécessite pas une grande base d'images, une seule image suffisant à créer l'atlas. Elle a l'inconvénient d'introduire un biais à travers l'anatomie de l'image considérée d'une part, et à travers sa segmentation d'autre part. Afin de minimiser l'impact de ce biais, il est recommandé de prendre comme atlas un patient ayant une anatomie « standard » par rapport à l'ensemble de la population, de façon à optimiser les recalages avec le maximum d'individus possibles. Suivant la variabilité anatomique présente dans la population, cette solution peut s'avérer plus ou moins satisfaisante. Si la variabilité anatomique est grande, l'atlas ainsi défini pourra être très proche de certains patients, mais très éloigné d'autres patients de la population étudiée.

3.4.1.2 Atlas moyen ou médian estimé à partir de plusieurs patients

Pour s'affranchir du biais introduit par la stratégie précédente, une solution consiste à estimer à partir d'une base d'images manuellement contourées une anatomie (c'est-à-dire une image et sa segmentation associée) moyenne ou « médiane » et à utiliser cette anatomie comme atlas (cf. figure 3.4).

Choix d'une anatomie « médiane » parmi les images de la base Au lieu de chercher à calculer directement une anatomie moyenne, certaines approches consistent à sélectionner de façon automatique parmi la base d'images celle qui représente le mieux l'anatomie « médiane », et à utiliser cette image et sa segmentation manuelle comme atlas. Ces méthodes sont en général basées sur le recalage deux à deux et le calcul d'une distance pour chaque couple d'images de la base. Dans les méthodes proposées par [Park 2005, Marsland 2003, Parraga 2007], cette distance quantifie d'une manière ou d'une autre la déformation obtenue par recalage des deux images. Park et al. [Park 2005] ne recherchent pas directement l'anatomie « médiane », mais l'anatomie la plus proche de l'anatomie moyenne selon un certain critère. Pour cela, ils appliquent un multidimensional scaling (MDS, [Sammon 1969]) sur la matrice de distance ainsi obtenue et calculent ensuite dans l'espace de représentation euclidienne obtenu la position de l'image moyenne, ainsi que l'image de la base qui lui est la plus proche dans cet espace. D'autres approches

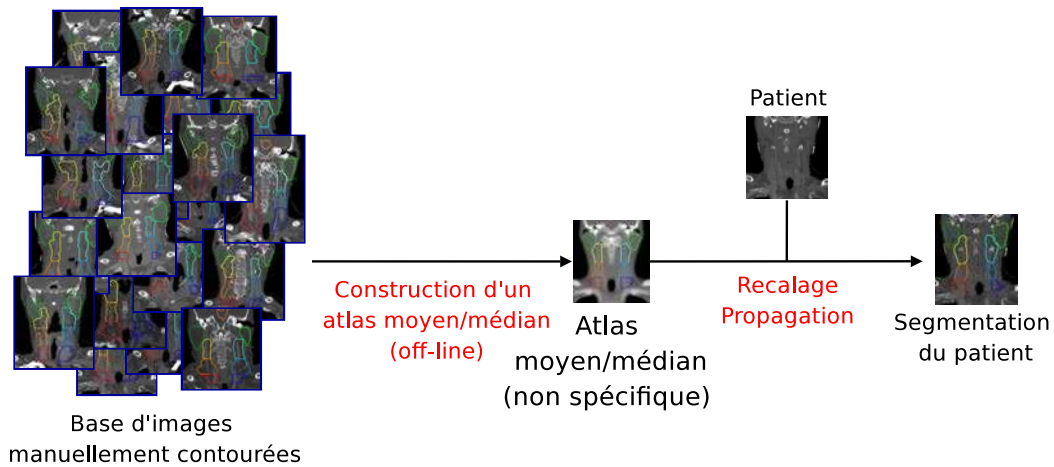


FIG. 3.4 – Utilisation d'un atlas moyen ou médian.

[Marsland 2003, Parraga 2007] choisissent comme atlas l'image de la base qui minimise la somme ou la moyenne de ses distances avec le reste des images de la base.

D'une manière générale, les méthodes citées ici fournissent une anatomie proche de la moyenne, mais leur principal inconvénient est qu'elles introduisent un biais lié à la variabilité de contourage inter- et intra-expert puisqu'alors la segmentation manuelle de l'image choisie est prise comme segmentation pour l'atlas.

Calcul d'une image moyenne et d'une segmentation moyenne Pour s'affranchir des inconvénients liés à la sélection de l'anatomie « médiane » parmi les images de la base, une solution consiste à calculer une anatomie moyenne, c'est-à-dire une image moyenne et une segmentation moyenne. Certaines approches estiment simultanément ces deux éléments [De Craene 2004, Craene 2005, Bhatia 2007, Zhuang 2010b], mais elles impliquent généralement des recalages entre segmentations (ou cartes de probabilités) plutôt qu'entre des images d'intensité, ce qui les rend peu adaptées à notre cas en raison de la grande variabilité de contourage des segmentations de notre base. La plupart des approches estiment de façon découplée l'image moyenne dans un premier temps, puis la segmentation moyenne ensuite (cf. figure 3.5). Une fois que l'image moyenne est construite, on peut déformer toutes les segmentations de la base dans le référentiel de l'image moyenne, et calculer à partir des segmentations normalisées une segmentation moyenne ou une segmentation probabiliste (c'est-à-dire des cartes de probabilité pour chaque structure). Un état de l'art détaillé des méthodes de fusion de segmentations est présenté dans le chapitre 4. Nous présentons ici les principales approches pour estimer l'image moyenne. Le calcul de l'image moyenne implique de mettre en correspondance les images de la base dans un référentiel commun optimal à identifier. Les inconnues de ce problème sont donc : le référentiel commun, et les transformations permettant de

déformer chaque image de la base dans le référentiel commun. L'image moyenne à proprement parler peut ensuite être calculée en déformant toutes les images dans le référentiel commun, et en moyennant leurs intensités dans ce référentiel.

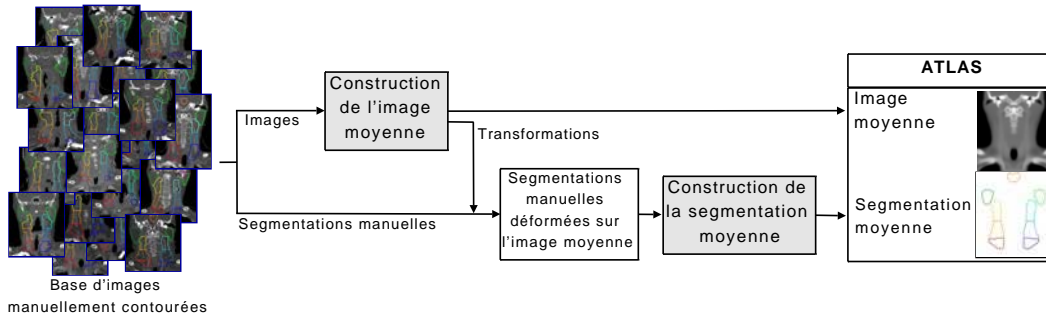


FIG. 3.5 – Construction d'un atlas moyen.

Le premier type d'approche pour calculer une image moyenne consiste à initialiser le référentiel commun par l'une des images de la base puis à calculer une image moyenne dans ce référentiel après y avoir déformé non-linéairement toutes les autres images de la base. L'image moyenne résultante est alors fortement biaisée par le choix de l'image de référence. Pour corriger ce biais, plusieurs méthodes proposent d'appliquer l'inverse de la transformation moyenne pour remettre a posteriori l'atlas dans son référentiel « naturel » [Guimond 2000, Rueckert 2001, Rueckert 2003]. Guimond et al. utilisent un schéma itératif qui alterne entre (1) recalage de toutes les images sur la référence et calcul d'une image d'intensité moyenne dans le référentiel de la référence, et (2) application de la transformation moyenne inverse à l'image d'intensité moyenne pour estimer la nouvelle référence. Une extension de cet algorithme a par ailleurs été proposée pour obtenir une image moyenne symétrique [Grabner 2006]. Rueckert et al. [Rueckert 2001, Rueckert 2003] appliquent le même principe que [Guimond 2000] mais avec une seule itération, et ils utilisent ensuite les principaux modes de variation des déformations des images de la base vers l'image moyenne pour générer plusieurs images moyennes correspondant à la variabilité anatomique observée dans la population.

Alors que les approches du premier type initialisent le référentiel commun « naturel » par l'une des images et cherchent à corriger a posteriori le biais ainsi introduit, les approches du second type cherchent d'emblée à estimer le référentiel commun « naturel » (ainsi que les transformations déformant chaque image dans ce référentiel commun) sans passer par le choix initial d'une référence. Ces méthodes sont généralement appelées *recalage par groupe* (*groupwise registration*), et prennent en entrée les images de la base recalées linéairement entre elles. A noter que le choix de la référence utilisée pour cette mise en correspondance initiale n'introduit pas de biais puisqu'il s'agit juste d'un recalage linéaire. Parmi ces approches, [Studholme 2003] et [Bhatia 2004] estiment le référentiel commun « naturel » en minimisant la somme des déformations des images de la base vers le référentiel commun (ou en lui imposant d'être nulle dans le cas de [Bhatia 2004]) tout en maximisant

une mesure de similarité entre les images ainsi mises en correspondance. [Joshi 2004] et [Lorenzen 2005] ont généralisé cette approche au cas des larges déformations. [Wu 2010] ont étudié la possibilité d'améliorer le recalage par groupe pour obtenir une image moyenne la moins floue possible : dans cette optique, ils ont proposé une extension de [Joshi 2004] qui donne des poids différents et variant au cours de l'optimisation à chaque image et à chaque voxel. [Wang 2010b] ont proposé de décomposer le problème complexe du recalage par groupe selon plusieurs niveaux hiérarchiques. Les images sont tout d'abord grossièrement normalisées puis regroupées de façon hiérarchique selon leur similarité en intensité. Ensuite, étant donnée la pyramide d'images, une image moyenne est calculée par une quelconque méthode de recalage par groupe à chaque niveau de pyramide en commençant par les nœuds inférieurs, jusqu'au nœud supérieur auquel une image moyenne globale est finalement obtenue.

Limitations de l'atlas moyen La construction d'un atlas moyen nécessite d'avoir une base suffisamment grande d'images manuellement contourées pour que la moyenne soit statistiquement significative. Une telle base peut être difficile à obtenir en raison du temps-médecin requis pour les contourages manuels. Cette approche permet en théorie de prendre en compte la variabilité anatomique présente dans la population étudiée, mais en pratique elle connaît ses limites, en particulier pour les localisations à grande variabilité anatomique comme la région ORL. En effet, en cas de variabilité anatomique importante, il est inévitable que certaines anatomies particulières ne soient représentées que de façon marginale dans la base d'images (par exemple patients corpulents ou patients ayant une flexion de colonne importante). Par effet de moyenne, ces anatomies marginales auront peu voire pas d'impact lors du calcul de l'atlas moyen. L'atlas moyen permettra donc de couvrir un plus large spectre d'anatomies que l'atlas défini à partir d'une seule image, mais il aura toujours des difficultés à segmenter correctement les anatomies extrêmes marginalement représentées dans la base d'images.

3.4.2 Construction d'atlas adaptés à l'anatomie du patient

3.4.2.1 Principe

Nous avons détaillé dans le paragraphe précédent les stratégies de segmentation utilisant un seul et même atlas pour segmenter tous les patients, et présenté leurs limites, notamment leur difficulté à segmenter correctement les anatomies peu représentées dans la base. Un moyen de dépasser ces limites est de ne plus utiliser le même atlas pour tous les patients, mais au contraire d'avoir à disposition plusieurs atlas et de choisir pour chaque nouveau patient le ou les atlas le(s) plus approprié(s).

3.4.2.2 Construction hors-ligne d'une base d'atlas

Tout d'abord, il est possible de construire hors-ligne une base de plusieurs atlas correspondant aux différents modes ou sous-populations observé(e)s dans la base d'images. Pour cela, deux stratégies différentes sont envisageables.

Partitionnement de la base d'images et création d'un atlas par sous-groupe

La première solution est de partitionner la base d'images en sous-groupes distincts (appelés *clusters*) et de construire un atlas moyen pour chaque sous-groupe (cf. figure 3.6). La base d'images est alors représentée par ses « modes » principaux sous la forme d'un ensemble restreint d'atlas. Une telle stratégie nécessite de définir (1) le critère utilisé pour partitionner la base d'images, et (2) une méthode pour sélectionner le ou les atlas le(s) plus approprié(s) pour chaque nouveau patient à segmenter. Le critère utilisé pour partitionner la base peut tout d'abord être basé sur des méta-informations, par exemple sujet sain ou malade. Il peut également être basé sur des informations images [Blezek 2007, Sabuncu 2007, Sabuncu 2008, Sabuncu 2009]. Blezek et al. ont introduit le principe de stratification d'atlas [Blezek 2007] en utilisant un algorithme de type « mean-shift » pour déterminer itérativement les modes de la population et une image moyenne pour chaque mode. Cette approche utilise des recalages deux à deux pour définir les distances pour chaque paire d'images. Sabuncu et al. sélectionnent au hasard K images de la base comme centres de sous-groupes, puis utilisent un algorithme EM pour estimer itérativement les probabilités d'appartenance de chaque image à chacun des K sous-groupes ainsi que l'image moyenne de chaque sous-groupe [Sabuncu 2007, Sabuncu 2008, Sabuncu 2009]. Cette méthode a l'avantage d'estimer un partitionnement « soft » (chaque image se voit attribuée une probabilité d'appartenance à chaque sous-groupe), ce qui peut être plus réaliste qu'un partitionnement « hard ». En revanche, le nombre K de sous-groupes doit être spécifié. D'autre part, contrairement à [Blezek 2007], l'algorithme inclut également la sélection du sous-groupe le plus approprié pour toute nouvelle image à segmenter [Sabuncu 2007]. Enfin, Frey et al. ont proposé un algorithme appelé « affinity-propagation » qui permet de partitionner une base d'images, et de déterminer pour chaque sous-groupe l'image qui est la plus représentative du sous-groupe [Frey 2007]. Il est alors envisageable d'utiliser ce sous-ensemble d'images les plus représentatives de la base et leurs segmentations comme une base d'atlas.

Création d'un continuum d'atlas

La seconde solution pour construire hors-ligne une base d'atlas consiste à construire un modèle de régression continu d'atlas selon un ou plusieurs critères, et à instancier un atlas spécifique à partir de ce modèle pour chaque nouveau patient à segmenter. Cette stratégie a été utilisée par Rohlfing et al. pour créer des atlas probabilistes du cerveau selon le sexe et l'âge du patient [Rohlfing 2009a, Rohlfing 2009b]. Murgasova et al. appliquent le même genre de méthode pour le cerveau du fœtus suivant son âge de gestation [Murgasova 2010]. Enfin, [Rueckert 2001] construisent une image moyenne à partir de l'ensemble de la population et calculent les principaux modes de variation des déformations des images de la base vers l'image moyenne. Un continuum d'images moyennes correspondant à la variabilité anatomique observée dans la base d'images peut alors être généré en faisant varier l'influence de chaque mode. La sélection de l'image moyenne la plus appropriée parmi ce continuum n'est cependant pas abordée.

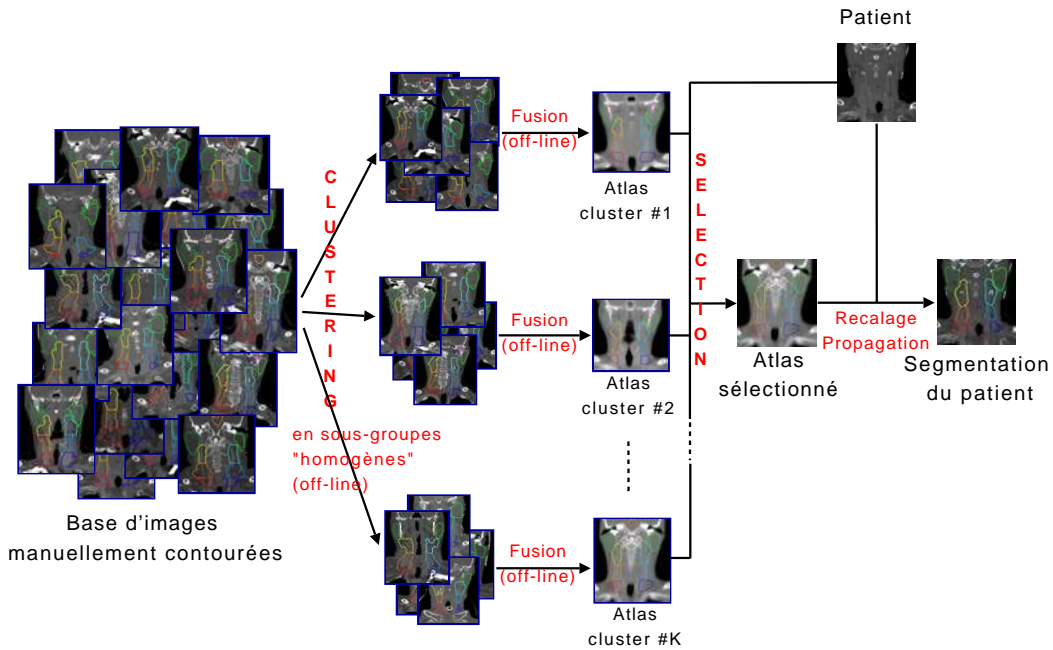


FIG. 3.6 – Partitionnement de la base d'images en sous-groupes homogènes, création d'un atlas par sous-groupe, et utilisation de l'atlas le plus approprié pour chaque patient.

3.4.2.3 Construction à la volée d'atlas adaptés au patient

Principe L'utilisation de l'atlas le plus approprié parmi un nombre restreint d'atlas représentant bien la population d'images peut ne pas suffire lorsque la variabilité anatomique est trop importante dans la population. Un moyen de dépasser ces limites est de ne plus se restreindre seulement aux modes principaux de la population, mais de pousser l'idée à l'extrême en considérant chaque image de la base et sa segmentation manuelle comme un atlas potentiel. Pour chaque nouveau patient à segmenter, il faut alors construire à la volée un atlas adapté à son anatomie à partir des images de la base qui lui sont suffisamment « similaires » (selon un critère à définir). Cette approche est généralement connue sous le nom de *segmentation multi-atlas*, et a été appliquée pour segmenter divers organes, notamment les structures de la sphère ORL [Commowick 2007a, Commowick 2009, Han 2008, Yang 2010], les structures du cerveau [Aljabar 2007, Aljabar 2009, Heckemann 2006, Artaechevarria 2009, Wu 2007], la prostate [Klein 2008], les lobes du poumon [van Rikxoort 2009], ou encore le cœur [Isgum 2009]. Elle nécessite d'une part d'être capable de quantifier la similarité entre l'image du patient et chacune des images de la base, et d'autre part de disposer d'une base d'images suffisamment grande pour être sûr d'y trouver un nombre suffisant d'images dont l'anatomie soit « similaire » à celle du patient à segmenter.

Différents schémas possibles Différentes approches ont été proposées dans la littérature pour construire des atlas adaptés à l'anatomie des patients à partir d'une base d'images contourées. Le schéma le plus simple consiste à sélectionner l'image de la base qui est la plus « similaire » à l'image du patient, et à utiliser cette image et sa segmentation associée comme atlas pour segmenter le patient [Commowick 2007a, Teguh 2010, van Rikxoort 2009, Wu 2007, Gorthi 2009a, Zhang 2009, Barnes 2008]. Cette stratégie est illustrée sur la figure 3.7. Elle a pour principaux inconvénients d'être fortement dépendante de la phase de sélection et de la variabilité de contourage inter- et intra-expert.

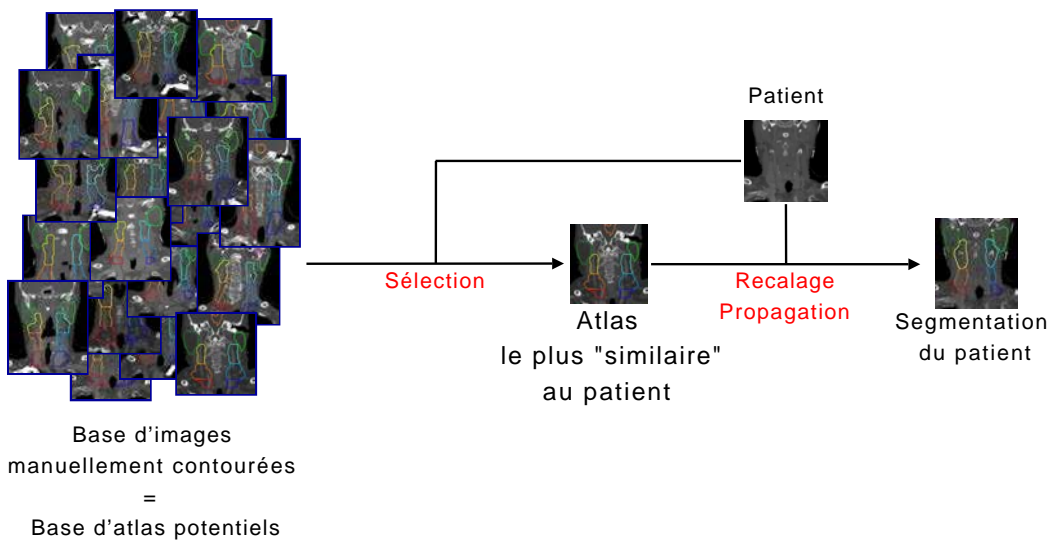


FIG. 3.7 – Sélection de l'image de la base la plus « similaire » au patient, et utilisation de cette image et de sa segmentation comme atlas.

Afin de dépasser ces inconvénients, de nombreuses études ont proposé de sélectionner parmi la base les K images les plus « similaires » à l'image du patient ainsi que leurs segmentations associées, puis de les fusionner d'une façon ou d'une autre pour segmenter le patient. En général, ces méthodes reposent toutes sur les trois étapes suivantes : sélection, fusion, propagation :

- **L'étape de sélection** consiste à sélectionner parmi les images de la base celles qui sont le plus « similaires » à l'image du patient à segmenter.
- **L'étape de fusion** désigne la combinaison des images et/ou segmentations sélectionnées à l'étape de sélection et mises en correspondance dans un référentiel commun.
- **L'étape de propagation** correspond à la déformation non-linéaire de segmentations vers le référentiel du patient à segmenter.

Nous abordons maintenant brièvement les problématiques et mises en œuvre possibles pour chacune de ces trois étapes.

Lors de l'étape dite de sélection, la sélection peut alternativement être remplacée par une pondération des images de la base selon leurs similarités à l'image du patient à segmenter. Dans les deux cas (sélection ou pondération), la similarité peut être mesurée d'un grand nombre de façons différentes : selon les intensités, selon les déformations, ou encore selon des méta-informations. Outre les cas de sélection selon des méta-informations, cette étape implique (1) de mettre en correspondance les images de la base avec le patient à segmenter, et (2) de mesurer leur similarité relative avec le patient à segmenter. Cela nécessite de définir :

- la façon dont sont mises en correspondance les images de la base avec le patient à segmenter (classe de transformations affine ou non-linéaire, algorithme de recalage utilisé, référentiel dans lequel se fait la sélection),
- le critère utilisé pour quantifier la similarité entre l'image du patient et chacune des images de la base après mise en correspondance,
- la méthode de sélection ou de pondération des images de la base en fonction de leur similarité à l'image du patient (nombre à sélectionner dans le cas de la sélection, ou définition des poids dans le cas de la pondération).

Ainsi, l'étape de sélection peut être mise en œuvre de différentes façons, mais elle reste quoi qu'il en soit le premier chaînon commun à toutes les méthodes. Ensuite, les étapes de fusion et de propagation peuvent quant à elles s'articuler différemment, avec des données d'entrée différentes et dans des ordres différents selon les stratégies.

Tout d'abord, certaines méthodes proposent de construire à la volée un atlas à partir des images sélectionnées et de leurs segmentations [Commowick 2009]. La phase de fusion inclut alors le calcul d'une image moyenne et d'une segmentation moyenne, et l'étape de propagation correspond à une simple segmentation par atlas avec l'atlas construit à la volée. Cette stratégie est illustrée sur la figure 3.8.

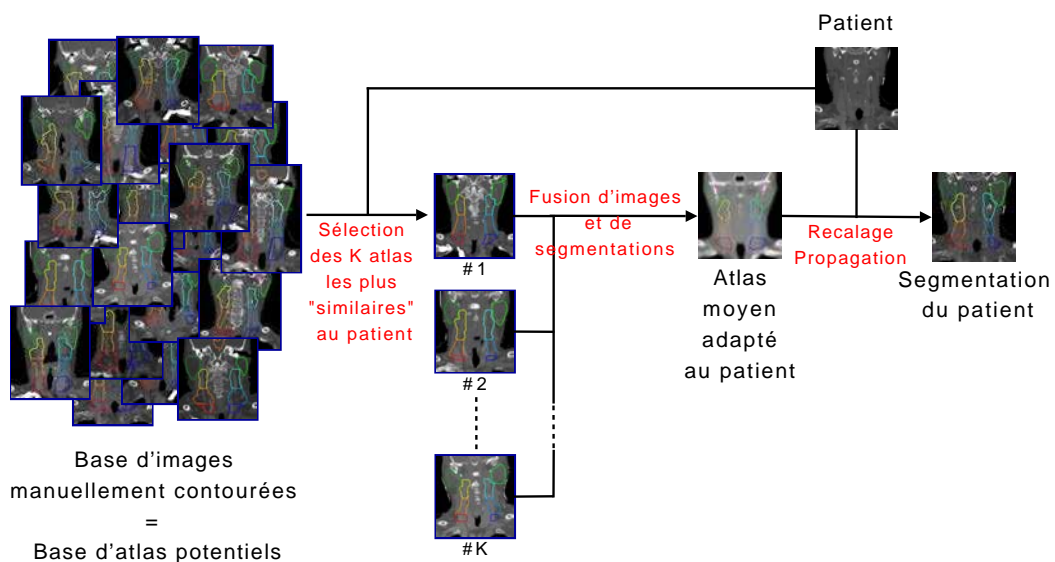


FIG. 3.8 – Sélection des K images de la base les plus « similaires » au patient, et construction à la volée d'un atlas moyen adapté au patient.

Cependant, la plupart des méthodes focalisent directement sur la segmentation de l'image du patient à partir des segmentations des images sélectionnées, sans chercher à construire une image moyenne à partir des images sélectionnées elles-mêmes. Pour ces méthodes, la phase de fusion se réduit alors à la fusion de segmentations mises en correspondance dans un référentiel commun (il n'y a pas de fusion d'images). Il n'y a donc pas à proprement parler *construction d'atlas* au sens où nous l'avons défini dans 3.2.2.4, mais nous utiliserons tout de même le terme de *construction d'atlas adapté au patient* pour désigner ce type d'approche. Le référentiel commun dans lequel sont fusionnées les segmentations des atlas sélectionnés est dans la majorité des études le référentiel du patient à segmenter. Dans ces cas-là, la phase de propagation est faite en amont de la fusion et vise à déformer de façon non-linéaire l'ensemble des segmentations sélectionnées sur l'image du patient. Cette stratégie est illustrée sur la figure 3.9. Le référentiel commun utilisé pour la fusion des segmentations peut également être un référentiel externe (par exemple le référentiel d'un atlas moyen), auquel cas la propagation se fait en deux temps : les segmentations sélectionnées sont d'abord déformées non-linéairement dans le référentiel commun, puis fusionnées dans ce même référentiel, et ensuite la segmentation résultat est déformée de façon non-linéaire sur l'image du patient. Cette stratégie est illustrée sur la figure 3.10. [Heckemann 2006] ont comparé les deux approches (fusion dans le référentiel du patient ou dans un référentiel externe) pour la segmentation des structures du cerveau sur des images IRM.

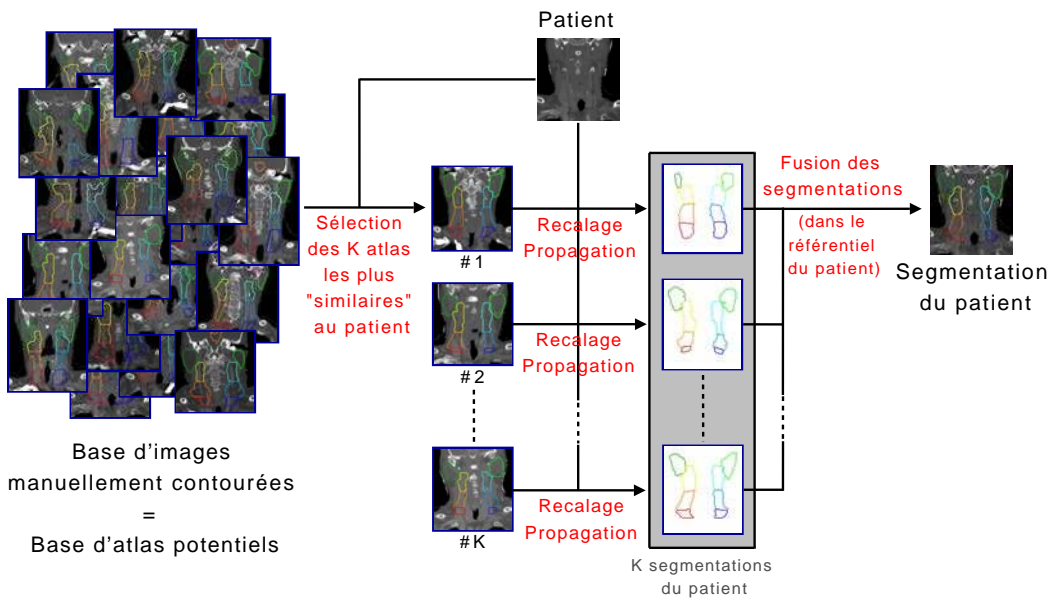


FIG. 3.9 – Sélection des K images de la base les plus « similaires » au patient, et fusion de leurs segmentations dans le référentiel du patient.

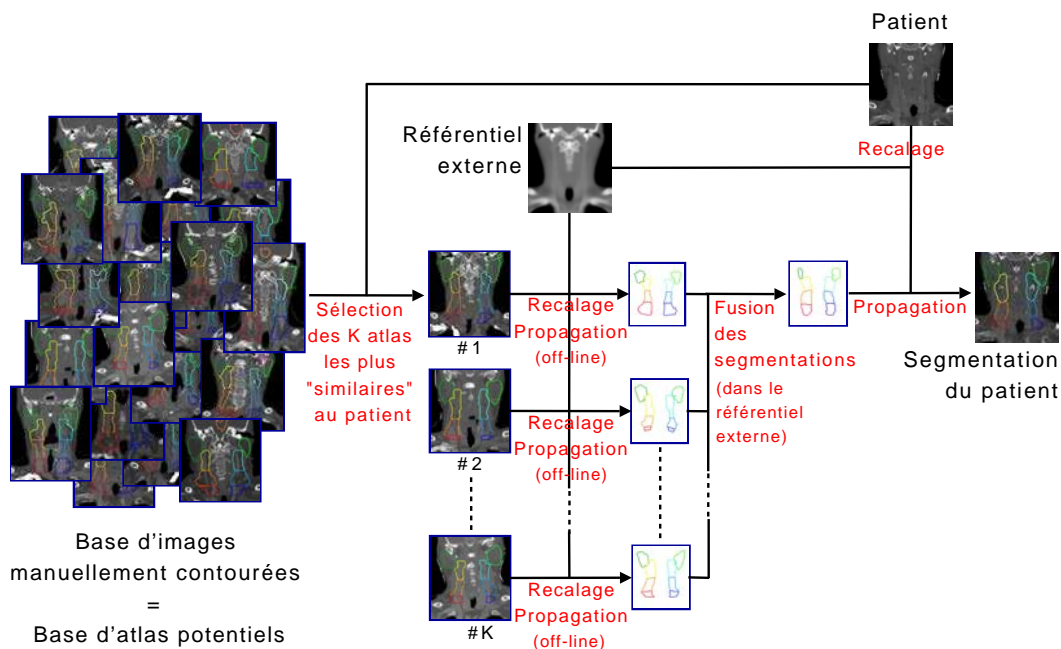


FIG. 3.10 – Sélection des K images de la base les plus « similaires » au patient, et fusion de leurs segmentations dans un référentiel externe.

3.4.3 Conclusion

Les techniques utilisant un seul et même atlas pour segmenter toute nouvelle image (détaillées en section 3.4.1) sont peu robustes à la variabilité anatomique. Les techniques de construction d'atlas adaptés à l'anatomie du patient (détaillées en section 3.4.2) sont en théorie plus efficaces pour segmenter les anatomies marginales mais elles ont en général l'inconvénient d'être beaucoup plus coûteuses en temps de calcul.

Dans notre approche, nous avons tout d'abord cherché à optimiser au mieux les techniques de construction d'un atlas moyen, et en particulier l'étape de construction de la segmentation moyenne. Ce travail est décrit dans le chapitre 4.

Nous nous sommes ensuite intéressés aux algorithmes de segmentation *multi-atlas adaptés au patient*. Les travaux correspondants sont détaillés dans les chapitres 5 et 6. De plus, le chapitre 5 présentera un état de l'art détaillé de ce type de méthodes, notamment les multiples possibilités pour les phases de sélection d'atlas (type de critère de similarité utilisé, sélection globale, régionale, locale des images les plus proches, etc) et de fusion d'atlas, ainsi que certains schémas hybrides qui combinent de façon alternée les phases de sélection et de fusion [van Rikxoort 2010, Langerak 2010].

Calcul d'une segmentation moyenne à partir d'une base incomplète

Sommaire

4.1	Introduction	62
4.1.1	Contexte et problématique	62
4.1.2	Notations	63
4.1.3	État de l'art	64
4.1.4	Organisation du chapitre	66
4.2	Utilisation de l'algorithme STAPLE	66
4.2.1	Description de l'algorithme	66
4.2.2	Gestion des structures manquantes	68
4.2.3	Résultats	68
4.2.4	Analyse de la sur-segmentation obtenue	71
4.2.5	Correction de la sur-segmentation par post-traitement	75
4.2.6	Correction de la gestion des structures manquantes	76
4.3	Algorithme de gestion des structures manquantes dans un calcul de segmentation probabiliste moyenne	77
4.3.1	Principe	77
4.3.2	Mise en œuvre avec STAPLE	78
4.3.3	Mise en œuvre avec un vote par majorité	79
4.4	Calcul d'une segmentation moyenne à partir du degré d'accord local	79
4.4.1	Principe	79
4.4.2	Algorithme global	79
4.5	Évaluation	84
4.5.1	Correction <i>a posteriori</i> de la sur-segmentation dans l'atlas	84
4.5.2	Algorithmes comparés	85
4.5.3	Évaluation quantitative	87
4.5.4	Évaluation qualitative	91
4.5.5	Robustesse aux structures manquantes	91
4.6	Limites de l'utilisation d'un atlas moyen	98
4.7	Conclusion et synthèse des contributions	100

Synthèse des contributions Les principales contributions présentées dans ce chapitre sont les suivantes :

- Analyse de la sur-segmentation obtenue avec l'algorithme STAPLE,
- Correction de la gestion des structures manquantes dans l'algorithme STAPLE,
- Proposition d'une nouvelle approche pour gérer les structures manquantes, et mise en œuvre de cette approche avec STAPLE et avec un vote par majorité,
- Proposition d'une nouvelle méthode de calcul de segmentation moyenne à partir de l'estimation du degré d'accord local entre les segmentations,
- Évaluation des méthodes proposées et de leur robustesse aux structures manquantes.

Publications associées Les travaux sur le calcul d'une segmentation moyenne à partir de l'estimation du degré d'accord local entre les segmentations ont été publiés dans la conférence ISBI 2009 et présentés oralement [Ramus 2009]. L'évaluation clinique de l'atlas moyen obtenu en corrigeant la sur-segmentation *a posteriori* a donné lieu à un résumé dans une conférence française de radiothérapie [Isambert 2008].

4.1 Introduction

4.1.1 Contexte et problématique

Les travaux présentés dans ce chapitre se situent dans le cadre de la construction d'un atlas moyen de la zone ORL à partir d'une base d'images manuellement contourées. La construction d'un tel atlas comprend la construction d'une image moyenne et la construction d'une segmentation moyenne. Nous supposons ici qu'une image moyenne a déjà été calculée, par exemple avec l'algorithme présenté dans [Guimond 2000] et la méthode de recalage présentée dans l'annexe A. Cet algorithme fournit non seulement une image moyenne mais aussi les transformations non-linéaires permettant de ré-échantillonner chacune des images de la base (et donc leur segmentation manuelle) sur l'image moyenne. Ce chapitre est focalisé sur la construction d'une segmentation moyenne à partir des segmentations manuelles de la base mises en correspondance dans le référentiel commun de l'image moyenne.

La base d'images utilisée est composée de 105 images scanner manuellement contourées par des radiothérapeutes et nous a été fournie par le professeur Vincent Grégoire de l'Université Catholique de Louvain (UCL). Les structures d'intérêt sont d'une part les principaux organes à risque de la région ORL (glandes parotides et glandes sous-maxillaires, moelle épinière, tronc cérébral et mandibule) et d'autre part les aires ganglionnaires II, III et IV. Cependant, nous utilisons les images de cette base de façon rétrospective, et comme elles ont été initialement contourées à des fins cliniques, elles ne comportent pas forcément la totalité des structures d'intérêt. Typiquement, en cas de tumeur de faible grade, seules les structures dans le voisinage proche de la tumeur sont prises en compte dans la planification. D'autre part, et sur les conseils du professeur Vincent Grégoire, certaines structures contourées ont été enlevées de notre base soit parce qu'elles incluaient la tumeur (typiquement pour

les aires ganglionnaires), soit parce qu’elles étaient trop déformées par la tumeur (typiquement pour les tumeurs de grosse taille). La figure 4.1 présente le nombre de contourages dont nous disposons pour chacune des 13 structures d’intérêt. Dans la suite du chapitre, nous utilisons parfois les abréviations suivantes pour se référer aux structures : Paro., SM, AG. II, AG. III, AG. IV, Mand., TC, moelle.

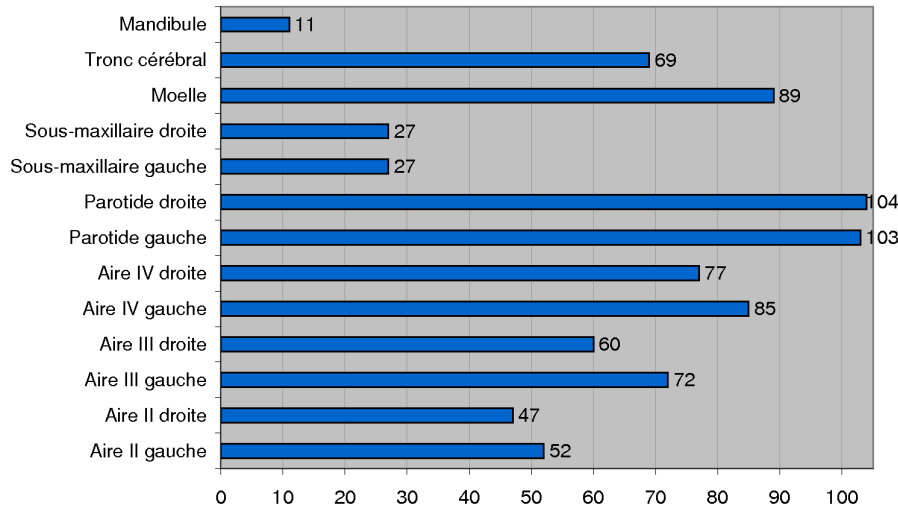


FIG. 4.1 – Nombre de contourages disponibles pour chaque structure d’intérêt parmi les 105 segmentations manuelles de notre base.

Ce chapitre présente d’abord un bref état de l’art sur les différents algorithmes proposés dans la littérature pour estimer une segmentation multilabel moyenne, puis il détaille les méthodes mises en œuvre dans le cadre de cette thèse pour répondre à cette problématique dans le cas particulier de segmentations multilabels incomplètes ainsi que les résultats obtenus pour les structures d’intérêt de la région ORL.

4.1.2 Notations

On note $\{S_k\}_{k \in [1 \dots K]}$ l’ensemble des K segmentations multilabels dont on dispose après normalisation dans le référentiel de l’image moyenne. Les labels $s \in [1 \dots L]$ représentent les structures anatomiques d’intérêt, et pour une segmentation S_k donnée, le label $s = 0$ est attribué aux voxels n’appartenant à aucune des structures anatomiques contourées. Le label 0 ne représente donc pas la même entité dans toutes les segmentations S_k , et cela constitue une difficulté importante lors du calcul d’une segmentation multilabel moyenne. On note \bar{S} la segmentation multilabel moyenne que l’on cherche à estimer. Pour chaque label $s \in [0 \dots L]$, on note K_s le nombre de segmentations multilabels dans lesquelles le label s est présent (on a $K_0 = K$ et $K_s \leq K$ pour tout label $s \in [1 \dots L]$).

4.1.3 État de l'art

De nombreuses méthodes ont été proposées dans la littérature pour construire une segmentation moyenne à partir d'une base de segmentations manuelles normalisées dans un référentiel commun. L'intérêt de ce calcul est que comme les segmentations fournies en entrée sont indépendantes, leur moyenne permet en théorie aux erreurs locales de se compenser. Nous détaillons d'abord les approches basées sur la forme, puis celles basées sur une combinaison voxel par voxel des segmentations S_k .

4.1.3.1 Approches basées sur la forme

Tout d'abord, une première façon d'obtenir la segmentation moyenne est de construire un modèle statistique de forme (Statistical Shape Model, SSM) à partir des segmentations normalisées. Cette approche a par exemple été utilisée pour le pelvis [Meller 2004], le foie [Lamecker 2002] ou encore le ventricule gauche [Cootes 1994]. Cependant, elle nécessite de paramétrer les contours pour en extraire des points caractéristiques. Cette tâche est très délicate pour la région ORL en particulier en raison de la très grande variabilité de forme des structures d'un patient à l'autre. De plus, ces approches sont en général utilisées pour des segmentations monolabels alors que nous avons ici affaire à des segmentations multilabels. Rohlfing et al. ont proposé un algorithme de calcul de segmentation moyenne également basé sur la forme des segmentations données en entrée, et plus particulièrement sur des cartes de distance des labels dans chacune des segmentations [Rohlfing 2005b, Rohlfing 2007]. En chaque voxel, une distance moyenne à chaque label est calculée à partir des K cartes de distances de ce label issues des segmentations S_k , et le label minimisant cette distance moyenne est choisi. Cet algorithme a l'avantage de produire une segmentation relativement lisse, tout comme l'ensemble des algorithmes basés sur la forme.

4.1.3.2 Approches basées sur la combinaison des segmentations voxel par voxel

Une alternative aux approches basées sur la forme consiste à combiner pour chaque voxel i les K votes $\{S_k(i)\}_{k \in [1 \dots K]}$ pour en déduire soit les probabilités d'appartenance de ce voxel à chacun des labels possibles, soit directement le label le plus probable. Cette approche s'inscrit dans le cadre plus général de la combinaison (ou fusion) de plusieurs classifieurs quels qu'ils soient. Un état de l'art sur les méthodes de base permettant de combiner plusieurs classifieurs est présenté dans [Kittler 1998, Kuncheva 2007]. Parmi les plus utilisées, nous mentionnons la règle du vote par majorité, qui consiste à attribuer à chaque voxel i le label le plus représenté parmi les K votes $\{S_k(i)\}_{k \in [1 \dots K]}$. Ceci est équivalent à calculer une carte de probabilité pour chaque label, puis à les seuiller à 0.5. Des variantes du vote par majorité peuvent être utilisées pour générer une segmentation moyenne plus lisse, ceci s'avère surtout utile lorsque le nombre de segmentations données en entrée est faible. Une première solution consiste à calculer un vote par majorité sur

un voisinage pour chaque voxel (utilisé par exemple dans [Qazi 2010]). Une seconde mise en œuvre assez similaire consiste à appliquer un filtrage gaussien aux cartes de probabilité avant de les seuiller à 0.5.

L'inconvénient du vote par majorité est qu'il ne permet pas de gérer les données aberrantes. Si certaines segmentations sont plus pertinentes que d'autres, il est alors intéressant de leur donner davantage de poids dans la moyenne. Par exemple dans notre cas, les segmentations d'entrée provenant de la déformation de la segmentation d'une image à l'anatomie atypique sur l'image moyenne peuvent être moins pertinentes à cause d'erreurs de recalage. Plusieurs algorithmes ont donc été proposés dans la littérature pour gérer efficacement les données aberrantes. Ces méthodes alternent typiquement entre deux étapes : (1) la quantification de la « qualité » de chacune des segmentations S_k , et (2) la moyenne pondérée des segmentations avec prises en compte de leurs « qualités » respectives. Parmi ces approches, l'algorithme STAPLE proposé par Warfield et al. [Warfield 2004] est une mise en œuvre itérative de type EM (Expectation-Maximization) de ces deux étapes. A chaque itération, la qualité d'une segmentation est estimée en la comparant à la version courante de l'estimation de la segmentation moyenne avec des mesures de type sensibilité/spécificité. Langerak et al. ont présenté une méthode alternative à STAPLE (appelée SIMPLE) dans laquelle la pondération de segmentations est remplacée par une sélection des segmentations les plus pertinentes au vu des paramètres de qualité [Langerak 2009, Langerak 2010]. Dans leur approche, la qualité de chaque segmentation est quantifiée par une mesure de chevauchement avec l'estimation courante de la segmentation moyenne, et un seuil sur cette mesure doit alors être choisi pour la sélection des segmentations les plus pertinentes. Enfin d'autres méthodes consistent à pondérer localement ou globalement chaque segmentation par une mesure quantifiant la qualité du recalage ayant permis de déformer la segmentation dans le référentiel de calcul (une mesure de similarité d'intensité entre les deux images recalées est en général utilisée). Cependant, ces approches ne sont pas adaptées pour notre problématique (calcul d'une segmentation moyenne dans le référentiel d'une image moyenne) car le flou de l'image moyenne rend peu pertinente l'utilisation d'une mesure de similarité en intensité.

4.1.3.3 État de l'art sur la gestion des structures manquantes

Les méthodes citées dans les paragraphes précédents sont typiquement conçues pour construire une segmentation multilabel moyenne à partir d'une base de segmentations multilabels complète, c'est-à-dire avec aucun label manquant dans toutes les segmentations. Plusieurs extensions de STAPLE ont été proposées pour prendre en compte les labels manquants. Tout d'abord, une initialisation adéquate combinée à une mise à jour partielle lors des itérations des paramètres de qualité des segmentations incomplètes a été proposée dans la thèse de Olivier Commowick [Commowick 2007b]. Landman et al. ont présenté une extension de STAPLE appelée STAPLER dans laquelle les voxels de labels 0 sont ignorés [Landman 2010]. Enfin, Commowick et al. ont proposé une autre extension de STAPLE pour gérer les labels

manquants. Cette extension consiste à contraindre les paramètres de qualité des segmentations à évoluer au cours des itérations selon une certaine loi fixée *a priori* et permettant de prendre en compte les structures manquantes [Commowick 2010].

4.1.4 Organisation du chapitre

La section 4.2 présente plus en détail l'algorithme STAPLE ainsi que les résultats obtenus en gérant les structures manquantes comme suggéré dans [Commowick 2007b, Commowick 2008b]. Nous nous attachons également dans cette partie à analyser le phénomène de sur-segmentation constaté, et à proposer des solutions pour la corriger. La section 4.3 propose un schéma permettant de gérer facilement les structures manquantes utilisable avec tout algorithme de calcul de segmentation moyenne probabiliste. Nous présentons sa mise en œuvre dans le cadre de STAPLE et dans le cadre d'un simple vote par majorité. Dans la section 4.4, nous proposons une nouvelle méthode de calcul de la segmentation moyenne basée sur l'estimation du degré d'accord local entre les segmentations données en entrée. Enfin, la section 4.5 compare les résultats obtenus avec les différentes méthodes proposées à ceux obtenus avec les méthodes de la littérature.

4.2 Utilisation de l'algorithme STAPLE

4.2.1 Description de l'algorithme

4.2.1.1 Principe

La méthode STAPLE (Simultaneous Truth And Performance Level Estimation) [Warfield 2004] est un algorithme de type EM (Expectation Maximization) [Dempster 1977, Leemput 1999] qui estime conjointement à partir d'un ensemble de K segmentations notées S_k : (1) la segmentation consensus \bar{S} (considérée comme variable cachée), et (2) des paramètres de performance quantifiant les qualités des segmentations données en entrée par rapport à la segmentation consensus \bar{S} . Des valeurs initiales doivent alors être données aux paramètres de qualité (traduisant les connaissances *a priori* sur la qualité de chacune des segmentations), puis les étapes d'espérance et de maximisation sont ensuite alternées jusqu'à convergence.

4.2.1.2 Notations

La qualité de la segmentation S_k à l'itération t est représentée par une matrice θ_k^t de taille $(L+1) \times (L+1)$ dont l'élément $\theta_{ks's}^t$ représente la probabilité $p(S_k(i) = s' | \bar{S}(i) = s)$ à l'itération t qu'un voxel i soit attribué au label s' dans la segmentation S_k alors qu'il appartient au label s . On note $\{W_{si}^t\}_{s \in [0 \dots L], i \in [1 \dots N]}$ la segmentation probabiliste associée à l'estimation de la segmentation consensus \bar{S} à l'itération t : elle représente la probabilité de chaque voxel $i \in [1 \dots N]$ d'appartenir à chaque label $s \in [0 \dots L]$ à l'itération t . Les probabilités *a priori* $\pi_s = p(\bar{S}(i) = s)$ d'avoir un

voxel i attribué à chacun des labels $s \in [0 \dots L]$ quelle que soit sa position spatiale sont calculées à partir des S_k .

4.2.1.3 Étape d'espérance

La phase d'espérance consiste à calculer voxel par voxel la segmentation probabiliste consensus à partir des segmentations données en entrée S_k et de leurs paramètres de qualité calculés à l'itération précédente θ_k^{t-1} :

$$W_{si}^t = p\left(\bar{S}(i) = s \mid \{S_k(i)\}_{k \in [1 \dots K]}, \{\theta_k^{t-1}\}_{k \in [1 \dots K]}\right) \quad (4.1)$$

Étant donnée la probabilité π_s a priori du label s , la loi de Bayes conduit successivement à :

$$\begin{aligned} W_{si}^t &= \frac{\pi_s \cdot p\left(\{S_k(i)\}_{k \in [1 \dots K]} \mid \bar{S}(i) = s, \{\theta_k^{t-1}\}_{k \in [1 \dots K]}\right)}{p\left(\{S_k(i)\}_{k \in [1 \dots K]} \mid \{\theta_k^{t-1}\}_{k \in [1 \dots K]}\right)} \\ &= \frac{\pi_s \cdot p\left(\{S_k(i)\}_{k \in [1 \dots K]} \mid \bar{S}(i) = s, \{\theta_k^{t-1}\}_{k \in [1 \dots K]}\right)}{\sum_{l=0}^L \pi_l \cdot p\left(\{S_k(i)\}_{k \in [1 \dots K]} \mid \bar{S}(i) = l, \{\theta_k^{t-1}\}_{k \in [1 \dots K]}\right)} \end{aligned} \quad (4.2)$$

De plus, en considérant que les K segmentations S_k sont indépendantes les unes des autres, on a aussi :

$$\begin{aligned} p\left(\{S_k(i)\}_{k \in [1 \dots K]} \mid \bar{S}(i) = s, \{\theta_k^{t-1}\}_{k \in [1 \dots K]}\right) &= \prod_{k=1}^K p\left(S_k = S_k(i) \mid \bar{S}(i) = s, \theta_k^{t-1}\right) \\ &= \prod_{k=1}^K \theta_{kS_k(i)s}^{t-1} \end{aligned} \quad (4.3)$$

On obtient donc :

$$\forall s \in [0 \dots L] \quad W_{si}^t = \frac{\pi_s \cdot \prod_{k=1}^K \theta_{kS_k(i)s}^{t-1}}{\sum_{l=0}^L \pi_l \cdot \prod_{k=1}^K \theta_{kS_k(i)l}^{t-1}} = \frac{w_{si}^t}{\sum_{l=0}^L w_{li}^t} \quad (4.4)$$

4.2.1.4 Étape de maximisation

La phase de maximisation détermine quant à elle les paramètres de qualité $\{\theta_k^t\}_{k \in [1 \dots K]}$ qui maximisent la log-vraisemblance $Q(\{\theta_k\}_{k \in [1 \dots K]}) = \sum_{i=1}^N \sum_{k=1}^K \sum_{s=0}^L W_{si} \cdot \log(\theta_{kS_k(i)s})$ étant donné l'estimation courante de la segmentation probabiliste consensus. On peut montrer que les $\{\theta_k^t\}_{k \in [1 \dots K]}$ définis comme suit sont solution de la maximisation :

$$\forall s \in [0 \dots L] \quad \forall s' \in [0 \dots L] \quad \theta_{ks's}^t = \frac{\sum_{i: S_k(i)=s'} W_{si}^t}{\sum_i W_{si}^t} \quad (4.5)$$

4.2.2 Gestion des structures manquantes

L'algorithme STAPLE ainsi présenté permet d'estimer la segmentation consensus \bar{S} à partir d'une base de segmentations $\{S_k\}_{k \in [1 \dots K]}$ complète, c'est-à-dire sans structures manquantes. Nous résumons dans ce paragraphe une adaptation de cet algorithme pour le cas des structures manquantes proposée dans [Commowick 2007b].

D'une part, le label 0 étant artificiellement présent à la place des structures non segmentées, il est nécessaire de modifier le calcul des probabilités *a priori* π_s de façon à ne pas sur-estimer sa probabilité *a priori* π_0 . Pour cela, la probabilité *a priori* de chaque structure $s \in [1 \dots L]$ est estimée en se restreignant aux K_s segmentations dans lesquelles cette structure a été segmentée :

$$\forall s \in [1 \dots L] \quad \pi_s = \frac{1}{K_s} \frac{1}{N} \sum_{i=1}^N \sum_{\substack{k \in [1 \dots K] \\ \text{ayant segmenté } s}} \delta(S_k(i), s) \quad (4.6)$$

où δ représente le symbole de Kronecker. La probabilité *a priori* du label 0 est alors obtenue par : $\pi_0 = 1 - \sum_{s=1}^L \pi_s$.

D'autre part, le paramètre $\theta_{ks's}$ doit être défini de façon particulière si le label s ou s' n'est pas segmenté dans la segmentation S_k . La solution mise en œuvre dans [Commowick 2007b] consiste à donner à ces paramètres les valeurs suivantes et à les garder fixes tout au long de l'algorithme :

- $\forall s \in [0 \dots L] \quad \theta_{ks's} = 0$ pour toute structure s' non segmentée dans S_k ;
- De plus, si une structure $s \in [1 \dots L]$ n'est pas présente dans S_k , on considère qu'il y a une forte probabilité α (par exemple $\alpha = 0.75$) pour que les voxels correspondant à la structure aient été attribués au label 0. En désignant par L_k le nombre de labels présents dans S_k (en comptant le label 0), cela se traduit par :

$$\begin{cases} \theta_{k0s} = \alpha \\ \theta_{ks's} = \frac{1-\alpha}{L_k-1} & \text{pour tout label } s' \in [1 \dots L] \text{ segmenté dans } S_k \\ \theta_{ks's} = 0 & \text{pour tout label } s' \in [1 \dots L] \text{ non segmenté dans } S_k \end{cases} \quad (4.7)$$

4.2.3 Résultats

Nous avons appliqué l'algorithme STAPLE adapté au cas de structures manquantes comme expliqué dans la section précédente sur notre base de 105 segmentations incomplètes mises en correspondance dans le référentiel de l'image moyenne obtenue avec l'algorithme de Guimond et al. [Guimond 2000]. L'évaluation de l'atlas moyen ainsi construit (image moyenne et segmentation moyenne) sur un patient P donné se fait selon les trois étapes suivantes :

- recalage non-linéaire entre l'image moyenne et l'image du patient P ,
- déformation de la segmentation moyenne sur le patient P ,
- comparaison de la segmentation automatique ainsi obtenue avec la segmentation manuelle de P .

En toute rigueur, cette évaluation doit se faire sur un ensemble de patients différents de ceux ayant servi à construire l'atlas moyen afin de n'introduire aucun biais. Cela requiert de mettre en œuvre un protocole de type *Leave-One-Out* : on extrait successivement chacun des 105 patients de la base et on construit à partir des 104 patients restants un atlas moyen (image moyenne et segmentation moyenne) que l'on évalue ensuite comme détaillé plus haut. La construction de 105 images moyennes selon l'algorithme de Guimond et al. s'avère cependant très lourde en terme de temps de calcul. Nous avons donc choisi de construire un seul et unique atlas moyen à partir des 105 patients de la base, et de l'évaluer sur le même ensemble de 105 patients. Nous faisons l'hypothèse que le biais introduit est négligeable étant donné le grand nombre de patients, et ceci a été confirmé par des comparaisons quantitatives avec et sans *Leave-One-Out* (non présentées ici).

4.2.3.1 Évaluation qualitative

La figure 4.2 présente des résultats visuels obtenus avec l'algorithme STAPLE. La ligne supérieure (a, b, c, d) présente différentes coupes de l'atlas moyen. La ligne du milieu (e, f, g, h) montre sur différentes coupes la segmentation automatique obtenue pour un patient donné en utilisant l'atlas moyen ainsi construit, tandis que la ligne du bas (i, j, k, l) montre la segmentation manuelle de ce même patient. Les structures sont globalement bien localisées mais elles sont plus larges que celles de la segmentation manuelle. Ceci est mis en évidence par les flèches blanches sur les images (e) à (l).

4.2.3.2 Évaluation quantitative

L'annexe B présente différentes mesures permettant de comparer quantitativement une segmentation automatique à une segmentation de référence (en l'occurrence la segmentation manuelle). La figure 4.3 présente les valeurs moyennes des mesures quantitatives obtenues pour les différentes structures sur les 105 patients : sensibilité et spécificité sont présentées sur le graphique de gauche, et le log-ratio entre les volumes automatique et manuel (mesure VOL explicitée dans l'équation B.1) est présenté sur le graphique de droite.

La sur-segmentation observée qualitativement est confirmée par les mesures quantitatives. Tout d'abord, le graphique de gauche de la figure 4.3 montre que l'ensemble des structures à l'exception de la mandibule se situent au-dessous de la droite $SENS = SPEC$. Comme mis en évidence dans l'annexe B (cf. figure B.2), cette configuration (sensibilité élevée et spécificité faible) traduit une sur-segmentation. Les aires ganglionnaires IV sont les structures pour lesquelles ce déséquilibre entre sensibilité et spécificité est le plus marqué, et sont donc *a priori* les structures les plus sur-segmentées. La mandibule au contraire semble légèrement sous-segmentée. Ces résultats sont confirmés par le graphique de droite de cette même figure. En effet, la mandibule est la seule structure pour laquelle la mesure VOL est négative, traduisant une sous-segmentation. L'ensemble des autres structures possèdent des

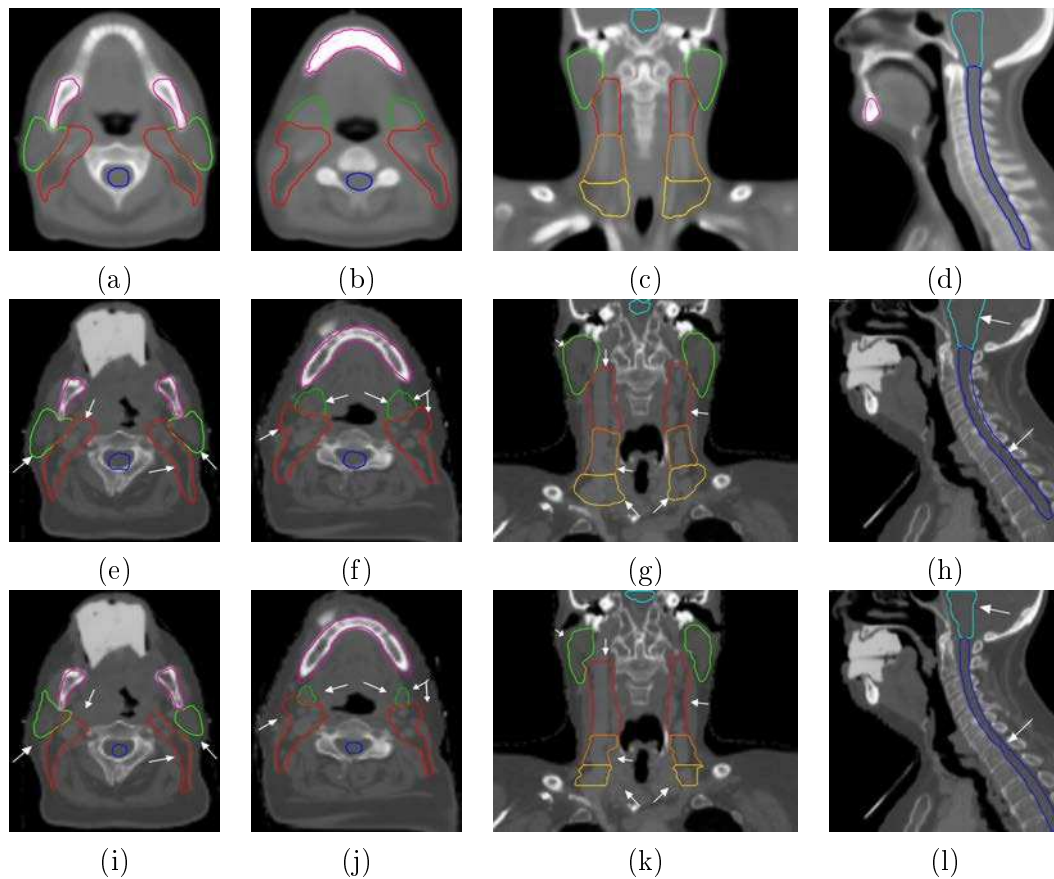


FIG. 4.2 – Résultats visuels obtenus avec l'algorithme STAPLE. La première ligne (a, b, c, d) montre des coupes de l'atlas moyen. La seconde ligne (e, f, g, h) montre la segmentation automatique obtenue pour un patient donné. La dernière ligne (i, j, k, l) montre la segmentation manuelle correspondante. Les aires ganglionnaires II, III et IV sont respectivement représentées en rouge, orange, jaune. Les parotides sont représentées en vert clair, et les sous-maxillaires en vert foncé. La mandibule, la moelle épinière et le tronc cérébral sont respectivement représentées en rose, bleu foncé et bleu clair.

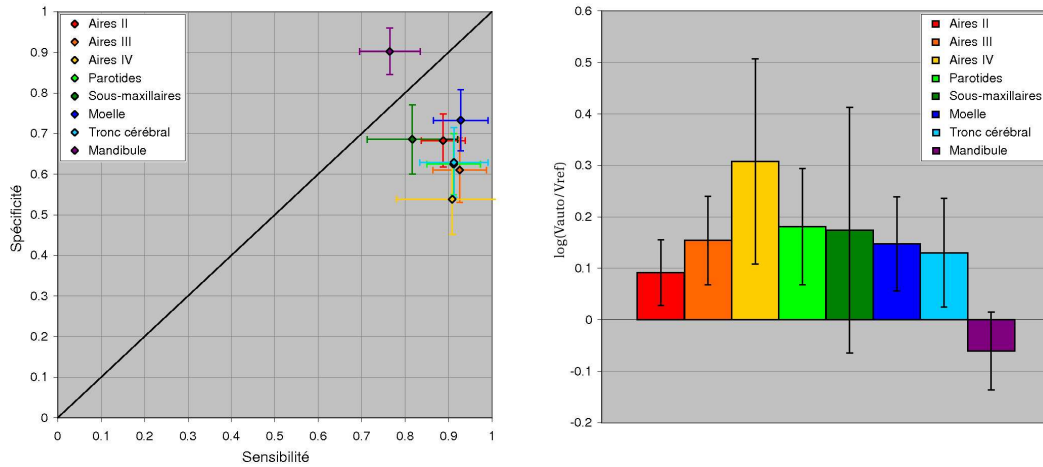


FIG. 4.3 – Mesures quantitatives moyennes obtenues pour les différentes structures sur les 105 patients. A gauche, sensibilité et spécificité. A droite, log-ratio entre les volumes automatique et manuel (mesure VOL).

valeurs significativement supérieure à 0 qui quantifient une sur-segmentation. Là encore, les structures les plus sur-segmentées sont les aires IV.

4.2.4 Analyse de la sur-segmentation obtenue

4.2.4.1 Positionnement du problème

Le problème de sur-segmentation ne semble pas être lié au recalage non-linéaire ou à la propagation de la segmentation puisqu'il est déjà observé dans l'atlas moyen lui-même (cf. première ligne de la figure 4.2). La sur-segmentation vient donc du calcul de la segmentation moyenne fait par STAPLE. Nous proposons dans ce paragraphe une analyse mettant en évidence l'origine du problème dans STAPLE.

Bien qu'observée à des degrés divers pour chaque structure, la sur-segmentation est cependant un problème présent aux frontières entre chaque structure $s \in [1 \dots L]$ et le fond de l'image (label 0). Plus précisément, elle se caractérise par des voxels classifiés comme appartenant à une structure $s \in [1 \dots L]$ alors qu'ils auraient dû être classifiés comme appartenant au label 0. En d'autres termes, et selon les notations utilisées dans l'équation 4.4, cela correspond à une sous-estimation de w_{0i} par rapport aux w_{si} pour les voxels i correspondants. Nous allons mettre en évidence le processus menant à cette sous-estimation.

4.2.4.2 Remarques préliminaires

Facteurs contribuant à un masque de grande taille Nous rappelons que le masque de calcul utilisé dans STAPLE est défini comme l'union des segmentations S_k . Dans notre cas, nous constatons que le masque de calcul est « grand par rapport aux segmentations », c'est-à-dire qu'on a une majorité de voxels du masque qui

n'appartiendront à aucune structure dans la segmentation consensus. Deux facteurs combinés permettent d'expliquer ce phénomène :

- la grande variabilité (de forme et de taille) parmi les segmentations S_k ,
- la configuration spatiale des structures : chaque structure est en majeure partie en contact avec le fond de l'image plutôt qu'avec une autre structure.

La configuration spatiale des structures implique que la variabilité de forme et de taille des segmentations se traduit directement par une augmentation du nombre de voxels du label 0 dans le masque.

Influence de la taille du masque sur le calcul des θ_{ks0}^t et θ_{k00}^t Le paramètre θ_{k00}^t peut s'exprimer de la façon suivante :

$$\theta_{k00}^t = \frac{\sum_{i:S_k(i)=0} W_{0i}^t}{\sum_{i:S_k(i)=0} W_{0i}^t + \sum_{i:S_k(i)\neq 0} W_{0i}^t} \quad (4.8)$$

Cette formule met en évidence que θ_{k00}^t correspond en fait à une généralisation probabiliste de la mesure de spécificité présentée dans l'équation B.4 de l'annexe B. Les Vrais Négatifs correspondent au terme $\sum_{i:S_k(i)=0} W_{0i}^t$, tandis que les Faux Positifs

correspondent au terme $\sum_{i:S_k(i)\neq 0} W_{0i}^t$.

A $\{W_{0i}\}_{i\in[1\dots N]}$ et $\{S_k(i)\}_{i\in[1\dots N]}$ fixés, plus le masque de calcul est grand, plus le nombre de Vrais Négatifs est élevé, et plus la spécificité est proche de 1. De la même façon, en version probabiliste : plus le masque est grand, plus le terme $\sum_{i:S_k(i)=0} W_{0i}^t$

est grand par rapport au terme $\sum_{i:S_k(i)\neq 0} W_{0i}^t$, et plus θ_{k00}^t est proche de 1 (et donc plus les θ_{ks0}^t pour tout $s \in [1 \dots L]$ sont proches de 0).

Influence des structures manquantes sur le calcul des θ_{ks0}^t et θ_{k00}^t Lorsqu'un label $s \in [1 \dots L]$ est absent dans une segmentation S_k :

- d'une part $\theta_{ks0}^t = 0$ à chaque itération t ;
- d'autre part, les voxels correspondant à la structure s étant étiquetés 0, cela contribue à augmenter artificiellement θ_{k00}^t .

Les structures manquantes dans les segmentations S_k vont donc accentuer artificiellement l'effet produit par la grande taille du masque expliqué précédemment, à savoir : $\theta_{k00}^t \gg \theta_{ks0}^t$ pour tout $s \in [1 \dots L]$.

Influence de la taille du masque sur les probabilités *a priori* π_s La grande taille du masque contribue également à augmenter la probabilité *a priori* π_0 du label 0 par rapport aux probabilités *a priori* π_s des autres labels $s \in [1 \dots L]$. Le facteur

π_0 a cependant moins d'impact que les θ_{ks0} dans le calcul de w_{i0} car il n'est présent qu'une seule fois (contrairement aux θ_{ks0} qui sont dans le produit).

4.2.4.3 Mise en évidence du processus menant à la sur-segmentation

Nous allons étudier le problème au niveau de la frontière entre une structure $s \in [1 \dots L]$ donnée et le fond de l'image (label 0). Pour simplifier, on considère que tous les voxels de cette zone ont été assignés dans les segmentations experts S_k soit au label s , soit au label 0. Cette hypothèse ne limite cependant pas la validité de l'explication qui va suivre.

Évolution de la sur-segmentation au cours des itérations Tout d'abord, on remarque expérimentalement que la taille de la segmentation consensus à l'issue de la première itération (notée $\bar{S}^{t=1}$) est approximativement correcte, il n'y a pas de sur-segmentation à ce stade. Les w_{si} correspondants ont été calculés à partir des $\theta_{ks's}$ donnés en initialisation. Cela signifie que la sur-segmentation n'est pas liée à une mauvaise initialisation. C'est au contraire un problème qui apparaît au fur et à mesure des itérations. On constate en effet que la segmentation consensus grossit au fur et à mesure des itérations, jusqu'à se stabiliser à sa taille définitive. Nous allons expliquer ce phénomène.

Calcul de $w_{0i}^{t=2}$ en bordure externe de $\bar{S}^{t=1}$ Soit i un voxel en bordure externe de la structure s dans la segmentation consensus $\bar{S}^{t=1}$ obtenue à l'issue de la première itération. A l'itération 2, $w_{0i}^{t=2}$ sera de la forme :

$$w_{0i}^{t=2} = \pi_0 \cdot \prod_{k:S_k(i)=s} \theta_{ks0}^{t=1} \cdot \prod_{k:S_k(i)=0} \theta_{k00}^{t=1} \quad (4.9)$$

On appelle $n_s^K(i) = \sum_{k=1}^K \delta(S_k(i), s)$ le nombre de segmentations parmi les K ayant attribué le voxel i au label s . En pratique, de par la grande variabilité parmi nos segmentations, $n_s^K(i)$ est proche de $\frac{K_s}{2}$. Dans le calcul de $w_{0i}^{t=2}$, on aura donc environ $\frac{K_s}{2}$ facteurs de type $\theta_{ks0}^{t=1}$, et $K - \frac{K_s}{2}$ facteurs de type $\theta_{k00}^{t=1}$.

Or nous avons précédemment mis en évidence que la grande taille de notre masque de calcul par rapport aux segmentations ainsi que les structures manquantes conduisaient à la configuration $\theta_{ks0}^t \ll \theta_{k00}^t$. De ce fait, un faible nombre de facteurs de type θ_{ks0}^t dans le calcul de $w_{0i}^{t=2}$ suffit à le faire chuter. Le calcul de $w_{si}^{t=2}$ fait quant à lui intervenir des facteurs de types θ_{kss}^t et θ_{k0s}^t qui ne dépendent ni de la taille du masque ni des structures manquantes, et ne présentent donc pas une telle disproportion. On a donc $w_{0i}^{t=2} \ll w_{si}^{t=2}$, puis naturellement $W_{0i}^{t=2} \ll W_{si}^{t=2}$, ce qui conduit le voxel i à être assigné à la structure s dans la segmentation consensus $\bar{S}^{t=2}$ à la deuxième itération.

Calcul de $w_{0i}^{t=2}$ plus loin de la bordure externe de $\bar{S}^{t=1}$ Considérons maintenant un voxel i un peu plus loin de la bordure de $\bar{S}^{t=1}$. Deux éléments diffèrent du

cas précédent (voxel i en bordure externe immédiate de $\bar{S}^{t=1}$). Les segmentations S_k ayant assigné ce voxel à s sont d'une part moins nombreuses (c'est-à-dire $n_s^K(i)$ plus faible). Et d'autre part, elles correspondent à des segmentations globalement plus larges, donc possédant moins de Vrais Négatifs et plus de Faux Positifs par rapport à $\bar{S}^{t=1}$, et donc pour lesquelles l'ampleur du phénomène $\theta_{ks0}^t \ll \theta_{k00}^t$ est moins marquée. Dans le calcul de $w_{0i}^{t=2}$ pour ce voxel i , on a donc d'une part moins de facteurs de type θ_{ks0}^t en comparaison du cas précédent, et d'autre part ces facteurs sont de moins faible valeur. Cette situation conduit à $W_{0i}^{t=2} > W_{si}^{t=2}$, et le voxel i est alors assigné au fond dans $\bar{S}^{t=2}$.

Évolution aux itérations suivantes Nous avons montré que la segmentation consensus de la structure s s'élargissait à ses voxels voisins lors de l'étape d'espérance (E-step) de l'itération $t = 2$: $\bar{S}^{t=2}$ est plus large que $\bar{S}^{t=1}$. Cela a deux conséquences :

1. Lors de l'étape de maximisation (M-step) de l'itération $t = 2$: les segmentations S_k auront moins de Faux Positifs avec $\bar{S}^{t=2}$ ($\sum_{i:S_k(i) \neq 0} W_{0i}^{t=2}$) qu'elles en avaient avec $\bar{S}^{t=1}$ à l'itération précédente ($\sum_{i:S_k(i) \neq 0} W_{0i}^{t=1}$). Ceci va contribuer à des valeurs θ_{k00} plus élevées et à des valeurs θ_{ks0} plus faibles : en d'autres termes, le phénomène $\theta_{ks0} \ll \theta_{k00}$ va être amplifié par rapport à l'étape de maximisation de l'itération précédente ;
2. Lors de l'étape d'espérance (E-step) de l'itération $t = 3$: les voxels i en bordure externe de $\bar{S}^{t=2}$ correspondront à des valeurs $n_s^K(i)$ plus faibles que les $n_s^K(j)$ des voxels j en bordure externe de $\bar{S}^{t=1}$ à l'itération précédente. On aura donc moins de facteurs de type θ_{ks0} dans le calcul des w_{0i} correspondants.

Ces deux facteurs vont avoir des effets contraires. En effet, si l'on compare les valeurs $w_{0i}^{t=3}$ pour i en bordure externe de $\bar{S}^{t=2}$ aux valeurs $w_{0j}^{t=2}$ pour j en bordure externe de $\bar{S}^{t=1}$ à l'itération précédente : le premier facteur (effet $\theta_{ks0} \ll \theta_{k00}$ amplifié) aura tendance à faire chuter $w_{0i}^{t=3}$ par rapport à $w_{0j}^{t=2}$, tandis que le second ($n_s^K(i) < n_s^K(j)$) aura l'effet inverse. A l'itération $t = 3$, le second effet sera encore peu marqué à cause de la grande variabilité parmi les S_k , et la segmentation va donc continuer à grossir. Cependant, à un certain stade de l'algorithme, la tendance va s'inverser, et le second effet prendra le pas sur le premier : la segmentation consensus arrêtera alors de grossir et on aura atteint la convergence.

4.2.4.4 Explication des différents degrés de sur-segmentation selon les structures

Comme nous l'avons mis en évidence dans le paragraphe 4.2.3, la sur-segmentation n'est pas observée avec le même degré selon les différentes structures. Pour expliquer ces différences, nous allons montrer que le degré de sur-segmentation dépend principalement de deux facteurs, qui ne sont pas totalement indépendants :

- d'une part, le degré de variabilité observé parmi les segmentations S_k pour la structure s considérée,
- d'autre part, le ratio $\frac{K_s}{K}$ de segmentations de la base dans lesquelles la structure s considérée est présente.

En effet, si la variabilité est faible pour la structure s , alors dès la seconde itération (ou peu après) on aura un nombre $n_s^K(i)$ faible pour les voxels i en bordure de la segmentation consensus courante, et donc ceux-ci seront assignés au label 0, empêchant ainsi la sur-segmentation de se produire. De plus, si la structure s considérée est peu représentée dans la base (ratio $\frac{K_s}{K}$ faible), alors on aura de fait beaucoup moins de chances d'avoir un facteur de type θ_{ks0}^t dans le calcul de w_{0i} , et donc moins de chances de sous-estimer w_{0i} . On aura donc une sur-segmentation moins marquée. A noter que la variabilité parmi les segmentations pour la structure s n'est pas complètement indépendante du ratio $\frac{K_s}{K}$.

En résumé, les structures ayant une plus faible variabilité de forme ou de taille ainsi que les structures peu représentées parmi les S_k présenteront une sur-segmentation moins marquée que les autres, voire pas de sur-segmentation du tout. Cela est cohérent avec les résultats présentés en section 4.2.3. En effet, la seule structure qui n'est pas sur-segmentée du tout est la mandibule : c'est la structure la moins représentée dans la base ($\frac{K_s}{K} = \frac{11}{105}$), et elle présente une variabilité de forme assez faible en comparaison des autres. D'autre part, les aires ganglionnaires IV qui sont les plus sur-segmentées présentent effectivement une très grande variabilité anatomique.

4.2.5 Correction de la sur-segmentation par post-traitement

Une première solution pour s'affranchir de la sur-segmentation obtenue avec STAPLE consiste à corriger *a posteriori* la segmentation moyenne par des post-traitements de morphologie mathématique. Cette correction *a posteriori* consiste dans un premier temps à appliquer une érosion de taille optimale pour chaque structure indépendamment. Cette première étape nécessite d'évaluer sur la base de 105 patients les atlas moyens construits en érodant plus ou moins les structures moyennes, et de déterminer ainsi pour chaque structure la taille d'érosion optimale qui fournit des segmentations automatiques les plus proches possibles des segmentations manuelles en terme de volume. Les structures moyennes ainsi érodées ont l'inconvénient de ne plus être jointives alors que certaines d'entre elles sont censées l'être. Une solution pour les rendre jointives est de les modifier au voisinage des jonctions en utilisant un squelette par zone d'influence (ou SKIZ). Cette opération de morphologie mathématique consiste à propager des graines à l'intérieur d'un masque, en assignant successivement les voxels du masque non étiquetés à la graine la plus proche. Dans notre cas, les graines sont les structures moyennes érodées, et un masque adapté peut être obtenu en dilatant N fois les structures moyennes érodées puis en les érodant N fois, avec N suffisamment grand pour que le masque soit correct au niveau des frontières. Ces dilatations et érosions ont cependant l'inconvénient de trop lisser les structures moyennes. Une façon de s'affranchir de ce problème consiste à utiliser ces érosions et dilatations uniquement pour calculer le masque aux niveau des frontières, le masque pouvant être défini par les structures moyennes elles-mêmes en dehors des frontières.

4.2.6 Correction de la gestion des structures manquantes

La méthode décrite dans le paragraphe précédent permet de corriger la sur-segmentation obtenue avec l'algorithme STAPLE mais elle n'est pas satisfaisante d'un point de vue méthodologique (car le problème n'est corrigé que *a posteriori*), ni d'un point de vue pratique (car elle nécessite de calculer la taille d'érosion optimale pour toute nouvelle structure potentielle). De plus, la correction *a posteriori* suppose que la sur-segmentation pour une structure donnée a la même amplitude quelle que soit la zone de la structure considérée, ce qui n'est pas forcément vrai.

Nous proposons dans ce paragraphe de modifier la façon dont sont calculés les paramètres $\theta_{ks's}$. En effet, nous avons montré dans le paragraphe 4.2.4.2 que la présence de structures manquantes dans la segmentation S_k contribue à accentuer la sur-estimation du paramètre θ_{k00} , et donc indirectement la sur-segmentation.

Pour s'affranchir de ce problème, on associe à chaque structure $s \in [1 \dots L]$ un masque \mathcal{M}_s correspondant à l'union des K_s segmentations manuelles de la structure s . Pour chaque segmentation S_k , on définit alors un masque M_k comme l'union des masques \mathcal{M}_s pour l'ensemble des structures s segmentées dans S_k :

$$\forall k \in [1 \dots K] \quad M_k = \bigcup_{\substack{s \in [1 \dots L] \\ \text{segmentées dans } S_k}} \mathcal{M}_s \quad (4.10)$$

Nous proposons alors de calculer les paramètres $\theta_{ks's}^t$ associés à la segmentation S_k en se restreignant à son masque M_k , c'est-à-dire l'équation 4.5 est remplacée par l'équation suivante :

$$\forall s \in [0 \dots L] \quad \forall s' \in [0 \dots L] \quad \theta_{ks's}^t = \frac{\sum_{i \in M_k: S_k(i)=s'} W_{si}^t}{\sum_{i \in M_k} W_{si}^t} \quad (4.11)$$

Ainsi, les voxels assignés artificiellement à 0 dans S_k du fait des structures manquantes ne sont plus comptabilisés, et ne contribuent donc plus à augmenter θ_{k00}^t . Il peut cependant arriver que certains de ces voxels soient malgré tout comptabilisés dans le calcul de θ_{k00}^t , si par exemple une structure manquante est voisine d'une structure s segmentée dans S_k (car alors le masque \mathcal{M}_s déborde un peu sur les voxels correspondant à la structure manquante). Cet effet reste cependant limité à la bordure du masque M_k et concerne donc peu de voxels.

4.3 Algorithme de gestion des structures manquantes dans un calcul de segmentation probabiliste moyenne

4.3.1 Principe

Nous présentons dans ce paragraphe une approche simple permettant de gérer efficacement les structures manquantes dans un algorithme quelconque de calcul de segmentation probabiliste moyenne. Comme nous l'avons déjà mis en évidence, la difficulté apportée par les structures manquantes est liée au statut particulier du label 0 par rapport aux labels $s \in [1 \dots L]$, puisque le label 0 représente des entités différentes dans les différentes segmentations S_k .

Dans les algorithmes de calcul d'une segmentation probabiliste moyenne, l'objectif est en général de déterminer pour chaque voxel i sa probabilité d'appartenance W_{si} à chacun des labels $s \in [0 \dots L]$, avec $\sum_{s \in [0 \dots L]} W_{si} = 1$. En d'autres termes, tous les labels sont mis en compétition simultanément en chaque voxel, et le plus probable est choisi.

Nous considérons ici le calcul d'une segmentation probabiliste moyenne selon un angle différent. Au lieu de mettre simultanément en compétition l'ensemble des labels $s \in [0 \dots L]$ les uns contre les autres, nous proposons de mettre successivement en compétition chaque label $s \in [1 \dots L]$ avec l'union notée \bar{s} des labels s' restants ($s' \in [0 \dots L]$, $s' \neq s$). Ainsi, étant donné un voxel i , on ne cherche plus à estimer directement si un label est plus probable qu'un autre en ce voxel, mais on examine successivement si le voxel i a plus de chances d'appartenir (probabilité W_{si}) ou de ne pas appartenir (probabilité $W_{\bar{s}i} = 1 - W_{si}$) à chacun des labels $s \in [1 \dots L]$. Pour chaque label $s \in [1 \dots L]$, les probabilités W_{si} et $W_{\bar{s}i}$ peuvent se calculer facilement en se restreignant aux K_s segmentations dans lesquelles ce label est présent. Ensuite, la probabilité W_{0i} d'appartenir au fond de l'image est définie comme la probabilité de ne pas appartenir à la structure la plus probable au vu des $\{W_{si}\}_{s \in [1 \dots L]}$:

$$W_{0i} = 1 - \max_{s \in [1 \dots L]} W_{si} = \min_{s \in [1 \dots L]} W_{\bar{s}i} \quad (4.12)$$

Le voxel i est ensuite assigné au label $\bar{S}(i) = \arg \max_{s \in [0 \dots L]} W_{si}$. En d'autres termes, cela revient à assigner au voxel i le label 0 uniquement si aucune des structures $s \in [1 \dots L]$ n'a une probabilité W_{si} supérieure à 0.5.

Dans cette approche, les termes $\{W_{si}\}_{s \in [0 \dots L]}$ ne caractérisent pas des probabilités d'appartenance dans l'absolu (c'est-à-dire on a $\sum_{s=0}^L W_{si} \neq 1$), ils sont seulement définis dans un objectif de classification des voxels : les $\{W_{si}\}_{s \neq 0}$ désignent des probabilités d'appartenance à chaque structure s en opposition à l'ensemble des autres labels ; W_{0i} désigne la probabilité d'appartenance au fond en opposition à la structure la plus probable.

4.3.2 Mise en œuvre avec STAPLE

Nous présentons ici la mise en œuvre de l'approche proposée dans la section précédente dans le cas où les probabilités $\{W_{si}\}_{s \in [1 \dots L]}$ sont calculées avec STAPLE.

Au lieu d'être associée à $(L + 1)^2$ paramètres de qualité $\{\theta_{ks's'}\}_{(s,s') \in [0 \dots L]^2}$, chaque segmentation S_k est maintenant associée à $4L$ paramètres de qualité notés $\{\theta_{kss}, \theta_{k\bar{s}s}, \theta_{k\bar{s}\bar{s}}, \theta_{ks\bar{s}}\}_{s \in [1 \dots L]}$. A partir de ces paramètres de qualité, les probabilités $\{W_{si}^t\}_{s \in [0 \dots L]}$ à l'itération t sont définies comme suit :

$$\begin{cases} W_{si}^t = \frac{\pi_s \cdot \prod_{k=1}^K \theta_{k\delta_k(i,s)s}^{t-1}}{\pi_s \cdot \prod_{k=1}^K \theta_{k\delta_k(i,s)s}^{t-1} + (1-\pi_s) \cdot \prod_{k=1}^K \theta_{k\delta_k(i,s)\bar{s}}^{t-1}} & \forall s \in [1 \dots L] \\ W_{0i}^t = 1 - \max_{s \in [1 \dots L]} W_{si}^t \end{cases} \quad (4.13)$$

où les probabilités *a priori* π_s sont définies comme dans l'équation 4.6 et où :

$$\forall s \in [1 \dots L] \quad \delta_k(i, s) = \begin{cases} s & \text{si } S_k(i) = s \\ \bar{s} & \text{sinon} \end{cases} \quad (4.14)$$

Le calcul des paramètres de qualité à partir des W_{si} est alors décrit par les équations suivantes :

$$\begin{aligned} \theta_{kss}^t &= \frac{\sum_{i \in \mathcal{M}_s: S_k(i)=s} W_{si}^t}{\sum_{i \in \mathcal{M}_s} W_{si}^t} & \theta_{k\bar{s}s}^t &= \frac{\sum_{i \in \mathcal{M}_s: S_k(i)=s} 1 - W_{si}^t}{\sum_{i \in \mathcal{M}_s} 1 - W_{si}^t} \\ \theta_{k\bar{s}\bar{s}}^t &= \frac{\sum_{i \in \mathcal{M}_s: S_k(i) \neq s} W_{si}^t}{\sum_{i \in \mathcal{M}_s} W_{si}^t} & \theta_{ks\bar{s}}^t &= \frac{\sum_{i \in \mathcal{M}_s: S_k(i) \neq s} 1 - W_{si}^t}{\sum_{i \in \mathcal{M}_s} 1 - W_{si}^t} \end{aligned} \quad (4.15)$$

où chaque masque \mathcal{M}_s est défini par l'union des K_s segmentations manuelles de la structure s , comme précisé dans la section 4.2.6.

Cette mise en œuvre de l'algorithme STAPLE permet de s'affranchir des inconvénients présents dans la version multilabel classique et mentionnés en section 4.2.4. Notamment, le fait d'utiliser des masques de calcul \mathcal{M}_s spécifiques à chaque structure $s \in [1 \dots L]$ pour calculer les paramètres $\{\theta_{kss}, \theta_{k\bar{s}s}, \theta_{k\bar{s}\bar{s}}, \theta_{ks\bar{s}}\}$ permet de réduire la taille du fond dans le masque de calcul. En effet, seuls les Vrais Négatifs au voisinage de la structure s considérée seront alors comptabilisés. Cela permet par conséquent de s'affranchir du déséquilibre $\theta_{k00} \gg \theta_{ks0}$ qui est à l'origine de la sur-segmentation obtenue avec la version multilabel classique de STAPLE. La mise en œuvre proposée dans ce paragraphe correspond en d'autres termes à une version régionale de l'algorithme STAPLE où chaque région correspond à une structure anatomique $s \in [1 \dots L]$.

4.3.3 Mise en œuvre avec un vote par majorité

L'approche de gestion des structures manquantes proposée en section 4.3.1 peut s'appliquer à tout algorithme de calcul de segmentation moyenne probabiliste. Nous l'appliquons ici au cas où les probabilités $\{W_{si}\}_{s \in [1 \dots L]}$ sont calculées selon le principe du vote par majorité. Les équations sont les suivantes :

$$\begin{cases} W_{si} = \frac{1}{K_s} \sum_{\substack{k \in [1 \dots K] \\ \text{ayant segmenté } s}} \delta(S_k(i), s) & \forall s \in [1 \dots L] \\ W_{0i} = 1 - \max_{s \in [1 \dots L]} W_{si} \end{cases} \quad (4.16)$$

4.4 Calcul d'une segmentation moyenne à partir du degré d'accord local

4.4.1 Principe

Les algorithmes présentés précédemment permettent de prendre en compte les structures manquantes directement lors du calcul de la segmentation probabiliste moyenne. Nous proposons ici d'estimer la segmentation moyenne par une approche différente basée sur l'estimation locale du degré d'accord entre les segmentations multilabels données en entrée. Cette approche repose sur l'hypothèse selon laquelle les frontières entre deux labels de la segmentation moyenne correspondent aux zones de plus grand désaccord entre les segmentations données en entrée. Inversement, l'intérieur des structures correspond aux zones de faible désaccord. L'approche est donc basée sur les deux étapes successives suivantes :

1. estimation locale du degré d'accord entre les segmentations multilabels,
2. estimation de la segmentation moyenne à partir de la carte d'accord local.

Ce schéma représente une alternative intéressante pour aborder le problème des structures manquantes dans les segmentations multilabels données en entrée. En effet, la gestion des structures manquantes se fait ici lors de l'estimation du degré d'accord local et non plus directement lors du calcul de la segmentation moyenne. L'algorithme proposé peut se classer parmi les approches basées sur la forme.

4.4.2 Algorithme global

Pour plus de clarté, nous présentons dans un premier temps la méthode d'estimation de l'accord local entre les segmentations dans le cas sans structures manquantes (paragraphe 4.4.2.1) puis sa généralisation au cas de structures manquantes (paragraphe 4.4.2.2). Enfin, le paragraphe 4.4.2.3 présente l'étape d'estimation de la segmentation moyenne à partir de la carte d'accord local.

4.4.2.1 Estimation du degré d'accord local entre plusieurs segmentations multilabels complètes

Chaque voxel i est associé à un ensemble de K votes $\{S_k(i)\}_{k \in [1 \dots K]}$. Pour tout label $s \in [0 \dots L]$, on note $n_s^K(i) = \sum_{k=1}^K \delta(S_k(i), s)$ le nombre de segmentations parmi les K ayant attribué le voxel i au label s (δ représentant le symbole de Kronecker).

Le calcul du degré d'accord local au voxel i (noté $\omega(i)$) est basé sur la façon dont se répartissent les votes parmi les $L + 1$ labels possibles. Intuitivement, l'accord sera fort si un label est nettement majoritaire parmi les K votes, tandis qu'il sera faible si deux labels se partagent la première place avec des nombres de votes similaires. Nous proposons donc de définir l'accord local au voxel i de la façon suivante :

$$\omega(i) = \frac{n_{s_1(i)}^K(i) - n_{s_2(i)}^K(i)}{K} \quad (4.17)$$

où $s_1(i)$ et $s_2(i)$ désignent respectivement le premier et le deuxième label majoritaire parmi les K votes :

$$\left\{ \begin{array}{l} s_1(i) = \arg \max_{s \in [0 \dots L]} n_s^K(i), \\ s_2(i) = \arg \max_{\substack{s \in [0 \dots L] \\ s \neq s_1(i)}} n_s^K(i) \end{array} \right. \quad (4.18)$$

Ainsi, le degré d'accord sera toujours compris entre 0 et 1. Il sera minimal pour $n_{s_1(i)}^K = n_{s_2(i)}^K$ et maximal pour $n_{s_1(i)}^K = K$ et $n_{s_2(i)}^K = 0$. Pour une même valeur de $n_{s_1(i)}^K$, il sera d'autant plus fort que les votes restants sont équi-répartis parmi les labels restants.

4.4.2.2 Estimation du degré d'accord local entre plusieurs segmentations multilabels incomplètes

Lorsqu'il y a des structures manquantes parmi les segmentations en entrée, on ne peut plus simplement se baser sur le nombre de votes en absolu $n_s^K(i)$ attribués à chaque label s au voxel i , et l'algorithme précédent n'est donc plus valide. Nous proposons de le généraliser au cas de structures manquantes selon les deux étapes suivantes. Pour chaque voxel i :

1. Identifier quels seraient les deux labels majoritaires au voxel i « s'il n'y avait pas de structures manquantes » ; nous qualifions ces deux labels de labels « principaux » et non plus « majoritaires » (car ils ne sont justement pas forcément majoritaires en absolu parmi les K votes) ;
2. Calculer le degré d'accord local au voxel i par une formule similaire à 4.17 mais en se restreignant aux segmentations dans lesquelles les deux labels principaux ont été segmentés.

Identification des deux labels « principaux » Nous définissons les deux labels « principaux » $l_1(i)$ et $l_2(i)$ au voxel i comme suit :

$$\left\{ \begin{array}{ll} l_1(i) = 0 & \text{si } \forall s \in [1 \dots L] \ n_s^K(i) = 0 \\ & = \arg \max_{s \in [1 \dots L]} \frac{n_s^K(i)}{K_s} & \text{sinon} \\ \\ l_2(i) = 0 & \text{si } \forall s \in [1 \dots L], s \neq l_1(i) \ n_s^{K_{l_1(i)}}(i) = 0 \\ & = \arg \max_{\substack{s \in [1 \dots L] \\ s \neq l_1(i)}} \frac{n_s^{K_{l_1(i)}}(i)}{K_{l_1(i)}s} & \text{sinon} \end{array} \right. \quad (4.19)$$

où :

- K_s désigne le nombre de segmentations dans lesquelles le label s est présent,
- $K_{l_1(i)}$ désigne le nombre de segmentations dans lesquelles le label $l_1(i)$ est présent,
- $K_{l_1(i)s}$ désigne le nombre de segmentations dans lesquelles les labels $l_1(i)$ et s sont tous les deux présents,
- $n_s^{K_{l_1(i)}}(i) = \sum_{\substack{k \in [1 \dots K] \\ \text{ayant segmenté } l_1(i)}} \delta(S_k(i), s)$.

Le premier label principal $l_1(i)$ est défini comme le label s maximisant le ratio entre le nombre de votes $n_s^K(i)$ pour ce label au voxel i et le nombre de segmentations K_s dans lesquelles ce label est présent. En effet, comparer les nombres de votes $n_s^K(i)$ en absolu ne permet pas de prendre en compte les structures manquantes. Par exemple, un voxel i situé à l'intérieur d'une structure anatomique s peu segmentée dans la base aura vraisemblablement un plus grand nombre de votes pour le label 0 que pour le label s . Pour autant, c'est bien le label « principal » (label s) et non le label « majoritaire » (label 0) que l'on veut déterminer.

D'autre part, la détermination du premier label principal doit également prendre en compte le statut particulier du label $s = 0$. En effet, comme les structures manquantes ne sont pas les mêmes d'une segmentation à l'autre, le label $s = 0$ ne représente pas la même entité dans toutes les segmentations. Il n'y a donc *a priori* pas de sens de définir la variable K_0 . Pour cette raison, nous recherchons $l_1(i)$ parmi les labels $s \in [1 \dots L]$, et nous assignons $l_1(i) = 0$ seulement dans le cas où les K votes ont désigné le label 0 au voxel i .

Le second label principal $l_2(i)$ est défini de façon analogue au premier label principal à une différence près. La différence est que l'on ne considère plus l'ensemble des K votes mais seulement les $K_{l_1(i)}$ segmentations dans lesquelles le label $l_1(i)$ est présent. La justification de ce choix est la suivante : étant donné le premier label principal $l_1(i)$ identifié, ce qui nous intéresse est de déterminer avec quel autre label les $K_{l_1(i)}$ experts ayant segmenté le label $l_1(i)$ ont hésité pour classifier le voxel i .

Estimation du degré d'accord local À l'issue de l'étape précédente, les deux labels principaux $l_1(i)$ et $l_2(i)$ ont été identifiés. Le cas $l_1(i) = l_2(i) = 0$ correspond

au cas où l'ensemble des votes ont désigné le label 0 pour le voxel i (accord parfait). Dans les autres cas, $l_1(i) \in [1 \dots L]$ et $l_2(i) \in [0 \dots L]$.

Afin de calculer le degré d'accord local au voxel i , l'idée est de se ramener localement au cas où aucune structure n'est manquante, et d'appliquer alors l'algorithme sans structures manquantes présenté dans la section 4.4.2.1. Nous allons maintenant montrer que selon certaines hypothèses, une façon de se ramener au cas sans structures manquantes est de se restreindre aux $K_{l_1(i)l_2(i)}$ segmentations dans lesquelles les labels $l_1(i)$ et $l_2(i)$ sont tous les deux présents.

L'hypothèse que nous faisons concerne la position relative des structures anatomiques considérées. Nous supposons ici que tout voxel est au maximum en contact avec deux structures anatomiques $s \in [1 \dots L]$. Cette hypothèse est correcte pour la configuration spatiale des structures ORL considérées. Cela signifie que lors de la classification d'un voxel, les experts ont hésité au maximum entre deux structures anatomiques $s \in [1 \dots L]$ et le label $s = 0$. La figure 4.4 illustre cette hypothèse.

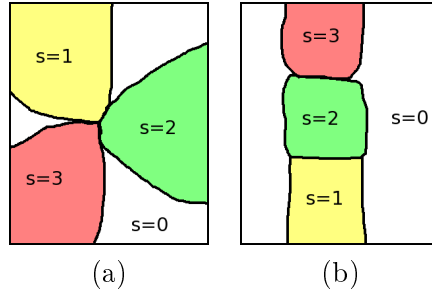


FIG. 4.4 – Différentes configurations spatiales possibles. (a) : configuration spatiale non gérée par notre méthode (mais absente pour les structures de la sphère ORL considérées) ; (b) : configuration spatiale correctement gérée par notre méthode.

Étant donnée cette hypothèse et les labels principaux $l_1(i)$ et $l_2(i)$ identifiés au voxel i , il en découle que les deux labels « majoritaires » en terme de nombre de votes en absolu parmi les $K_{l_1(i)l_2(i)}$ segmentations considérées sont forcément parmi les labels $l_1(i)$, $l_2(i)$ et 0. Le label 0 représente bien localement la même entité dans les $K_{l_1(i)l_2(i)}$ segmentations considérées, cette entité étant le complémentaire de $l_1(i) \cup l_2(i)$. On peut donc appliquer les équations 4.18 et 4.17 aux $K_{l_1(i)l_2(i)}$ segmentations. En notant $n_s^{K_{l_1(i)l_2(i)}}(i) = \sum_{\substack{k \in [1 \dots K] \\ \text{ayant segmenté } l_1(i) \text{ et } l_2(i)}} \delta(S_k(i), s)$, on a :

$$\begin{cases} s_1(i) = \arg \max_{s \in [0 \dots L]} n_s^{K_{l_1(i)l_2(i)}}(i), \\ s_2(i) = \arg \max_{\substack{s \in [0 \dots L] \\ s \neq s_1(i)}} n_s^{K_{l_1(i)l_2(i)}}(i) \end{cases} \quad (4.20)$$

$$\omega(i) = \frac{n_{s_1(i)}^{K_{l_1(i)l_2(i)}}(i) - n_{s_2(i)}^{K_{l_1(i)l_2(i)}}(i)}{K_{l_1(i)l_2(i)}} \quad (4.21)$$

4.4.2.3 Estimation de la segmentation moyenne à partir de la carte de degré d'accord local

A ce stade, on a estimé en chaque voxel i le degré d'accord local $\omega(i)$ entre les segmentations. La figure 4.5 montre la carte d'accord local obtenue pour différentes coupes. Les maxima du degré d'accord correspondent à l'intérieur des structures et au fond de l'image, tandis que les minima correspondent aux frontières entre les différents labels.

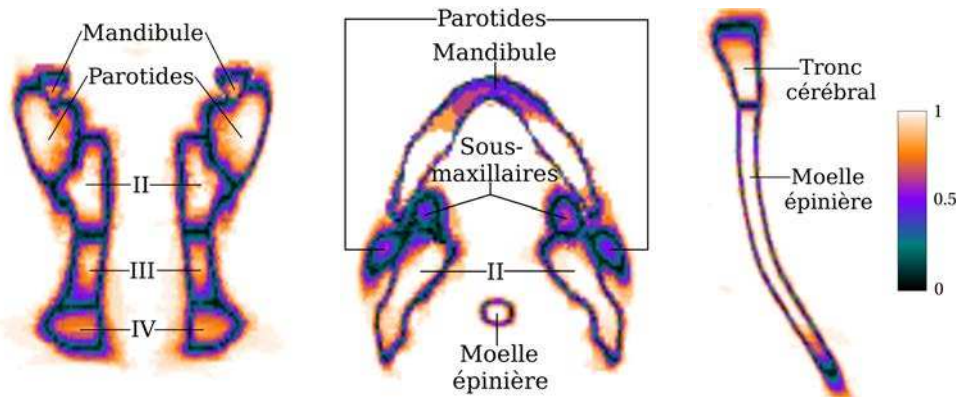


FIG. 4.5 – Carte d'accord local obtenue pour différentes coupes.

Nous proposons d'utiliser l'algorithme de ligne de partage des eaux [Vincent 1991] pour déduire la segmentation moyenne à partir de la carte du degré d'accord local. Cet algorithme de détection de contours utilise des opérations morphologiques pour propager au sein de l'image des graines initialement fournies par l'utilisateur, et ce jusqu'à détection des lignes de crête entre les différents labels. En général, il est appliqué à l'image de gradient de façon à ce que les contours recherchés correspondent aux maxima d'intensité. Dans notre cas, les contours recherchés correspondent aux minima d'intensité, nous allons donc appliquer l'algorithme sur l'inverse de la carte d'accord (c'est-à-dire la carte des $(1 - \omega(i))$ puisque les $\omega(i)$ sont toujours compris entre 0 et 1). Alternativement, il est aussi possible d'inverser une version lissée de la carte d'accord (obtenue par exemple par filtrage gaussien, ou encore en moyennant pour chaque voxel i les degrés d'accord obtenus sur un voisinage du voxel i).

Plusieurs solutions sont possibles quant à la définition des graines. La première consiste à extraire les maxima locaux d'intensité sur la carte des $\omega(i)$ vu que ces maxima locaux correspondent aux intérieurs des labels. Une autre solution consiste à définir la graine de chaque label comme l'intersection des segmentations fournies en entrée pour ce label. Nous avons testé les deux solutions et obtenu des résultats similaires. La figure 4.6 montre la segmentation moyenne ainsi obtenue superposée à la carte du degré d'accord local.

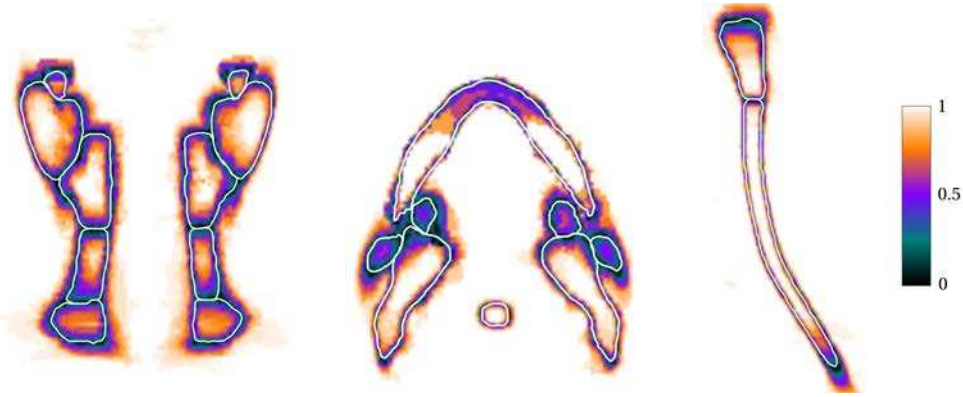


FIG. 4.6 – Visualisation de la segmentation moyenne obtenue (contours blancs) superposée à la carte d'accord local pour différentes coupes.

4.5 Évaluation

Nous présentons dans cette partie une évaluation détaillée des différentes méthodes citées et proposées dans les sections précédentes 4.2, 4.3 et 4.4. Nous utilisons les mesures d'évaluation DICE, SENS, SPEC, DIST, VOL et RHD_{τ} présentées dans l'annexe B.

4.5.1 Correction *a posteriori* de la sur-segmentation dans l'atlas

Comme détaillé dans la partie 4.2.5, nous avons dans un premier temps cherché à déterminer l'érosion optimale à appliquer à chaque structure anatomique moyenne de l'atlas. Nous avons évalué des érosions avec des boules euclidiennes de rayons respectifs 1, 2 et 3 voxels. En pratique, on peut montrer que pour ces valeurs de rayons, les érosions correspondantes peuvent être réalisées en première approximation par des successions adéquates d'érosions utilisant des éléments structurants de connexité 6, 18, ou 26. Par exemple, l'érosion de rayon 1 voxel correspond à une érosion avec un élément structurant de connexité 18, celle de rayon 2 voxels correspond à une érosion avec un élément structurant de connexité 26 suivie d'une érosion avec un élément structurant de connexité 6.

Afin d'évaluer l'effet des différentes tailles d'érosion pour chaque structure, nous avons utilisé les mesures VOL, DICE, DIST et $RHD_{90\%}$. Lorsque les résultats quantitatifs ne permettaient pas de trancher clairement entre deux tailles d'érosion, nous avons choisi l'érosion de plus petite taille. Les érosions optimales ainsi déterminées sont les suivantes :

- mandibule, moelle épinière, aires ganglionnaires II : pas d'érosion ;
- parotides, aires ganglionnaires III et IV, sous-maxillaires, tronc cérébral : érosion de rayon euclidien 1 voxel.

La figure 4.7 permet de comparer visuellement les contours moyens obtenus avec STAPLE sans aucune érosion (contours rouges) aux contours moyens après érosion

optimale pour chaque structure et post-traitement au niveau des jonctions entre structures (contours verts).

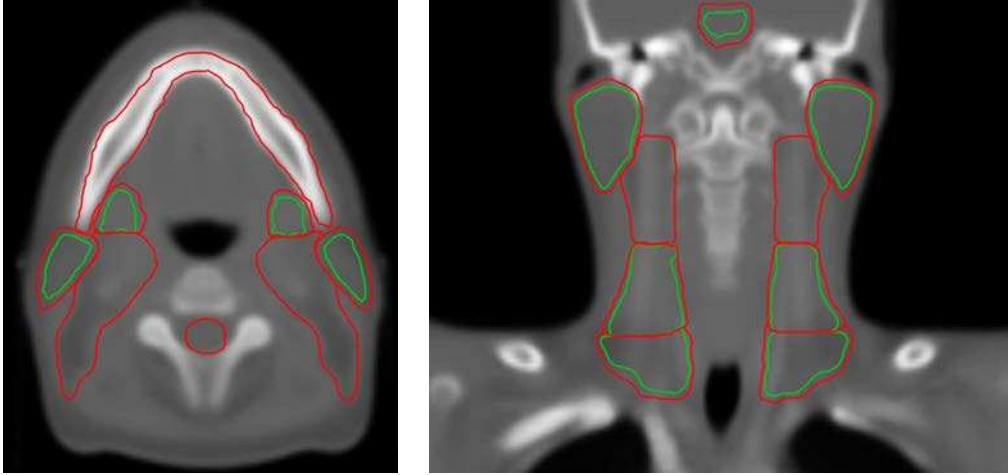


FIG. 4.7 – Comparaison qualitative entre les contours moyens obtenus avec STAPLE (en rouge) et ces mêmes contours après l'érosion de chaque structure selon la taille optimale indiquée par les mesures VOL, DICE, DIST et $RHD_{90\%}$ (en vert) et un post-traitement au niveau des jonctions.

Par ailleurs, une évaluation qualitative des différentes tailles d'érosion a été menée par un médecin radiothérapeute. Cette évaluation qualitative préconisait des tailles d'érosion plus grandes que celles déterminées quantitativement grâce aux mesures VOL, DICE, DIST et $RHD_{90\%}$: érosion de rayon euclidien de 2 voxels pour les parotides et l'ensemble des aires ganglionnaires, de 1 voxel pour la moelle, le tronc cérébral et les sous-maxillaires, et pas d'érosion pour la mandibule.

Ainsi, les mesures quantitatives ne semblent pas refléter fidèlement les appréciations qualitatives fournies par le radiothérapeute. Ce décalage met en évidence la difficulté d'évaluation des segmentations automatiques pour les structures de la sphère ORL. Nous avons choisi dans cette évaluation de considérer malgré tout les tailles d'érosions déterminées quantitativement selon les mesures VOL, DICE, DIST et $RHD_{90\%}$ plutôt que les tailles d'érosions préconisées qualitativement par le radiothérapeute. Cependant, lors de la comparaison entre ces contours érodés et les contours moyens obtenus selon diverses méthodes, nous gardons à l'esprit que cette version érodée est peut être encore légèrement sur-segmentée selon l'appréciation visuelle du médecin.

4.5.2 Algorithmes comparés

4.5.2.1 Notations

Dans cette section, nous comparons les segmentations moyennes obtenues avec l'algorithme STAPLE ainsi qu'avec les différentes méthodes proposées dans les sec-

tions 4.2, 4.3 et 4.4. Nous comparons aussi nos résultats à un algorithme basique présenté en section 4.5.2.2. Le tableau 4.8 définit les notations utilisées dans la suite pour se référer à chacun de ces algorithmes.

STAPLE_STD	Algorithme STAPLE standard, avec gestion des structures manquantes comme détaillé dans la section 4.2.2	4.2.2
STAPLE_STD_EROSION	STAPLE_STD suivi d'une érosion optimale pour chaque structure et d'un post-traitement pour garder les structures jointives	4.2.5
STAPLE_STD_COR	Algorithme STAPLE standard avec gestion des structures manquantes comme détaillé dans la section 4.2.6	4.2.6
STAPLE_REG	Approche de gestion des structures manquantes appliquée à l'algorithme STAPLE (l'abréviation REG faisant écho à l'application régionale de STAPLE, structure par structure)	4.3.2
MAJVOTE_REG	Approche de gestion des structures manquantes appliquée au vote par majorité	4.3.3
ACC_LOC	Algorithme estimant la segmentation moyenne à travers le degré d'accord local	4.4
MAJVOTE_SMOOTH_MONO	Algorithme de vote par majorité lissé pour chaque structure suivi d'un post-traitement pour corriger les chevauchements	4.5.2.2

FIG. 4.8 – Notations utilisées pour les différentes méthodes de calcul de segmentation moyenne comparées.

4.5.2.2 Algorithme de vote par majorité lissé

Nous présentons ici un algorithme permettant de mettre en œuvre structure par structure une version lissée du vote par majorité, puis de combiner les segmentations obtenues en gérant leurs éventuels chevauchements à l'aide d'opérations de morphologie mathématique de base. Les étapes de cet algorithme sont les suivantes :

1. Pour chaque structure anatomique $s \in [1 \dots L]$:
 - Calculer la carte de probabilité à partir des K_s segmentations dans lesquelles le label s est présent (moyenne voxel par voxel)
 - Appliquer un filtrage gaussien à la carte de probabilité
 - Seuiller la carte de probabilité lissée pour obtenir une segmentation binaire moyenne $T(s)$ de la structure s
2. A chaque jonction entre deux structures anatomiques s et s' voisines :
 - Calculer $T_{\text{disj}}(s) = T(s) \setminus (T(s) \cap T(s'))$
 - Calculer $T_{\text{disj}}(s') = T(s') \setminus (T(s) \cap T(s'))$
 - Effectuer un SKIZ (squelette par zone d'influence) de $T_{\text{disj}}(s)$ et $T_{\text{disj}}(s')$ à l'intérieur de $T(s) \cup T(s')$

Cet algorithme dépend principalement de deux paramètres : l'écart-type du filtre gaussien et la valeur utilisée pour seuiller la carte de probabilité lissée. Pour ces deux

paramètres, nous avons choisi parmi un ensemble de valeurs raisonnables le couple de valeurs optimal au sens des mesures quantitatives DICE, DIST et VOL sur les 105 patients de notre base. Les valeurs retenues et utilisées dans ce chapitre sont un écart-type de 1 pour le filtrage gaussien, et un seuil de 0.4.

4.5.3 Évaluation quantitative

Dans les sections 4.2.5 et 4.5.1 nous avons proposé et mis en œuvre une méthode de correction *a posteriori* de la sur-segmentation obtenue avec STAPLE (STAPLE_STD_ERO). Nous avons ensuite proposé plusieurs méthodes permettant soit de s'affranchir *a priori* des problèmes à l'origine de la sur-segmentation (c'est le cas des méthodes STAPLE_STD_COR, STAPLE_REG et MAJVOTE_REG), soit d'utiliser une approche alternative pour calculer la segmentation moyenne en se basant sur l'estimation préalable du degré d'accord local (ACC_LOC). Pour l'évaluation quantitative, nous procédons en deux temps.

4.5.3.1 Évaluation en terme de volume

Tout d'abord, nous voulons comparer les méthodes proposées pour corriger *a priori* la sur-segmentation à la méthode de correction *a posteriori*. Pour cela, nous utilisons la mesure d'évaluation VOL car elle permet de quantifier facilement la sur- ou sous-segmentation. La figure 4.9 présente pour chaque structure les moyennes obtenues sur l'ensemble des 105 patients avec les différentes méthodes. Dans cette étude, les structures symétriques (parotides, sous-maxillaires, aires ganglionnaires II, III et IV) ont été groupées dans une même catégorie. Les moyennes obtenues sur l'ensemble des structures sont également présentées (Global).

Tout d'abord, nous constatons que la correction *a posteriori* de la sur-segmentation (barres grises, STAPLE_STD_ERO) permet de réduire la sur-segmentation par rapport à l'algorithme STAPLE_STD (barres noires) mais elle ne permet pas de s'en affranchir complètement : la moelle épinière ainsi que les aires ganglionnaires II et IV restent sur-segmentées. Pour ces structures, nous nous retrouvons typiquement dans le cas où l'érosion de taille choisie n'était pas suffisante, mais l'érosion de taille supérieure était trop importante au vu des mesures quantitatives. Comme précisé dans le paragraphe 4.5.1, nous avons alors choisi l'érosion la plus petite.

D'autre part, la correction apportée à la gestion des structures manquantes dans l'algorithme STAPLE (barres blanches, STAPLE_STD_COR) permet effectivement de réduire la sur-segmentation par rapport à l'algorithme STAPLE_STD. Cela est vrai pour toutes les structures, mais l'amélioration reste marginale en comparaison de la correction *a posteriori* STAPLE_STD_ERO. Cela signifie que la sur-segmentation dans STAPLE_STD n'est imputable aux structures manquantes que dans une très faible proportion. D'après notre analyse présentée en section 4.2.4, la sur-segmentation viendrait donc essentiellement de la grande taille du masque de calcul. Selon cette hypothèse, les résultats devraient donc être meilleurs avec la méthode

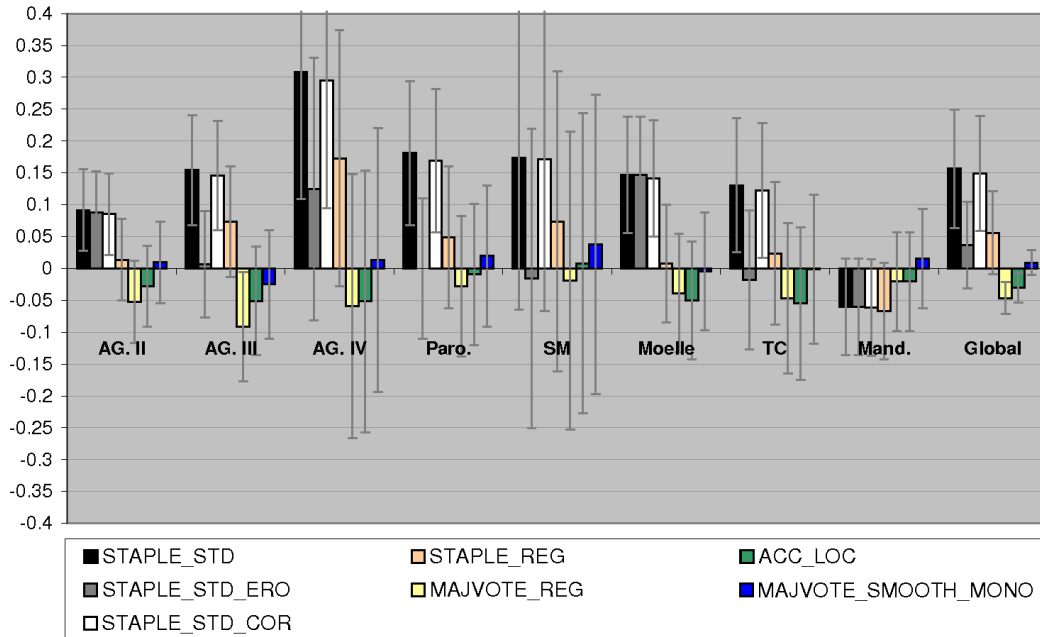


FIG. 4.9 – Valeurs moyennes obtenues pour la mesure VOL (log-ratio entre les volumes automatique et manuel) pour chaque structure sur l'ensemble des 105 patients. Les notations utilisées pour les méthodes sont celles du tableau 4.8.

STAPLE_REG : en effet, avec cette méthode il n'y a plus un seul et unique grand masque de calcul commun à l'ensemble des structures, mais un masque de calcul par structure. Les Vrais Négatifs dans chaque masque sont donc considérablement réduits, et la sur-segmentation devrait l'être également. Cela est confirmé par les résultats quantitatifs (barres oranges). La méthode STAPLE_REG a l'avantage de réduire la sur-segmentation mieux que la correction *a posteriori* (STAPLE_STD_ERO) sur les structures pour lesquelles l'érosion optimale se situait à cheval entre deux tailles (par exemple les aires ganglionnaires II et la moelle épinière). Pour les autres structures en revanche, la correction *a posteriori* donne de meilleurs résultats (aires ganglionnaires III, parotides et sous-maxillaires).

Concernant la méthode ACC_LOC (barres vertes), les résultats suggèrent une légère sous-segmentation des volumes. L'amplitude de cette sous-segmentation est cependant bien inférieure à celle de la sur-segmentation observée avec l'algorithme STAPLE_STD pour toutes les structures. En moyenne sur toutes les structures (Global), l'amplitude de sous-segmentation avec ACC_LOC est également inférieure à celle de l'amplitude de sur-segmentation avec STAPLE_REG. Seuls le tronc cérébral et la moelle épinière présentent la tendance inverse, c'est-à-dire sont davantage sous-segmentés par ACC_LOC que sur-segmentés par STAPLE_REG. En moyenne sur toutes les structures, la méthode ACC_LOC semble donc fournir des volumes légèrement sous-segmentés, mais de taille plus proche de la segmentation

manuelle que ceux obtenus avec les différentes variantes de STAPLE (STAPLE_STD, STAPLE_STD_ERO, STAPLE_REG). De plus, il faut garder à l'esprit que les érosions préconisées qualitativement par le médecin correspondaient à des volumes légèrement sous-segmentés selon la mesure VOL, ce qui représente un autre argument en faveur de la méthode ACC_LOC.

D'autre part, la méthode ACC_LOC permet de réduire la sous-segmentation de la mandibule constatée avec les méthodes précédemment citées. Elle apparaît donc comme un bon compromis, permettant à la fois de réduire la sous-segmentation de la mandibule, tout en réduisant également la sur-segmentation des autres structures. La méthode MAJVOTE_REG (barres jaunes) présente une tendance similaire mais avec une sous-segmentation plus prononcée que ACC_LOC.

Enfin, bien que basique sur le plan méthodologique, la méthode MAJVOTE_SMOOTH_MONO (barres bleues) est celle qui fournit des volumes automatiques les plus proches en taille des volumes manuels. Cela est vrai pour toutes les structures sauf pour les parotides et les sous-maxillaires pour lesquelles les meilleurs résultats sont obtenus avec la méthode ACC_LOC.

Du point de vue clinique, il est difficile de dire s'il est préférable d'avoir une sur- ou une sous-segmentation car les volumes considérés sont ici de deux types : les aires ganglionnaires sont des volumes cibles, alors que les autres structures sont des organes à risque. Pour les organes à risque, une sur-segmentation semblerait préférable (une sous-segmentation entraînerait une irradiation non souhaitée de ces organes), mais pour les volumes cibles, c'est l'inverse (une sur-segmentation pourrait engendrer l'irradiation non souhaitée de tissus ou d'organes sensibles). Quoi qu'il en soit, qu'elles soient trop larges (sur-segmentation) ou trop petites (sous-segmentation), les segmentations automatiques sont en pratique clinique contrôlées et corrigées par le radiothérapeute. Ce qui importe est donc l'amplitude de la sur- ou de la sous-segmentation, qui doit être aussi faible que possible de façon à être acceptable d'un point de vue clinique.

4.5.3.2 Évaluation en terme de précision

Nous évaluons maintenant les contours automatiques en terme de précision de segmentation. Pour cela, nous utilisons les mesures DICE, DIST et $RHD_{\tau=90\%}$. Les résultats obtenus sur les 105 patients pour les aires ganglionnaires IV, la mandibule et en moyenne sur toutes les structures (Global) sont présentées dans la figure 4.10. Nous présentons les résultats sur un nombre restreint de structures pour ne pas surcharger, et nous avons choisi ces structures car ce sont celles pour lesquelles la tendance à la sur-segmentation (aires IV) ou à la sous-segmentation (mandibule) est la plus marquée.

Tout d'abord, l'amélioration marginale fournie par la méthode STAPLE_STD_COR (barres blanches) par rapport à STAPLE_STD (barres noires) est confirmée. L'amélioration significative fournie par la méthode STAPLE_REG (barres oranges) par rapport à STAPLE_STD est également confirmée en moyenne sur toutes les structures.

Les différences entre les méthodes STAPLE_REG, MAJVOTE_REG, ACC_LOC et MAJVOTE_SMOOTH_MONO (barres oranges, jaunes, vertes, bleues) ne sont pas significatives en moyenne sur toutes les structures pour les mesures DICE et DIST. Elles le sont en revanche pour $RHD_{\tau=90\%}$, où STAPLE_REG fournit de meilleurs résultats que les trois autres.

Conformément aux conclusions tirées pour la mesure VOL, les méthodes dérivées de STAPLE (c'est-à-dire STAPLE_STD, STAPLE_STD_ERO, STAPLE_STD_COR, STAPLE_REG) fournissent des résultats moins bons pour la mandibule en comparaison des autres méthodes, que ce soit pour DICE, DIST ou $RHD_{\tau=90\%}$. Comme mis en évidence en section 4.2.4.4, cela est dû au fait que la mandibule est très peu représentée dans la base (11 segmentations seulement sur les 105 patients).

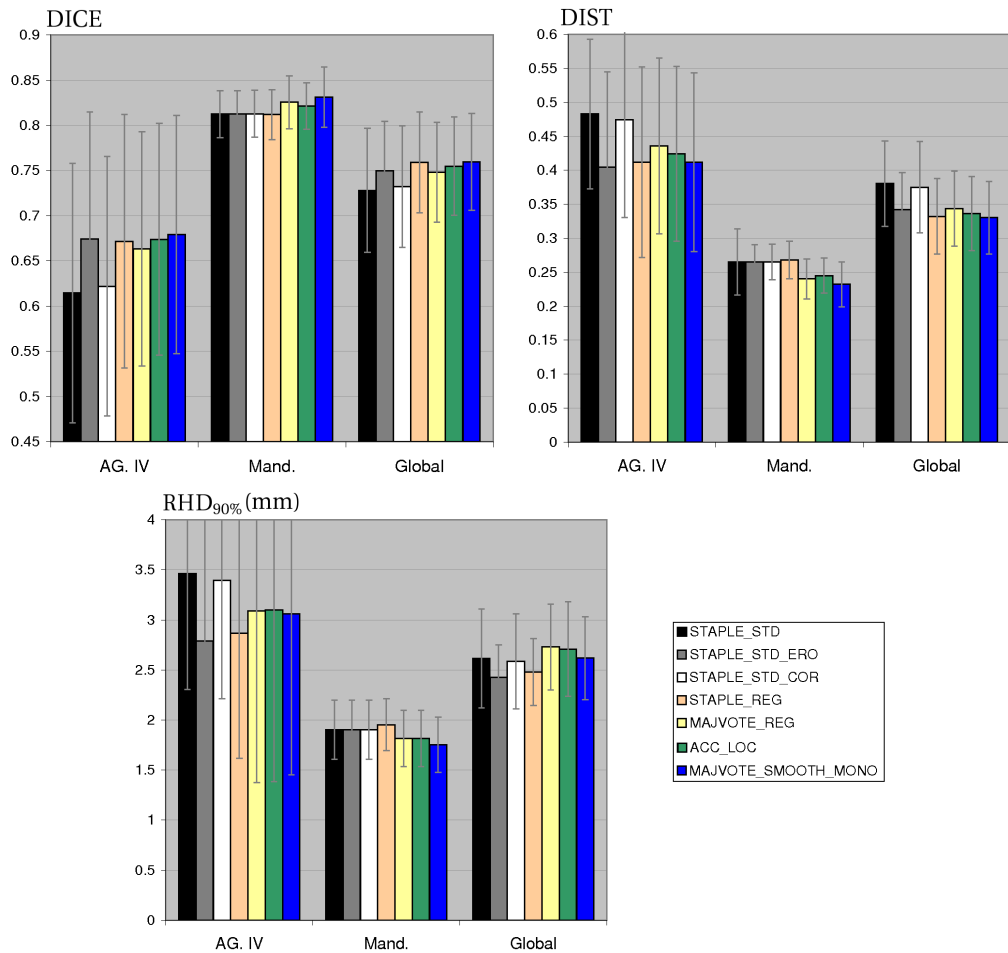


FIG. 4.10 – Valeurs moyennes obtenues pour les mesures DICE, DIST et $RHD_{\tau=90\%}$ pour les aires ganglionnaires IV, la mandibule et en moyenne sur toutes les structures (Global) sur l'ensemble des 105 patients. Les notations utilisées pour les méthodes sont celles du tableau 4.8.

4.5.4 Évaluation qualitative

Nous présentons dans cette partie des résultats visuels obtenus avec les différentes méthodes de calcul de segmentation moyenne.

4.5.4.1 Segmentations moyennes

Tout d’abord, la figure 4.11 montre des coupes de l’image moyenne de l’atlas sur laquelle nous avons représenté les contours moyens obtenus avec les méthodes STAPLE_STD (en rouge), STAPLE_REG (en orange), ACC_LOC (en jaune) et MAJVOTE_SMOOTH_MONO (en bleu).

Les tendances en terme de sur-segmentation mentionnées lors de l’évaluation quantitative se vérifient visuellement. STAPLE_STD présente les contours les plus larges, suivi de STAPLE_REG puis de ACC_LOC et MAJVOTE_SMOOTH_MONO qui sont tous deux très proches. De façon cohérente avec les résultats de la mesure VOL présentés en figure 4.9, ces différences sont les plus marquées pour les aires ganglionnaires IV.

4.5.4.2 Segmentation automatique pour un patient

La figure 4.12 compare la segmentation manuelle d’un patient particulier (en bleu clair) aux segmentations automatiques obtenues avec les atlas construits selon les méthodes STAPLE_STD, STAPLE_STD_ERO, STAPLE_REG, ACC_LOC et MAJVOTE_SMOOTH_MONO. Les conclusions en terme de largeur de segmentation des différentes méthodes sont confirmées sur ce patient. Comme pour la figure 4.11, cela est le plus visible sur les aires ganglionnaires IV (structures les plus basses sur la coupe de la troisième colonne). Pour ces structures, les atlas construits avec STAPLE_STD, STAPLE_STD_ERO, STAPLE_REG fournissent une segmentation automatique trop large, tandis que MAJVOTE_SMOOTH_MONO et en particulier ACC_LOC produisent des segmentations automatiques de taille correcte. La conclusion est la même pour les sous-maxillaires (structures sphériques dans la coupe de la colonne du milieu). On remarque que certaines spécificités anatomiques des segmentations manuelles ne sont correctement segmentées par aucune des segmentations automatiques. C’est le cas par exemple des lobes des parotides (coupe de la première colonne).

4.5.5 Robustesse aux structures manquantes

4.5.5.1 Importance de ce critère

Nous présentons dans cette section une analyse de la robustesse des différentes méthodes présentées face au taux de structures manquantes dans la base de segmentations. Ce critère a de l’importance car il conditionne l’applicabilité des différentes méthodes sur des bases de segmentations plus ou moins complètes. Par exemple, dans notre cas, nous sommes amenés à intégrer progressivement de nouvelles structures dans l’atlas, par exemple les aires ganglionnaires IB et V. En pratique, il est

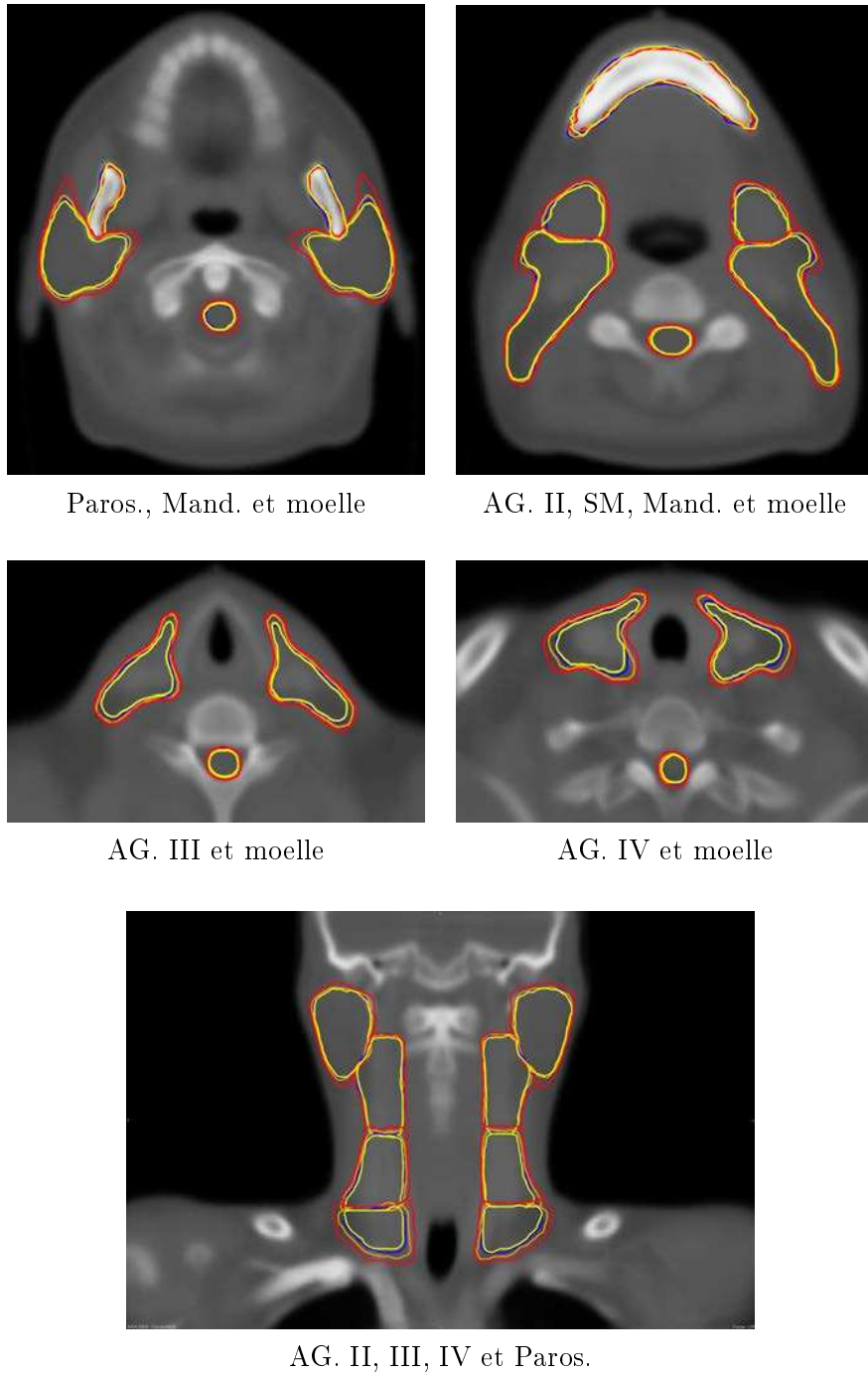


FIG. 4.11 – Comparaison qualitative entre les contours moyens obtenus avec STAPLE_STD (en rouge), STAPLE_REG (en orange), ACC_LOC (en jaune) et MAJ-VOTE_SMOOTH_MONO (en bleu).

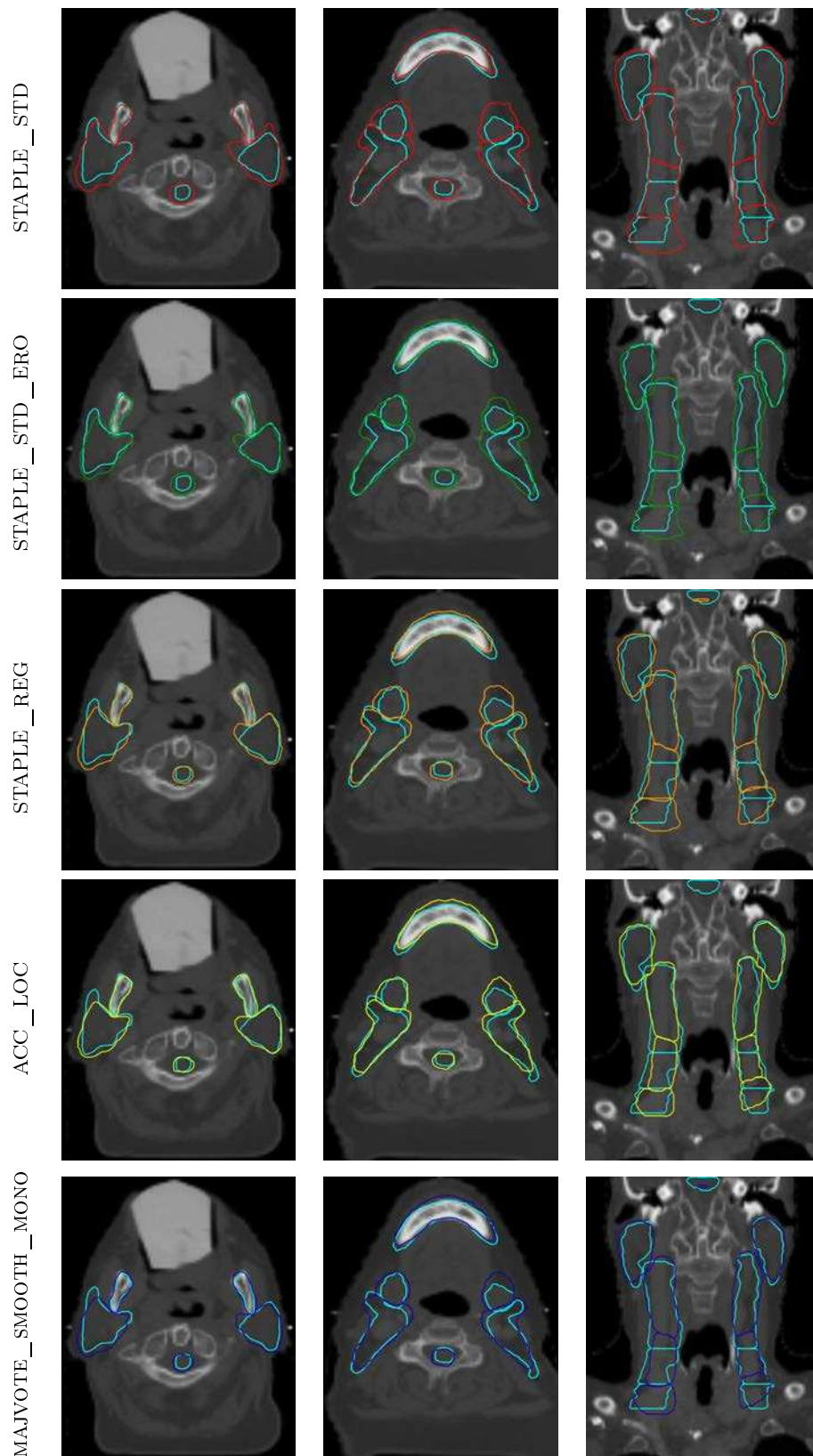


FIG. 4.12 – Résultats visuels obtenus sur trois coupes d'un patient particulier avec les méthodes STAPLE_STD (ligne 1), STAPLE_STD_ERO (ligne 2), STAPLE_REG (ligne 3), ACC_LOC (ligne 4) et MAJVOTE_SMOOTH_MONO (ligne 5). Sur chaque ligne sont représentées la segmentation manuelle en bleu clair et la segmentation automatique d'une autre couleur.

souvent compliqué de demander au radiothérapeute de re-contourer ces nouvelles structures sur les images de notre base. Il est beaucoup plus simple par contre de récupérer de nouvelles images dans lesquelles ces nouvelles structures sont contourées en plus des structures classiques. Cela contribue à augmenter le taux de structures manquantes dans la base de segmentations, car alors le nombre total de segmentations augmente, et un faible nombre seulement (uniquement les nouvelles) ont les nouvelles structures segmentées. Nous avons donc besoin d'une méthode de calcul de segmentation moyenne qui soit robuste dans le cas où certaines structures sont très peu représentées dans la base.

4.5.5.2 Sensibilité des méthodes basées sur STAPLE

Nous avons mis en évidence que les méthodes utilisant STAPLE (STAPLE_STD, STAPLE_STD_COR, STAPLE_REG) étaient assez sensibles à un fort taux de structures manquantes, notamment lorsque une structure en particulier est très peu représentée. En effet, les résultats obtenus pour la mandibule (dont la base ne comporte que 11 segmentations sur 105) sont significativement moins bons avec ces méthodes qu'avec les méthodes ACC_LOC et MAJVOTE_SMOOTH_MONO (cf. figures 4.9 et 4.10). En pratique, ces résultats restent acceptables en terme de qualité de segmentations comme on peut le voir sur les figures 4.11 et 4.12. Cependant, cela pose tout de même problème.

En effet, lors de la détermination des valeurs optimales à donner aux $\theta_{k's}$ lors de l'initialisation, nous avons été confronté à une forte sensibilité de STAPLE à la faible proportion de contourages de mandibule dans la base de segmentations. Nous avons évalué quantitativement selon les mesures de l'annexe B les résultats obtenus avec différentes initialisations. Les paramètres optimaux pour l'ensemble des structures sauf la mandibule donnaient des segmentations aberrantes pour la mandibule (dans la plupart des cas, les labels assignés à la mandibule dans la segmentation consensus se situaient en bordure des autres structures). Afin d'obtenir une segmentation correcte pour la mandibule, nous avons été contraints d'utiliser des paramètres d'initialisation qui n'étaient pas optimaux pour l'ensemble des autres structures.

D'autre part, nous avons également testé l'algorithme présenté dans [Commowick 2010] qui est une adaptation de STAPLE au cas de structures manquantes dans la base. Une étude exhaustive sur les valeurs données aux paramètres de l'algorithme n'a pas permis de trouver une configuration dans laquelle la mandibule était segmentée correctement. Suivant les cas, soit elle était absente de la segmentation consensus, soit elle était bien localisée mais englobait également les voxels en bordure des autres structures.

Nos expériences montrent que ce qui perturbe l'algorithme STAPLE est l'existence d'un fort déséquilibre dans les pourcentages de structures manquantes, quand en particulier une ou plusieurs structures est/sont représentée(s) à un taux largement inférieur au reste des structures. L'algorithme a en revanche tendance à être assez robuste au cas où toutes les structures sont manquantes selon un même taux raisonnable.

4.5.5.3 Sensibilité de la méthode ACC_LOC basée sur le degré d'accord local

La méthode ACC_LOC présentée en section 4.4 consiste à déterminer en chaque voxel les deux structures les plus représentées dans la base en ce voxel, et à calculer un degré d'accord local en se restreignant aux segmentations de la base ayant ces deux structures segmentées. Elle dépend donc *a priori* du taux de structures manquantes dans la base, et en particulier du taux de structures manquantes pour chaque couple de structures voisines.

Le tableau 4.13 présente pour chaque couple de structure (s, s') le nombre de segmentations parmi les 105 dans lesquelles les structures s et s' sont simultanément segmentées. Les cases violettes représentant les structures voisines permettent de visualiser les pires cas de figure pour la méthode ACC_LOC. Les pires cas de figure sont les cas II-mandibule pour lesquels seulement 7 ou 9 segmentations de la base possèdent les deux contourages. Cependant, ces cas de figure sont peu probables car ces deux structures ne sont voisines que ponctuellement. Toutes les autres cases violettes correspondent en revanche à des frontières non ponctuelles.

	II g.	II d.	III g.	III d.	IV g.	IV d.	Paro g.	Paro d.	SM g.	SM d.	Moelle	TC	Mand.
II g.	52	25	49	31	48	41	52	52	15	15	41	32	9
II d.	25	47	33	44	39	37	45	47	11	11	42	32	7
III g.	49	33	72	43	65	54	71	72	20	21	60	47	10
III d.	31	44	43	60	49	52	58	60	18	18	53	41	9
IV g.	48	39	65	49	85	71	84	85	22	24	71	57	10
IV d.	41	37	54	52	71	77	76	76	24	22	64	52	10
Paro g.	52	45	71	58	84	76	103	102	27	27	88	67	11
Paro d.	52	47	72	60	85	76	102	104	26	27	88	69	11
SM g.	15	11	20	18	22	24	27	26	27	25	19	13	10
SM d.	15	11	21	18	24	22	27	27	25	27	20	13	10
Moelle	41	42	60	53	71	64	88	88	19	20	89	64	10
TC	32	32	47	41	57	52	67	69	13	13	64	69	5
Mand.	9	7	10	9	10	10	11	11	10	10	10	5	11

FIG. 4.13 – Nombres de segmentations parmi les 105 dans lesquelles deux structures sont simultanément segmentées. Les cases roses en diagonale correspondent aux nombres de segmentations K_s pour chaque structure s , et les cases violettes correspondent aux cas de structures potentiellement voisines deux à deux.

Pour étudier l'influence des structures manquantes, considérons deux structures voisines s et s' , et plaçons nous dans le pire cas, c'est-à-dire le cas où seulement deux segmentations de la base ont ces deux structures segmentées. Au voisinage de la frontière entre s et s' , les deux structures les plus probables seront s et s' , et le degré d'accord local calculé en se restreignant aux deux segmentations pertinentes vaudra alors soit 0.5 (si les deux segmentations sont en désaccord pour ce voxel) soit 1 (si les deux segmentations sont en accord pour ce voxel). Si les deux segmentations

sont d'accord localement au niveau de la frontière, alors celle-ci ne correspondra pas à un minimum d'accord, et la ligne de partage des eaux reliera l'extérieur de la structure à son intérieur. Ce risque n'est pas négligeable, et peut donc amener à des résultats dégradés, en particulier si seulement deux segmentations ont s et s' segmentées. Le risque devient moins probable dès qu'un nombre suffisant (3 ou 4) de segmentations ont les structures s et s' segmentées. Un moyen de réduire ce risque est de calculer une version lissée de la carte d'accord local : au lieu de définir l'accord local au voxel i comme $\omega(i)$, on le définit comme la moyenne des $\omega(j)$ pour tout voxel j dans le voisinage de i . Ainsi, le risque d'avoir un accord local de 1 sur la frontière est considérablement réduit. Nous notons cette version lissée `ACC_LOC_SMOOTH` (voisinage de taille $3 \times 3 \times 3$ centré en chaque voxel).

4.5.5.4 Comparaison quantitative de la robustesse des différentes méthodes

Afin d'étudier quantitativement la robustesse des différentes méthodes aux structures manquantes, nous avons simulé différents taux de structures manquantes dans notre base, puis évalué les performances des atlas construits avec les différentes méthodes pour chaque taux.

Simulation de différents taux de structures manquantes Nous avons voulu simuler différents taux de structures manquantes, allant du cas sans structures manquantes au cas où un fort déséquilibre est présent entre les taux de présence des différentes structures. Pour cela, nous avons restreint le nombre de structures étudiées afin de pouvoir simuler le cas sans aucune structures manquantes. Plus précisément, nous avons créé une base de 83 segmentations multilabels dans laquelle toutes les aires ganglionnaires gauches et la parotide gauche étaient segmentées :

- d'une part, nous avons utilisé les 46 segmentations de notre base dans lesquelles ces 4 structures étaient présentes,
- d'autre part, nous avons symétrisé selon leur axe sagittal médian les 37 images de la base ayant les aires ganglionnaires droites et la parotide droite segmentées ; nous les avons ensuite recalées avec l'image moyenne afin de déformer leurs segmentations dans le référentiel de l'atlas moyen.

A partir de cette base de 83 segmentations complètes, nous avons simulé différents taux de structures manquantes en retirant plus ou moins de structures dans les 83 segmentations, chaque taux étant relié à un facteur λ . $\lambda = 0$ correspond à la base de segmentation complète, sans aucune structure manquante. $\lambda = 1$ correspond à la base de segmentation complète à laquelle on a retiré des structures de façon à avoir, pour chaque sous-ensemble de structures possible (par exemple, II. g./IV. g./Paro. g. ou encore II. g./Paro. g. ou encore III. g.) le même ratio d'images ayant segmenté exactement ce sous-ensemble de structures (ni plus ni moins) que dans la base de 105 patients. Pour les valeurs $0 < \lambda < 1$ et $\lambda > 1$, nous avons procédé de façon similaire en assurant une progression linéaire. Les bases de segmentations ainsi générées pour différents taux λ sont présentées sur la figure 4.14.

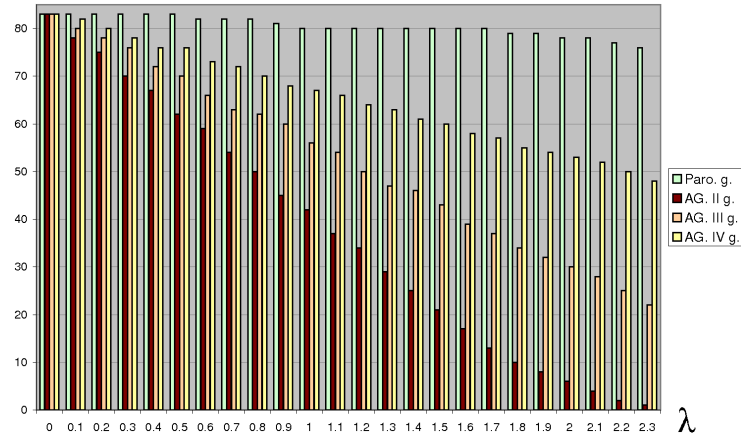


FIG. 4.14 – Simulation de différents taux de structures manquantes à partir de la base de 83 segmentations sans structure manquante ayant les aires ganglionnaires II, III et IV gauches ainsi que la parotide gauche contourées (la base complète correspond au cas $\lambda = 0$).

Résultats Pour chaque taux de structures manquantes λ , nous avons construit la segmentation moyenne correspondante avec les différentes méthodes proposées, puis nous avons utilisé les atlas moyens ainsi créés pour segmenter automatiquement les 83 patients. Nous avons ensuite évalué les performances de chaque atlas pour chaque λ avec la mesure DICE calculée entre les contours automatiques et manuels.

Nous nous intéressons ici à la qualité de la segmentation automatique de l'aire ganglionnaire II, puisque c'est celle qui a le taux de structures manquantes le plus important quel que soit λ . La figure 4.15 présente l'évolution du DICE moyen sur l'aire ganglionnaire II des 83 patients en fonction de λ , et ce pour différentes méthodes de calcul de segmentation moyenne.

Pour $\lambda < 1.6$, chaque méthode semble peu affectée par l'évolution du taux de structures manquantes, et affiche un DICE moyen environ constant sur cette plage. Entre $\lambda = 1.7$ et $\lambda = 2.3$, on note une diminution progressive du DICE moyen pour l'ensemble des méthodes à l'exception de MAJVOTE_SMOOTH_MONO. Cette diminution n'a pas la même amplitude ni la même pente suivant les méthodes. Pour STAPLE_STD et STAPLE_STD_COR (courbes noire et blanche), la diminution apparaît de façon marquée dès $\lambda = 1.8$, avec une forte amplitude et une forte pente. Cela confirme la haute sensibilité de STAPLE au cas où une structure est faiblement représentée dans la base (l'aire II est présente dans 10 segmentations sur 83 pour $\lambda = 1.8$). Pour les méthodes basées sur l'accord local (deux courbes vertes), la chute du DICE moyen apparaît pour des valeurs plus grandes de λ ($\lambda = 2$, correspondant à 6 segmentations sur 83 pour l'aire II), et elle est de moindre amplitude (pour $\lambda = 2.3$, DICE ≈ 0.4 pour ACC_LOC et ACC_LOC_SMOOTH, alors que DICE ≈ 0.1 pour STAPLE_STD et STAPLE_STD_COR). De plus, on remarque que comme attendu, l'utilisation de la version lissée de la carte d'accord

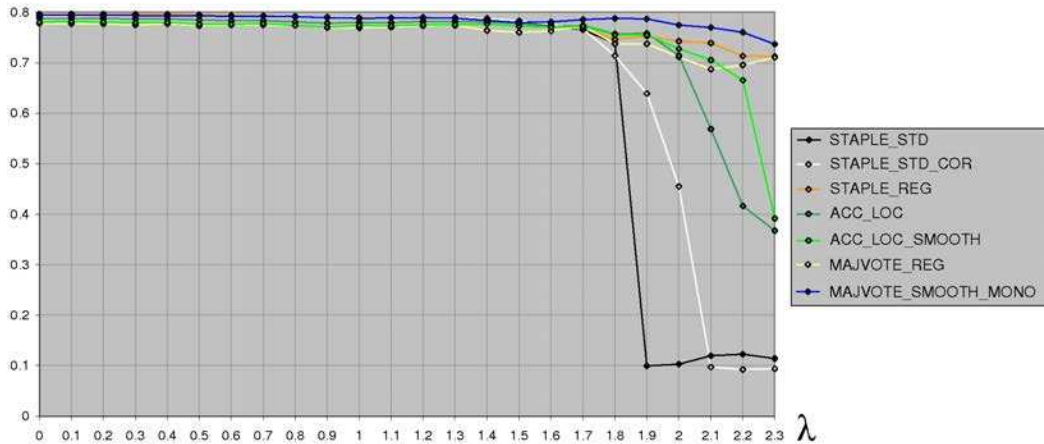


FIG. 4.15 – Évolution du DICE moyen sur l'aire ganglionnaire II en fonction du facteur λ pour les atlas moyens construits selon différentes méthodes.

local (ACC_LOC_SMOOTH) est plus robuste aux structures manquantes que la version voxelique (ACC_LOC). Enfin, les trois dernières méthodes (STAPLE_REG, MAJVOTE_REG et MAJVOTE_SMOOTH_MONO, en orange, jaune, bleu) sont les plus robustes à un fort déséquilibre dans les structures manquantes. En effet, la diminution du DICE moyen avec λ est très faible pour ces courbes, et la qualité de segmentation reste correcte (DICE > 0.7) même pour le cas extrême $\lambda = 2.3$, qui correspond à 1 seule segmentation de l'aire II parmi les 83 segmentations multilabels.

4.6 Limites de l'utilisation d'un atlas moyen

Nous avons présenté dans les sections précédentes des méthodes permettant de construire une segmentation moyenne à partir d'une base de segmentations incomplètes mises en correspondance. Chacune de ces méthodes fournit ainsi un atlas moyen capable de segmenter une image quelconque par recalage puis déformation de la segmentation moyenne sur l'image du patient. Cependant, cette approche a ses limites.

En effet, nous avons mis en évidence, notamment sur la figure 4.12, que l'ensemble des méthodes proposées présentaient des difficultés pour segmenter certaines particularités anatomiques présentes sur les images. Les différences entre les segmentations moyennes obtenues selon les différentes méthodes relevaient en effet davantage d'un facteur d'échelle que de vraies différences anatomiques. Cette difficulté à segmenter correctement les détails anatomiques est due à deux types de variabilité : d'une part une variabilité anatomique réelle à l'intérieur de la population, et d'autre part une variabilité de contourage intra- et inter-experts (décrite dans l'annexe C).

Concernant la variabilité de contourage, nous sommes malheureusement tributaires de ce facteur puisque nous utilisons les segmentations manuelles comme vérité terrain lors de l'évaluation des segmentations automatiques. S'en affranchir néces-

siterait de mettre en œuvre des processus d'évaluation clinique des différentes méthodes testées, mais ceci est difficile en pratique quand le nombre de méthodes que l'on veut comparer est élevé.

Concernant la variabilité anatomique, le recalage non-linéaire devrait en théorie permettre de récupérer les détails anatomiques de chaque patient, mais en pratique, cela n'est vrai que dans une certaine mesure. La première raison est liée au recalage non-linéaire en lui-même : celui-ci doit en effet fournir un champ de déformation suffisamment lisse pour que la transformation soit cohérente par rapport à l'anatomie sous-jacente. La seconde raison est liée au fait que l'image moyenne de l'atlas avec laquelle est recalée le patient est relativement floue, et possède peu de détails anatomiques contrastés en dehors des os.

Afin de rendre l'image moyenne moins floue, nous avons mis en œuvre l'algorithme de Guimond et al. [Guimond 2000] en appliquant un rehaussement de contraste à chaque itération. La figure 4.16 compare l'image moyenne ainsi obtenue à l'image moyenne obtenue sans rehaussement de contraste. L'apport du rehaussement de contraste se situe à deux niveaux : d'une part au niveau de la construction de la segmentation moyenne (car la mise en correspondance des segmentations de la base est potentiellement améliorée), et d'autre part au niveau du recalage entre l'image moyenne et l'image du patient à segmenter. Afin d'évaluer l'apport du rehaussement de contraste, nous avons comparé les performances des différents atlas moyens ainsi créés en terme d'indice de DICE, et ce pour les différents algorithmes de calcul de segmentation moyenne proposés. Le rehaussement de contraste apporte une amélioration marginale mais significative lorsque la segmentation moyenne est construite avec STAPLE_REG (p-value de 10^{-8}). En revanche, lorsque la segmentation moyenne est construite avec MAJVOTE_SMOOTH_MONO, l'amélioration n'est pas significative (p-value de 0.506). Enfin, lorsque la segmentation moyenne est construite avec ACC_LOC, l'utilisation du rehaussement de contraste dégrade les résultats de façon marginale mais significative (p-value de 10^{-10}).

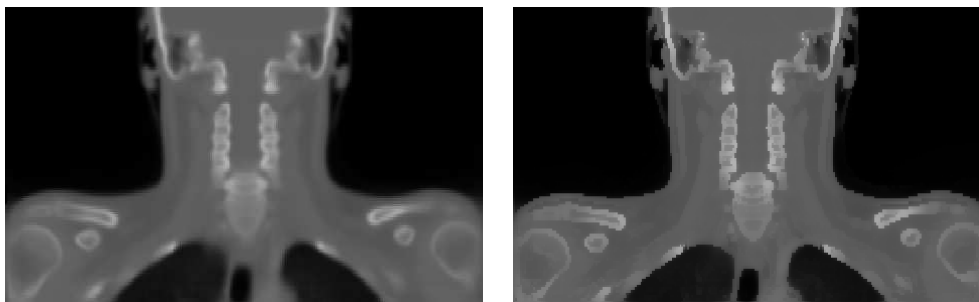


FIG. 4.16 – Image moyenne construite selon l'algorithme de Guimond et al. [Guimond 2000] avec (droite) et sans (gauche) rehaussement de contraste.

La grande variabilité anatomique présente pour les structures de la sphère ORL limite donc la capacité de l'atlas moyen à segmenter les images. Le chapitre 5 présente des stratégies alternatives pour tenter de dépasser cette limite.

4.7 Conclusion et synthèse des contributions

Dans ce chapitre, nous avons comparé et proposé différentes méthodes pour construire une segmentation multilabel moyenne à partir d'une base de segmentations incomplètes, c'est-à-dire possédant des contourages manquants. Dans un premier temps, nous avons mis en évidence une sur-segmentation en utilisant l'algorithme STAPLE [Warfield 2004]. Nous avons analysé en détail l'origine de cette sur-segmentation et présenté une explication cohérente. Basé sur cette analyse, nous avons proposé une correction de la gestion des structures manquantes dans STAPLE. Cette correction s'est avérée n'apporter qu'une amélioration marginale sur les résultats quantitatifs et qualitatifs.

Nous avons ensuite présenté une méthode de correction *a posteriori* de la sur-segmentation dans la segmentation moyenne basée sur des opérations de morphologie mathématique. Cette correction a permis d'améliorer les résultats mais pas pour l'ensemble des structures.

Nous avons alors proposé une approche permettant de gérer efficacement les structures manquantes lors du calcul de la segmentation moyenne. Cette approche peut s'appliquer à n'importe quelle méthode de calcul de segmentation moyenne probabiliste. Nous l'avons mis en œuvre avec STAPLE ainsi qu'avec le vote par majorité. Dans la mise en œuvre avec STAPLE, nous avons montré quantitativement que l'utilisation de cette approche permettait de réduire significativement la sur-segmentation par rapport à l'algorithme STAPLE classique.

Enfin, nous avons présenté un algorithme original permettant de calculer une segmentation moyenne en estimant le degré d'accord local entre les segmentations. Une évaluation quantitative a mis en évidence l'aptitude de cette méthode à fournir des contours moyens non sur-segmentés. Les résultats ont montré une tendance à la sous-segmentation, mais avec une amplitude moindre en comparaison de la sur-segmentation obtenue avec les méthodes utilisant STAPLE. Nous avons également montré que cet algorithme était davantage robuste à un fort taux de structures manquantes dans la base que les méthodes basées sur STAPLE, notamment lorsqu'une structure en particulier est très peu représentée par rapport aux autres.

Nous avons également mis en œuvre un algorithme de vote par majorité lissé qui s'est avéré fournir des résultats de qualité comparable (voire supérieure) à la méthode basée sur l'accord local.

Algorithmes de segmentation multi-atlas adaptés au patient

Sommaire

5.1	Introduction	102
5.1.1	Motivations et vocabulaire	102
5.1.2	Notations	104
5.2	Méthodes de sélection des atlas les plus adaptés au patient	104
5.2.1	Revue des critères de sélection basés sur des méta-informations	104
5.2.2	Revue des critères de sélection basés intensité	104
5.2.3	Revue des critères de sélection basés déformation	106
5.2.4	Revue des critères de sélection basés sur des segmentations	107
5.2.5	Contributions relatives aux critères de sélection	108
5.2.6	Mise en œuvre des méthodes de sélection	109
5.3	Évaluation non biaisée des méthodes de sélection	113
5.3.1	Corrélation moyenne de classement entre méthodes de sélection	114
5.3.2	Algorithme de partitionnement par affinité-propagation	114
5.3.3	Algorithme d'évaluation non biaisée	115
5.4	Différentes stratégies pour la fusion des atlas sélectionnés	116
5.4.1	Préambule	116
5.4.2	Construction hors-ligne de sous-atlas	117
5.4.3	Création à la volée d'atlas adaptés au patient	119
5.4.4	Fusion des segmentations dans le référentiel du patient	122
5.5	Conclusion et synthèse des contributions	128

Synthèse des contributions Les principales contributions présentées dans ce chapitre sont les suivantes :

- Proposition de deux critères de sélection d'atlas : (1) d'après des méta-informations adaptées à notre problématique (localisation et stade de tumeurs), et (2) d'après des mesures de contraction et de dilatation des structures,
- Proposition d'un protocole d'évaluation non biaisée des méthodes de sélection d'atlas,
- Proposition d'une méthode de construction à la volée d'atlas moyennement adapté au patient,
- Proposition d'une méthode de pondération locale des segmentations des atlas.

Publications associées Les travaux présentés dans ce chapitre ont donné lieu à plusieurs articles présentés dans des conférences internationales avec comité de lecture. Tout d’abord, la construction et l’utilisation des sous-atlas par localisation et stade de tumeur a été présentée à l’ESTRO’10 [Ramus 2010c]. L’approche d’évaluation non biaisée des méthodes de sélection d’atlas a été présentée à ISBI’10 [Ramus 2010e]. La construction d’un atlas moyen composite adapté au patient a été présenté à MICCAI’10 [Ramus 2010d]. La méthode de segmentation par pondération locale des segmentations de la base a été présentée au workshop satellite de MICCAI’10 dans le cadre du challenge de segmentation des parotides [Ramus 2010a].

5.1 Introduction

5.1.1 Motivations et vocabulaire

La segmentation par atlas repose sur l’hypothèse selon laquelle il est possible d’estimer une transformation non-linéaire permettant de superposer de façon exacte l’image de l’atlas et l’image du patient. Cependant, nous souhaitons également que la transformation non-linéaire permettant de mettre en correspondance les deux images soit anatomiquement cohérente. Une façon d’y parvenir est de rechercher une transformation non-linéaire relativement régulière. Un compromis doit alors être trouvé entre la régularité de la transformation et la précision du recalage. Un moyen de faciliter la détermination de ce compromis (et donc d’améliorer la précision du recalage et de la segmentation automatique résultante) est d’utiliser un atlas qui soit le plus similaire possible à l’image du patient.

Dans le chapitre précédent, nous avons présenté des méthodes permettant de construire un atlas moyen à partir d’une base d’images manuellement contournées. L’image moyenne ainsi construite présente alors une anatomie typique par rapport à la population étudiée, ce qui lui permet d’être relativement similaire à un grand nombre de patients. Cependant, comme nous l’avons évoqué l’approche consistant à utiliser un seul et même atlas pour segmenter toute nouvelle image ne permet pas de prendre en compte de façon robuste la forte variabilité anatomique présente dans la région ORL. Notamment, elle a des difficultés à segmenter correctement les patients ayant une anatomie atypique.

Étant donné une base d’images manuellement contournées, une alternative simple consiste à considérer chaque image de la base comme un atlas potentiel et à sélectionner celui qui est le plus similaire au patient. Cette procédure dite de *sélection d’atlas* peut être appliquée de différentes façons comme nous le détaillons en section 5.2. Elle permet en théorie de rendre le recalage entre l’atlas et le patient plus précis, et donc d’améliorer la qualité de la segmentation du patient. Cependant, la sélection d’une seule image de la base (et donc d’une seule segmentation) en tant qu’atlas introduit un fort biais lié à la variabilité de contourage manuel. Afin de s’affranchir de ce biais, une solution consiste alors non plus à sélectionner une seule image dans la base, mais à sélectionner un certain nombre d’images de la base parmi les plus similaires au patient, et à les combiner pour estimer la segmentation du pa-

tient. Cela fait cette fois-ci appel à deux ingrédients, illustrés sur la figure 5.1 : la *sélection d'atlas* et la *fusion d'atlas*. Dans tout le chapitre, nous utilisons le terme *fusion d'atlas* pour qualifier la combinaison des segmentations et éventuellement des images des atlas sélectionnés pour estimer la segmentation du patient. Cette procédure peut se mettre en œuvre de multiples façons, qui sont détaillées dans la section 5.4. Si les atlas sont indépendants (ce qui est le cas si on utilise une base d'images correspondant à des patients différents), les erreurs locales apportées par chaque atlas (dues par exemple à une erreur de segmentation manuelle ou encore à une erreur de recalage) sont indépendantes, et la fusion de plusieurs atlas permet en théorie d'améliorer la segmentation automatique.

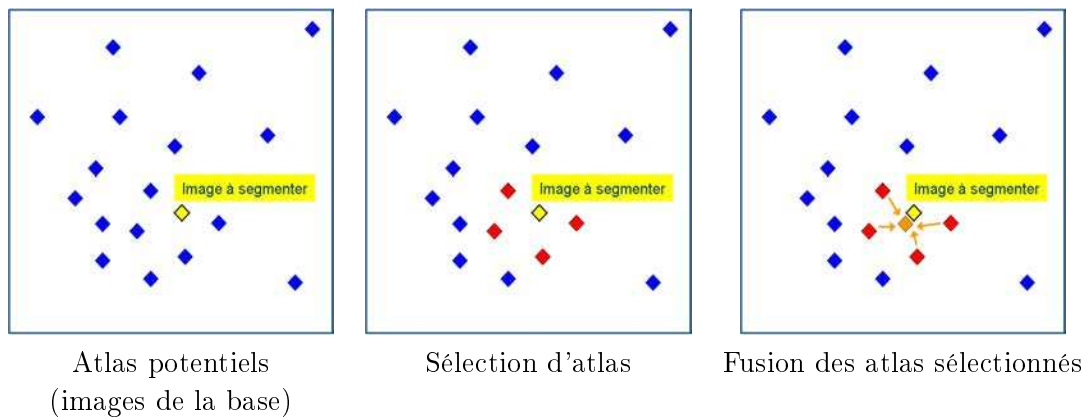


FIG. 5.1 – Illustration des principes de sélection et de fusion d'atlas.

Dans la littérature, le terme *multi-atlas* est en général utilisé pour parler du recalage direct de plusieurs atlas avec l'image du patient, et de la fusion de leurs segmentations propagées dans le référentiel du patient. Dans ce chapitre, nous utilisons ce terme dans un sens plus large pour qualifier tout algorithme de fusion d'atlas au sens où nous l'avons défini plus haut. Lorsque les atlas fusionnés sont issus d'un processus de sélection d'atlas, nous utilisons le terme d'algorithme *multi-atlas adapté au patient*.

L'objectif de ce chapitre est de présenter un état de l'art approfondi des principales approches de sélection et de fusion d'atlas présentées dans la littérature, et de proposer des solutions nouvelles pour tenter d'améliorer les résultats dans le cadre de la segmentation des structures ORL. Nous présentons en sections 5.2 et 5.3 l'état de l'art des méthodes de sélection d'atlas ainsi que nos contributions relatives à cette étape, à savoir : deux nouveaux critères de sélection et un algorithme permettant d'évaluer de façon non biaisée les méthodes de sélection d'atlas. Nous abordons ensuite en section 5.4 l'état de l'art relatif à l'étape de fusion ainsi que nos contributions, à savoir : un algorithme de construction d'atlas moyen composite régionalement adapté au patient, et un algorithme de pondération locale des segmentations dans le référentiel du patient. Les résultats relatifs aux différentes méthodes présentées dans ce chapitre sont exposés dans le chapitre 6.

5.1.2 Notations

Dans tout le chapitre, on note $T_{A \leftarrow B}$ la transformation permettant de ré-échantillonner l'image A sur l'image B. Cette transformation se décompose en une composante affine notée $T_{A \leftarrow B}^{aff}$ et une composante non-linéaire notée $T_{A \leftarrow B}^{nl}$: $T_{A \leftarrow B} = T_{A \leftarrow B}^{aff} \circ T_{A \leftarrow B}^{nl}$. Les transformations affines et non linéaires sont estimées en utilisant les algorithmes de recalage par blocs détaillés dans l'annexe A. La transformation $T_{A \leftarrow B}$ permet de déformer l'image A ainsi que sa segmentation S_A sur l'image B, et les notations utilisées sont alors $A \circ T_{A \leftarrow B}$ et $S_A \circ T_{A \leftarrow B}$.

On note P l'image d'un patient à segmenter, $\{I_k\}_{k \in [1 \dots N]}$ les N images de la base (aussi appelées atlas) et $\{S_k\}_{k \in [1 \dots N]}$ leurs segmentations manuelles.

5.2 Méthodes de sélection des atlas les plus adaptés au patient

La sélection des atlas les plus adaptés au patient nécessite de quantifier la similarité entre l'image du patient et chacune des images de la base. Cette similarité peut être basée sur des méta-informations (cf. section 5.2.1) ou bien sur des mesures calculées directement à partir des images. Dans ce dernier cas, l'objectif est d'identifier les images de la base pour lesquelles la mise en correspondance par recalage non-linéaire sera la plus précise possible. La mesure de similarité peut alors être basée sur les intensités (cf. section 5.2.2), sur les déformations (cf. section 5.2.3), ou encore sur des segmentations (cf. section 5.2.4). D'autre part, quel que soit le critère de similarité choisi, il peut être mis en œuvre de différentes façons (cf. section 5.2.6).

5.2.1 Revue des critères de sélection basés sur des méta-informations

Dans la littérature, les approches de sélection basées sur des méta-informations sont en général utilisées pour des images IRM du cerveau, et le critère de sélection est alors l'âge [Ericsson 2008] et éventuellement le sexe du patient. Dans notre cas, l'âge n'est pas un critère adapté car il n'a *a priori* pas d'impact sur les structures de la région ORL. Des critères de sélection basés sur la pathologie peuvent aussi être utilisés, mais peu d'études s'y sont intéressés. La seule information exploitée est en général le statut sain ou pathologique des images disponibles.

5.2.2 Revue des critères de sélection basés intensité

Les méthodes basées intensité consistent à comparer l'image du patient avec chacune des images de la base, et ainsi identifier parmi les images de la base celles qui sont le plus similaires à celle du patient. L'hypothèse sous-jacente à ce type de méthodes est que la similarité entre l'image du patient et celle d'un atlas déformé sur le patient est un bon prédicteur de la qualité du recalage entre les deux images, et donc de la qualité de la propagation des labels de l'atlas sur le patient.

Comme illustré sur la figure 5.2, il est nécessaire, pour pouvoir quantifier la similarité entre l'image du patient P et chacune des images I_k de la base :

- de les mettre en correspondance dans un référentiel commun par recalage, ce qui nécessite de choisir le référentiel commun ainsi que le type de mise en correspondance, c'est-à-dire l'algorithme de recalage utilisé (cf. section 5.2.2.1),
- de définir une mesure de similarité entre les images normalisées \tilde{P} et \tilde{I}_k (cf. section 5.2.2.2).

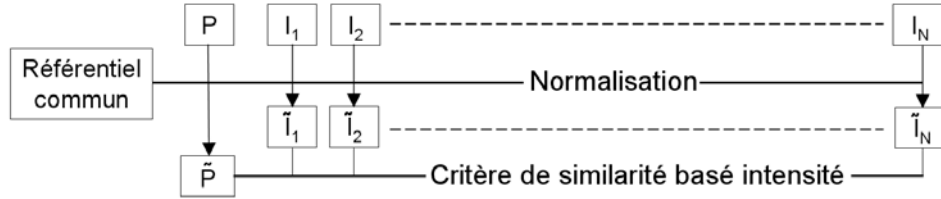


FIG. 5.2 – Sélection d’atlas d’après un critère de similarité basé intensité.

5.2.2.1 Référentiel de sélection et mise en correspondance

La très grande majorité des articles de la littérature utilisent comme référentiel commun l’image du patient à segmenter : la mesure de similarité est alors calculée entre l’image du patient à segmenter et chacune des images de la base déformées de façon affine [Aljabar 2009, Gorthi 2009a, Gorthi 2010, Teguh 2010, Zhuang 2010a, Leung 2010, Han 2008] ou non-linéaire [Klein 2008, Zhang 2009] sur le patient. Le choix d’une mise en correspondance de type affine pour la sélection suppose que la similarité entre les deux images à la suite du recalage affine est suffisamment corrélée à leur similarité après recalage non-linéaire. Cela constitue une approximation, mais a l’avantage d’accélérer grandement la phase de sélection, surtout quand le nombre d’images dans la base est important. Certains auteurs ont mis en évidence de meilleurs résultats avec une sélection après recalage non-linéaire [Lötjönen 2009], mais d’autres ont obtenu de meilleurs résultats avec une sélection après recalage affine [Zhuang 2010a]. Un compromis intéressant peut consister à utiliser une mise en correspondance de type non-linéaire, mais avec un faible nombre d’itérations [van Rikxoort 2009, van Rikxoort 2010].

Il est également possible de calculer la similarité dans un référentiel externe au patient à segmenter, comme par exemple une image artificielle [Aljabar 2007, Aljabar 2009] ou encore un atlas moyen calculé à partir de l’ensemble de la base [Shi 2010]. Bien que peu utilisée dans la littérature, cette approche présente l’avantage de ne nécessiter qu’un seul recalage (celui entre le patient à segmenter et l’image de référence), puisque les recalages entre l’image de référence et les images de la base peuvent être faits à l’avance une fois pour toutes. Yang et al. mettent en correspondance l’ensemble des atlas dans un référentiel externe dans lequel ils procèdent ensuite à une analyse par composantes principales à partir des informations image des régions concernées [Yang 2010]. Cela leur permet de sélectionner hors-ligne par

Leave-One-Out sur la base d'atlas un nombre restreint d'atlas qui sont les plus représentatifs. Cette méthode pourrait cependant être appliquée à la volée pour sélectionner un ou des atlas adapté(s) au patient. Par ailleurs, le référentiel externe peut aussi être utilisé lorsque la mesure de similarité est faite dans le référentiel du patient : il sert alors de référentiel intermédiaire pour la propagation indirecte des atlas dans le référentiel du patient.

Les approches citées précédemment utilisent uniquement le référentiel externe commun comme outil de mise en correspondance de l'image du patient avec les atlas : l'image de ce référentiel commun (image moyenne ou image artificielle par exemple) n'est utilisée que pour faire les recalages mais elle n'intervient pas dans la mesure de similarité, qui se fait *in fine* entre l'image du patient et les atlas normalisés dans un même référentiel. Van Rikxoort et al. [van Rikxoort 2008] utilisent également un référentiel commun pour la sélection d'atlas, mais dans une optique un peu différente. Le référentiel externe ne sert plus ici à la mise en correspondance, mais il sert d'étalon pour estimer de façon indirecte les similarités entre l'image du patient et les images des atlas. D'une part l'image externe est déformée sur l'image patient, et une mesure de similarité (SAD, Sum of Absolute Differences) est alors calculée dans le référentiel du patient. D'autre part, l'image externe est déformée sur chacun des atlas, et des mesures de similarités (SAD) sont calculées dans les référentiels respectifs des atlas. Le ou les atlas fournissant une valeur de similarité avec l'image externe la plus proche de la valeur de similarité obtenue pour l'image patient est/sont alors sélectionné(s).

5.2.2.2 Mesure de similarité

Dans la littérature, le critère de similarité utilisé est le plus souvent une information mutuelle ou une information mutuelle normalisée [Zhang 2009, Lötjönen 2009, Klein 2008, Teguh 2010, Zhuang 2010a, Aljabar 2009, Shi 2010]. Certains auteurs utilisent des coefficients de corrélation [Leung 2010, Artaechevarria 2009], ou encore une SSD (Sum of Squared Differences) [Gorthi 2009a, Gorthi 2010].

5.2.3 Revue des critères de sélection basés déformation

Les méthodes de sélection basées intensité représentent les méthodes les plus intuitives puisqu'elles quantifient directement la ressemblance visuelle après mise en correspondance. Plusieurs auteurs ont proposé une alternative consistant à se baser non plus sur les intensités mais sur la déformation. Ces alternatives font le postulat qu'un atlas est d'autant plus adapté au patient que la déformation non-linéaire nécessaire à la mise en correspondance entre les deux images est de faible amplitude. Rohlfing et al. ont proposé de quantifier cette déformation non-linéaire sur l'ensemble de l'image et en utilisant un recalage non-linéaire direct entre l'image du patient et chaque atlas [Rohlfing 2004]. Ils ont étudié l'utilisation de la déformation moyenne et de la déformation maximale. Commowick et al. ont proposé de quantifier cette déformation non-linéaire en passant par un référentiel externe constitué

d'un atlas moyen [Commowick 2007a]. Cette approche a l'avantage de ne nécessiter qu'un seul recalage non-linéaire, celui de l'image du patient sur l'atlas moyen, puisque les recalages non-linéaires des atlas potentiels sur l'atlas moyen peuvent être faits hors-ligne.

5.2.4 Revue des critères de sélection basés sur des segmentations

Certains auteurs ont présenté des méthodes de sélection basées sur des comparaisons de segmentations de structures secondaires (c'est-à-dire autres que les structures d'intérêt). Par exemple, dans le cadre de la segmentation des lobes des poumons, van Rikxoort et al. obtiennent dans un premier temps la segmentation des fissures des poumons sur l'image du patient par une autre méthode de segmentation. Ensuite, ils sélectionnent l'atlas qui fournit après recalage non-linéaire sur le patient les segmentations des fissures les plus proches (selon une mesure de distance) de celles calculées précédemment sur l'image du patient [van Rikxoort 2009]. Teng et al. [Teng 2007] utilisent également ce type d'approche pour la segmentation des aires ganglionnaires IB et IIB dans la région ORL : la mandibule, l'os hyoïde, les veines jugulaires et le contour externe sont préalablement segmentées sur l'image du patient par une autre technique. Des comparaisons de volume et de position de centroïdes sont faits entre ces volumes et les volumes correspondants des atlas disponibles après recalage rigide sur le patient et permettent de sélectionner l'atlas le plus adapté.

Langerak et al. [Langerak 2009, Langerak 2010] utilisent également une sélection par comparaison de segmentations, mais cette fois-ci, les segmentations comparées ne sont pas des structures secondaires, mais les structures d'intérêt elles-mêmes. Ils initialisent la segmentation du patient grâce à un vote par majorité à partir des segmentations de l'ensemble des atlas, puis ils excluent de façon itérative les atlas qui fournissent les segmentations les moins proches de l'estimation courante de la segmentation du patient. Cette approche est basée sur l'estimation des performances de chaque segmentation (au même titre qu'elle avait été introduite dans l'algorithme STAPLE [Warfield 2004]). A la différence de STAPLE, l'estimation de performance ne sert pas seulement à pondérer les segmentations dans le processus de fusion d'atlas, mais intervient également dans le processus de sélection d'atlas pour exclure les segmentations les moins performantes.

Enfin, Heckemann et al. ont proposé un critère de sélection d'atlas qui a pour objectif de quantifier la fiabilité du recalage entre chaque atlas et le patient à segmenter [Heckemann 2009]. Ce critère est défini comme le chevauchement entre la segmentation de l'atlas et cette même segmentation déformée deux fois (en utilisant un recalage non-linéaire de l'atlas sur le patient, puis un recalage non-linéaire du patient sur l'atlas). Cette méthode de sélection requiert donc deux recalages non-linéaires par atlas, mais les temps de calcul peuvent être réduits en utilisant des recalages non-linéaires à faible résolution.

5.2.5 Contributions relatives aux critères de sélection

Nous proposons deux nouveaux critères de sélection d’atlas, l’un basé sur des méta-informations (cf. section 5.2.5.1), l’autre basé sur les déformations (cf. section 5.2.5.2).

5.2.5.1 Sélection d’atlas d’après la localisation et le stade de tumeur

Nous proposons un critère de sélection basé sur des méta-informations adaptées à la segmentation des structures ORL pour la radiothérapie. Ce critère est défini par la localisation et le stade de la tumeur primitive. L’hypothèse sous-jacente à ce critère est que la présence de la tumeur a un impact sur les structures environnantes. Pour la localisation, nous considérons cinq sites différents illustrés sur la figure 2.1 (larynx, hypopharynx, oropharynx, nasopharynx et cavité orale) ainsi que le côté (droit ou gauche) de la tumeur, constituant ainsi 10 sites tumoraux. Pour le stade de la tumeur, nous dissociions les stades N0 et N1 (pas d’atteinte des ganglions lymphatiques ou atteinte limitée) du stade N2. Au total, nous avons ainsi 20 catégories. Étant donné un patient à segmenter, les atlas correspondant à la même catégorie que le patient peuvent être sélectionnés pour l’étape de fusion d’atlas.

5.2.5.2 Sélection d’atlas d’après le degré de contraction ou de dilatation des structures

Nous proposons un critère de sélection d’atlas basé sur les déformations, et plus précisément sur la contraction ou dilatation des structures anatomiques. Étant donné un atlas moyen M déjà construit à partir de l’ensemble des atlas de la base, l’objectif de notre critère est de repérer les atlas dont les déformations subies pour se déformer sur l’atlas moyen M sont similaires aux déformations subies par l’image du patient P pour se déformer sur M . A la différence de [Commowick 2007a], nous proposons de quantifier la déformation non plus par son amplitude, mais par le degré de contraction ou de dilatation calculé pour chaque région d’intérêt.

Étant donné les images de la base $\{I_k\}_{k \in [1 \dots N]}$ et les transformations $\{T_{I_k \leftarrow M} = T_{I_k \leftarrow M}^{aff} \circ T_{I_k \leftarrow M}^{nl}\}_{k \in [1 \dots N]}$ permettant de les ré-échantillonner sur l’atlas moyen M , notre critère de sélection nécessite de calculer hors ligne (une fois pour toutes) les jacobiens $J_{I_k \leftarrow M}$ correspondant aux parties non-linéaires des déformations $T_{I_k \leftarrow M}^{nl}$. A partir de ces jacobiens, nous pouvons estimer sur chaque région d’intérêt R_l définie sur M le degré moyen de contraction ou de dilatation subi par l’atlas I_k pour se déformer sur M avec la formule suivante :

$$\bar{J}_{R_l}(I_k \leftarrow M) = \frac{1}{\text{card}(R_l)} \sum_{x \in R_l} \log(J_{I_k \leftarrow M}(x)) \quad (5.1)$$

Lors de la moyenne, le logarithme permet de considérer de façon symétrique une dilatation de volume d’un facteur r et une contraction de volume d’un facteur $\frac{1}{r}$. Dans la formule 5.1, nous avons choisi une approche voxelique (calcul du degré de dilatation/contraction moyen sur tous les voxels de R_l). Une approche régionale

aurait également été possible en prenant le logarithme de la moyenne au lieu de moyenner les logarithmes. Cette approche aurait alors quantifié la variation de volume de la région dans son ensemble. Après avoir recalé non-linéairement le patient P sur l’atlas M , nous pouvons calculer le jacobien $J_{P \leftarrow M}$ de la partie non-linéaire $T_{P \leftarrow M}^{nl}$ ainsi que sa moyenne sur chaque région R_l :

$$\bar{J}_{R_l}(P \leftarrow M) = \frac{1}{\text{card}(R_l)} \sum_{x \in R_l} \log(J_{P \leftarrow M}(x)) \quad (5.2)$$

Pour chaque région R_l , chaque atlas peut maintenant être comparé à l’image du patient en utilisant la distance suivante :

$$d_{R_l}(I_k, P) = \|\bar{J}_{R_l}(P \leftarrow M) - \bar{J}_{R_l}(I_k \leftarrow M)\| \quad (5.3)$$

Les atlas ayant les distances les plus petites au patient P sont alors sélectionnés région par région pour l’étape de fusion. Un avantage du critère proposé est qu’il est extrêmement rapide à calculer. En effet, il ne nécessite qu’un seul recalage non-linéaire et un calcul de jacobien. A titre de comparaison, beaucoup de critères de sélection nécessitent soit un recalage non-linéaire par atlas [Klein 2008, Zhang 2009] soit des calculs de composition et d’inversion de champ [Commowick 2007a], ce qui rend peu adaptée leur utilisation pour les grandes bases d’atlas.

Nous avons appliqué ce critère de sélection de façon régionale, structure par structure. Les régions d’intérêt R_l sont alors définies par des dilatations des structures moyennes de l’atlas moyen M . Cependant, bien que nous ne l’ayons pas testé, il serait envisageable de mettre en œuvre ce critère de façon locale en le calculant pour chaque voxel sur un certain voisinage. Dans ce dernier cas, les voisinages doivent avoir une taille suffisante pour que la moyenne du jacobien soit pertinente.

A noter que l’utilisation du jacobien pour caractériser la similarité entre deux images a aussi été proposée par Klein et al., mais leur approche est différente [Klein 2010]. Étant donné deux images et un recalage non-linéaire direct effectué entre ces deux images, ils utilisent l’écart-type du jacobien de la partie non-linéaire de la transformation sur un masque global du cerveau pour caractériser la dissimilarité entre les deux anatomies comparées.

5.2.6 Mise en œuvre des méthodes de sélection

Quels qu’ils soient, les critères de sélection basés image peuvent être mis en œuvre de différentes façons.

5.2.6.1 Sélection globale, régionale ou locale

Tout d’abord, le critère de sélection peut être appliqué à différentes échelles.

Sélection globale Il peut être appliqué de façon globale, auquel cas la sélection d’un atlas implique que l’ensemble des segmentations manuelles de ses

structures sont utilisées pour estimer la segmentation du patient. Dans la littérature, plusieurs auteurs ont utilisé cette solution [Lötjönen 2009, Teguh 2010, Zhuang 2010a, Commowick 2007a, van Rikxoort 2009, Rohlfing 2004, Aljabar 2009, Han 2008, Klein 2008]. Suivant les cas, le masque de sélection utilisé est défini par l'image entière, le contour externe du patient, un masque de l'union des structures d'intérêt, ou encore une boîte englobant l'ensemble des structures. Cependant, la sélection globale ne permet pas de prendre en compte le fait qu'un atlas peut avoir une anatomie proche de celle du patient dans une certaine zone, mais assez différente dans d'autres zones. En d'autres termes, elle ne permet pas de considérer que le recalage puisse avoir une qualité variable suivant les zones de l'image.

Sélection régionale Afin de dépasser ces limitations, beaucoup de publications utilisent une sélection régionale des atlas les plus adaptés via des masques appropriés pour chaque structure d'intérêt [Wu 2007, Commowick 2009, Aljabar 2007, Heckemann 2009, Langerak 2009]. Pour chaque structure, la segmentation du patient est alors estimée à partir des segmentations des atlas sélectionnés sur le masque correspondant. Pour chaque structure, le masque peut être défini par exemple par dilatation de l'union des segmentations de l'ensemble des atlas mis en correspondance. C'est le choix couramment utilisé, mais il est assez sensible à un éventuel échec de recalage pour l'un des atlas. De façon plus robuste, le masque peut aussi être défini par la propagation grossière (affine) d'une segmentation moyenne. Certains auteurs mettent en œuvre une sélection régionale qui est basée sur une parcellisation de l'image indépendamment des structures d'intérêt. Par exemple, Shi et al. définissent les régions de sélection en parcellisant une image moyenne construite à partir de l'ensemble des atlas [Shi 2010]. Dans le même style, van Rikxoort et al. partitionnent la structure d'intérêt (le cœur en l'occurrence) en 8 blocs de même taille [van Rikxoort 2010, van Rikxoort 2008]. Plus spécifique et plus ciblée que la sélection globale, la sélection régionale des atlas les plus adaptés au patient peut donc en théorie conduire à une segmentation plus précise. La principale difficulté introduite par la sélection régionale est la gestion des frontières entre structures lors de la fusion des atlas sélectionnés. Nous parlerons plus en détail de ce problème et des solutions possibles dans la section 5.4. À noter que dans le cas où les masques de sélection ne correspondent pas aux structures, le problème de continuité ne se pose plus seulement à la frontière entre deux structures, mais aussi au sein même de chaque structure (puisque alors une structure peut comporter des zones de plusieurs masques de sélection).

Sélection locale Il est également possible d'effectuer une sélection encore plus ciblée que la sélection régionale : c'est la sélection locale (voxel par voxel). Cela permet de prendre en compte le fait qu'un atlas puisse être similaire au patient sur une zone donnée d'une structure, mais pas sur le reste de la structure. Par exemple, pour les parotides, la sélection locale peut en théorie permettre de sélectionner plus finement les atlas qui ont un lobe accessoire si le patient en a un. La sélection locale a été utili-

sée dans [Isgum 2009, Artaechevarria 2009]. Dans ces deux articles, les atlas ne sont pas sélectionnés à proprement parler, mais leurs segmentations sont pondérées dans le référentiel du patient en fonction de la similarité locale entre l'image du patient et chacune des images d'atlas déformées sur le patient. Le principal inconvénient de la sélection locale est qu'elle est très sensible au bruit dans les images, surtout dans les zones peu contrastées. En pratique, plusieurs artifices permettent d'assurer une segmentation finale suffisamment lisse. Tout d'abord, la mesure de similarité entre le patient et chaque atlas gagne à être calculée sur un voisinage de chaque voxel, ce qui permet d'ajouter une régularisation intrinsèque à la segmentation. Plus le voisinage de calcul est grand, plus on se rapproche de la sélection régionale. Cependant, ce lissage intrinsèque ne suffit souvent pas. En supplément, il est alors possible d'appliquer un lissage gaussien additionnel soit aux valeurs de similarités soit à la carte de probabilité [Isgum 2009] ou segmentation résultante. Un autre exemple de pondération locale a été proposé par Sdika [Sdika 2010], mais dans ce cas les cartes de pondérations associées aux atlas sont calculées hors-ligne et traduisent l'accord local en terme de segmentation de chaque atlas par rapport aux autres. Enfin, comme pour la sélection régionale, la gestion des frontières entre structures pose problème. Nous aborderons ce problème lors de la fusion en section 5.4. A noter que la sélection/pondération locale s'applique préférentiellement avec les critères de similarités basés intensité (tels que ceux présentés en section 5.2.2) plutôt qu'avec des critères de similarité basés déformation (le critère de similarité basé sur les segmentations n'ayant par nature pas de sens à être appliqué localement).

5.2.6.2 Fonction de sélection ou de pondération

Les différents critères de sélection présentés en section 5.2.2, 5.2.3, et 5.2.4 permettent de classer, de façon globale, régionale ou locale, les atlas disponibles du plus similaire au moins similaire à l'image du patient. A partir de là, plusieurs approches sont envisageables pour exploiter ce ou ces classement(s). Les deux grandes catégories sont la sélection et la pondération, la sélection pouvant être vue comme un cas particulier de pondération (fonction escalier). La première solution consiste à sélectionner les atlas les mieux classés et à leur donner ensuite un poids égal dans la fusion, les atlas en bas de classement étant alors ignorés. La seconde solution consiste à garder tous les atlas en compte dans l'étape de fusion, mais à pondérer leurs contributions en fonction de leur similarité respective à l'image du patient. L'inconvénient majeur de la pondération par rapport à la sélection est qu'elle nécessite de recalculer la totalité des atlas de façon non-linéaire sur le patient. Dans la littérature, la pondération a très peu été utilisée, et elle l'a principalement été lorsque la similarité était quantifiée de façon locale, voxel par voxel [Artaechevarria 2009, Isgum 2009]. La grande majorité des approches utilisent une sélection d'atlas, appliquée alors de façon globale ou régionale.

Dans le cas de la sélection d'atlas, la principale question à traiter est le nombre d'atlas sélectionnés. Tout d'abord, certains auteurs se sont dans un premier temps intéressés à la sélection unique de l'atlas le plus similaire au pa-

tient, et à son utilisation dans un schéma de segmentation par atlas classique [Commowick 2007a, Wu 2007, Teng 2007]. Cependant, cette solution s'avère très sensible à la phase de sélection d'une part et à la variabilité de contourage manuel d'autre part (puisqu'alors une seule segmentation manuelle est sélectionnée). De nombreuses études ont montré que la sélection d'un plus grand nombre d'atlas parmi les plus similaires permettait de s'affranchir de ces inconvénients et d'améliorer la précision des segmentations [Aljabar 2009, Heckemann 2009, van Rikxoort 2008, van Rikxoort 2010, Leung 2010, Zhuang 2010a]. Plus précisément, ces études ont montré qu'il existe généralement un nombre d'atlas optimal en deçà et au-delà duquel la qualité de la segmentation finale est dégradée ou stable : la qualité de la segmentation augmente rapidement et de façon significative jusqu'à atteindre le nombre optimal d'atlas, puis lorsque davantage d'atlas sont sélectionnés, les études reportent un pallier et dans certains cas une légère dégradation. Ainsi, utiliser un nombre trop élevé d'atlas peut non seulement représenter un inconvénient du fait du temps de calcul, mais cela peut également dégrader la qualité de la segmentation. Le nombre d'atlas optimal est fortement dépendant de l'application et du reste de l'algorithme, notamment de la règle utilisée pour la fusion de segmentation, comme mis en évidence dans [Leung 2010].

Par ailleurs, plusieurs schémas ont été proposés pour mettre en œuvre la sélection à partir des classements d'atlas :

- sélectionner un nombre fixe d'atlas [Leung 2010],
- sélectionner un pourcentage des atlas les plus similaires au patient [Klein 2008],
- sélectionner tous les atlas qui satisfont un certain critère (par exemple critère de similarité supérieur ou inférieur à un seuil [Langerak 2009, Langerak 2010]),
- sélectionner ou exclure les atlas de façon progressive.

Ce dernier type de schéma (sélection/exclusion progressive) est en général basé sur une alternance entre sélection et fusion, ce qui peut être vu comme une fonction de sélection adaptative. Le principe consiste à partir d'une estimation initiale de la segmentation, et à l'améliorer de façon itérative en raffinant les atlas sélectionnés [van Rikxoort 2010, van Rikxoort 2008, Langerak 2009, Langerak 2010]. Ce raffinement de sélection d'atlas peut se faire soit en sélectionnant initialement tous les atlas, et en écartant itérativement les moins pertinents [Langerak 2009, Langerak 2010], soit au contraire en sélectionnant initialement un très faible nombre d'atlas, et en ajoutant itérativement les plus pertinents [van Rikxoort 2010, van Rikxoort 2008].

Dans [van Rikxoort 2008, van Rikxoort 2010], la segmentation initiale est estimée en propageant les labels d'un atlas de la base pris au hasard [van Rikxoort 2008] ou de façon optimale [van Rikxoort 2010], ainsi que pour chaque région les labels de l'atlas le plus similaire au patient sur cette région (soit une initialisation avec deux atlas propagés par région). À une itération donnée, étant donné une région et les segmentations d'un certain nombre d'atlas déjà propagées sur cette région, on considère alors que l'utilisation d'un atlas supplémentaire serait susceptible d'apporter une amélioration pour cette région si le pourcentage dans cette région de voxels dits *indécis* est supérieur à un certain seuil. Les voxels *indécis* sont ici définis comme les

voxels ayant été classifiés comme appartenant à la structure dans la moitié des labels déjà propagés, et comme appartenant au fond dans le reste des labels déjà propagés. Pour chaque région, des atlas supplémentaires sont ainsi itérativement sélectionnés pour affiner le calcul de la segmentation du patient jusqu'à ce que le pourcentage de voxels indécis soit inférieur au seuil fixé. Le seuil utilisé par les auteurs est de 8% pour le cœur [van Rikxoort 2008, van Rikxoort 2010], ce qui représente en moyenne quatre atlas sélectionnés par région. Appliqué globalement et non par région pour le noyau caudé, le seuil utilisé est de 20% [van Rikxoort 2010], ce qui correspond aussi à en moyenne 4 atlas sélectionnés (la structure étant plus petite que le cœur, elle a davantage de voxels situés à la frontière et donc potentiellement indécis).

Dans [Langerak 2009, Langerak 2010], la segmentation est initialisée par une méthode quelconque (vote par majorité éventuellement pondéré) à partir des labels propagés provenant de l'ensemble des atlas. Cette segmentation est ensuite améliorée d'itération en itération en écartant un nombre croissant d'atlas parmi les moins performants (c'est-à-dire ceux dont les labels propagés sont les moins proches de l'estimation courante de la segmentation au sens d'une mesure de chevauchement). Ce filtrage progressif des atlas utilisés peut être régi de différentes façons. Dans [Langerak 2009], un seuillage sur les performances des atlas permet de garder à chaque itération les atlas plus performants. Ce seuil est fonction de la performance maximale des atlas, et est incrémenté linéairement à chaque itération de façon à ne pas écarter des atlas trop rapidement dans le processus. Dans [Langerak 2010], le principe est le même mais à partir d'un certain nombre d'itérations, les atlas écartés ne sont plus reconsidérés ce qui permet d'assurer la convergence de l'algorithme.

5.3 Évaluation non biaisée des méthodes de sélection

Nous avons présenté dans la section précédente différentes méthodes de sélection automatique d'atlas que nous souhaitons maintenant évaluer. Notre objectif est double : d'une part nous souhaitons les comparer entre elles de façon à pouvoir déterminer dans quelle mesure certaines d'entre elles sont équivalentes. D'autre part, nous voulons aussi les évaluer à proprement parler, c'est-à-dire déterminer celle(s) qui fournira(ont) *in fine* les segmentations automatiques les plus proches des segmentations manuelles.

Dans la littérature, les méthodes de sélection d'atlas ne sont en général pas évaluées indépendamment du reste de l'algorithme. L'approche la plus courante consiste en effet à évaluer l'algorithme de segmentation dans son ensemble en comparant les segmentations automatiques obtenues avec les segmentations manuelles. Cela ne permet pas d'évaluer la méthode de sélection d'atlas indépendamment de la méthode utilisée pour fusionner les atlas sélectionnés, ainsi que du nombre d'atlas sélectionnés.

Nous proposons dans ce paragraphe un protocole permettant d'évaluer différentes méthodes de sélection d'atlas de façon non biaisée, c'est-à-dire en s'affranchissant de l'étape de fusion et du nombre d'atlas fusionnés. Notre approche est

basée sur le partitionnement des méthodes de sélection à partir des corrélations entre les classements d'atlas qu'elles fournissent, du plus adapté au moins adapté au patient. Nous présentons les mesures de corrélation entre méthodes de sélection en section 5.3.1, l'algorithme de partitionnement utilisé en section 5.3.2, et la mise en œuvre de l'algorithme dans sa globalité en section 5.3.3.

5.3.1 Corrélation moyenne de classement entre méthodes de sélection

Soit un patient à segmenter. Chaque méthode de sélection fournit un classement des atlas disponibles du plus adapté au moins adapté au patient en question. Étant donné deux méthodes de sélection d'atlas, nous proposons de les comparer en quantifiant la corrélation entre les classements d'atlas qu'elles fournissent. Pour quantifier cette corrélation, nous utilisons le coefficient ρ de Spearman qui est défini de la façon suivante :

$$\rho = 1 - \frac{6 \sum_{i=1}^k d_i^2}{k(k^2 - 1)} \quad (5.4)$$

où k est le nombre d'atlas dans les classements et d_i représente l'écart entre les rangs de l'atlas i dans les deux classements comparés. Ce coefficient vaut 1 si les deux classements sont les mêmes, -1 si les classements sont parfaitement inversés, et 0 si les classements sont complètement décorrélés. Nous aurions également pu utiliser le coefficient de Pearson, et calculer alors une corrélation entre les valeurs plutôt qu'entre les rangs. L'inconvénient principal de cette solution est que le coefficient de Pearson permet uniquement de prendre en compte les dépendances de type affine entre les deux listes de valeurs. L'avantage du coefficient de Spearman par rapport au coefficient de Pearson est qu'il permet de prendre en compte des dépendances de type non-linéaires entre les deux classements comparés.

Afin d'estimer un coefficient de corrélation moyenne pour chaque couple de méthodes de sélection, nous utilisons un schéma de type Leave-One-Out : chaque patient P de la base est exclu successivement, et les patients restants sont considérés comme des atlas potentiels, et classés selon chacune des méthode de sélection du plus adapté au moins adapté au patient exclu. A partir de ces classements, nous calculons un coefficient de Spearman noté $\rho(P, M_i, M_j)$ pour chaque couple de méthodes de sélection (M_i, M_j) . Nous recommençons la procédure en excluant successivement chaque patient de la base. Pour chaque couple de méthodes de sélection (M_i, M_j) , nous calculons alors une corrélation moyenne à partir des coefficients de Spearman obtenus en excluant chaque patient successivement : $\bar{\rho}(M_i, M_j) = \frac{1}{N} \sum_{n=1}^N \rho(I_n, M_i, M_j)$.

5.3.2 Algorithme de partitionnement par affinité-propagation

Étant donné un ensemble de méthodes de sélection d'atlas, nous proposons de les regrouper d'après leurs corrélations moyennes de classements deux à deux

$\bar{\rho}(M_i, M_j)$. Comme algorithme de partitionnement, nous utilisons l'algorithme d'affinité-propagation proposé par Frey et al. [Frey 2007]. De façon générale, étant donné un ensemble de points à partitionner, cet algorithme prend en entrée des valeurs de similarités deux à deux entre ces points, et identifie de façon itérative les sous-groupes optimaux permettant de maximiser les valeurs de similarités intra-groupes tout en minimisant les valeurs de similarités inter-groupes. De plus, cet algorithme fournit également l'exemplaire le plus représentatif pour chaque sous-groupe. Le principe de cet algorithme consiste à réévaluer à chaque itération la capacité de chaque point d'être un exemplaire (c'est-à-dire un centre de sous-groupe) d'après des échanges de « messages » entre les points. Les « messages » échangés sont de deux types : *responsabilités* et *disponibilités*. Les *responsabilités* sont envoyées par les points aux exemplaires potentiels, et caractérisent la préférence relative des points pour chaque exemplaire potentiel par rapport aux autres exemplaires potentiels. En d'autres termes, chaque point envoie à chaque exemplaire potentiel un message indiquant dans quelle mesure cet exemplaire serait un bon représentant pour lui. Les *disponibilités* sont envoyées par les exemplaires potentiels aux points, et traduisent la disponibilité de chaque exemplaire pour chaque point d'après les crédits qui lui sont apportés par les autres points. Après convergence, les messages échangés sont combinés pour décider quels points vont jouer le rôle d'exemplaires, et pour chaque point, à quel exemplaire il va se rapporter. Le nombre de sous-groupes est déterminé automatiquement par l'algorithme. Il dépend essentiellement des valeurs d'auto-similarités (c'est-à-dire les valeurs sur la diagonale de la matrice de similarité). L'auto-similarité d'un point caractérise la probabilité *a priori* que ce point soit son propre exemplaire. Dans notre cas, les points à partitionner sont les méthodes de sélection d'atlas, et nous utilisons comme valeur de similarité les coefficients de Spearman moyens $\bar{\rho}(M_i, M_j)$. Nous considérons qu'aucune des méthodes de sélection n'est *a priori* plus apte que les autres à être un exemplaire, donc nous leur attribuons toutes une même valeur commune d'auto-similarité. Plus cette valeur d'auto-similarité commune est grande, plus le nombre de sous-groupes sera élevé, chaque méthode de sélection tendant alors à être son propre exemplaire.

5.3.3 Algorithme d'évaluation non biaisée

Nous avons présenté la mise en œuvre d'un algorithme de partitionnement des méthodes de sélection d'atlas à partir des corrélations moyennes entre les classements d'atlas qu'elle fournissent. Nous montrons maintenant comment ceci peut être utilisé pour la comparaison et l'évaluation non biaisée des méthodes de sélection.

Tout d'abord, notre approche peut être utilisée pour comparer plusieurs méthodes de sélection automatique entre elles. Cela permet d'identifier des groupes de méthodes équivalentes, et pour chaque groupe, la méthode de sélection la plus représentative du groupe. Cela peut permettre de sélectionner un nombre restreint de méthodes de sélection pour une évaluation plus approfondie, par exemple pour l'évaluation de l'algorithme de segmentation multi-atlas dans son ensemble.

D'autre part, afin d'évaluer les méthodes de sélection automatique, nous propo-

sons d'appliquer l'algorithme d'évaluation proposé en incluant en plus des méthodes de sélection automatique des méthodes de sélection dites *de référence*. Alors que les méthodes de sélection automatique classent les atlas en fonction de leur similarité image au patient à segmenter, les méthodes de référence doivent classer les atlas en fonction de leur aptitude à segmenter correctement le patient dans un processus de segmentation par atlas classique, c'est-à-dire : recalage non-linéaire entre le patient à segmenter et l'atlas, déformation de la segmentation manuelle de l'atlas sur le patient, et comparaison de la segmentation automatique ainsi obtenue avec la segmentation manuelle du patient. Comme mesures d'évaluation de qualité de segmentations, nous pouvons utiliser toute mesure définie dans l'annexe B. Quand la sélection est globale, on utilise la moyenne sur toutes les structures de la mesure choisie. Ces mesures de référence nous donnent pour chaque patient de la base des classements de référence des atlas. Comme précédemment, nous pouvons alors calculer des coefficients de Spearman moyens pour chaque couple de méthodes qu'elles soient automatiques ou de référence. Lorsque nous appliquons l'algorithme de partitionnement, nous constatons alors que les méthodes de référence se réunissent ensemble dans un ou plusieurs sous-groupe(s). Afin d'identifier quel sous-groupe de méthodes automatiques est le plus corrélé au(x) sous-groupe(s) de référence, nous proposons de calculer des corrélations moyennes inter-groupes de la façon suivante. Étant donné les valeurs de similarités entre méthodes de sélection (coefficients de Spearman moyens), nous définissons la corrélation inter-groupe entre deux sous-groupes c_{auto} et c_{ref} comme la similarité moyenne calculée sur tous les couples de méthodes ($M_a \in c_{auto}, M_r \in c_{ref}$) :

$$\bar{\rho}(c_{auto}, c_{ref}) = \frac{1}{N_a \cdot N_r} \sum_{M_a \in c_{auto}} \sum_{M_r \in c_{ref}} \bar{\rho}(M_a, M_r) \quad (5.5)$$

où N_a et N_r représentent les nombres de méthodes de sélection dans les sous-groupes c_{auto} et c_{ref} . Le sous-groupe de méthodes de sélection automatique le plus pertinent sera celui qui à la meilleure corrélation inter-groupe avec le(s) sous-groupe(s) de référence.

L'algorithme ainsi décrit sera utilisé dans le chapitre suivant pour évaluer et comparer les méthodes de sélection d'atlas. Les résultats seront présentés en section 6.2.2.

5.4 Différentes stratégies pour la fusion des atlas sélectionnés

5.4.1 Préambule

Dans la section précédente, nous avons présenté des méthodes permettant de sélectionner les atlas les plus adaptés à un patient donné. Les informations provenant des atlas sélectionnés (segmentations et éventuellement images) doivent ensuite être combinées pour estimer la segmentation du patient. Cela nécessite d'une part leur

mise en correspondance et leur *fusion*, et d'autre part une étape de *propagation* des informations (avant ou après fusion) vers le référentiel du patient.

Nous classons les approches de la littérature selon deux catégories principales : les approches de type « fusion - propagation » et les approches de type « propagation - fusion ». Dans le premier type, les informations provenant des atlas sélectionnés sont fusionnées dans un référentiel externe, puis le résultat de cette fusion est propagé sur l'image du patient. Dans le second type, tous les atlas sélectionnés sont propagés sur l'image du patient et la phase de fusion a lieu dans le référentiel du patient.

Nous présentons dans cette section trois stratégies de fusion différentes :

- un schéma de type « fusion hors-ligne - propagation » : il s'agit de la construction de sous-atlas hors ligne dont le principe est illustré sur la figure 3.6 (cf. section 5.4.2) ;
- un schéma de type « fusion à la volée - propagation » : il s'agit de la construction à la volée d'un atlas adapté au patient dont le principe est illustré sur la figure 3.8 (cf. section 5.4.3) ;
- les différentes mises en œuvre envisageables pour le schéma de type « propagation - fusion » : il s'agit de la fusion directe des segmentations des atlas sélectionnés dans le référentiel du patient (cf. section 5.4.4).

5.4.2 Construction hors-ligne de sous-atlas

5.4.2.1 Principe

Nous présentons dans ce paragraphe une approche que l'on peut classer dans la catégorie « fusion - propagation ». Cette approche est basée sur les étapes suivantes :

1. **Fusion** hors-ligne :
 - Partitionner la base d'atlas en sous-groupes homogènes selon un critère,
 - Construire un atlas représentatif pour chaque sous-groupe.
2. A la volée, pour un patient donné à segmenter :
 - **Sélectionner** l'atlas représentatif du sous-groupe adapté au patient,
 - **Propager** la segmentation de l'atlas sélectionné sur le patient.

Cette approche a deux particularités par rapport au reste des algorithmes présentés dans ce chapitre. D'une part, la fusion d'atlas se fait hors-ligne. D'autre part, la sélection d'atlas se fait alors non plus parmi l'ensemble des atlas initiaux mais à partir d'un nombre réduit de sous-atlas représentatifs de la base initiale et construits à partir de sous-groupes de cette base. Ces sous-atlas représentatifs peuvent être des images particulières de la base sélectionnées comme telles ou des atlas moyens construits pour chaque sous-groupe. Cette approche repose sur l'hypothèse selon laquelle la base d'atlas peut être représentée par un nombre restreint d'atlas représentatifs, chacun d'entre eux correspondant à un mode de variation précis. Ce principe est celui de la *stratification d'atlas*, introduit par [Blezek 2007].

5.4.2.2 État de l'art sur la construction hors-ligne de sous-atlas

Le nombre de modes présents dans la base d'atlas ainsi que la répartition des atlas dans chaque mode peuvent être déterminés par des algorithmes de partitionnement appliqués à partir d'information image. Par exemple, Blezek et al. utilisent un algorithme de type mean-shift pour regrouper les atlas disponibles en sous-groupes homogènes [Blezek 2007]. Un atlas moyen est alors calculé pour chaque sous-groupe, mais la sélection de l'atlas le plus approprié pour un patient donné n'est pas traitée.

Les sous-groupes peuvent aussi être définis selon des méta-données. Par exemple, un atlas moyen est construit par catégorie d'âge pour la pédiatrie dans [Fonov 2011].

Shi et al. ont également proposé un algorithme de type « fusion hors-ligne - propagation » [Shi 2010]. A la différence des deux approches citées précédemment, leur méthode est régionale et non globale, avec un certain nombre d'atlas représentatifs pour chaque région. Dans un premier temps, une image moyenne est construite à partir de l'ensemble des atlas disponibles et parcellisée en régions. Les atlas de la base sont alors normalisés dans ce référentiel, et un partitionnement de type affinité-propagation [Frey 2007] leur est appliqué région par région en utilisant des mesures de similarité basées intensité. Pour chaque région, chaque sous-groupe est alors représenté par le morceau d'atlas le plus représentatif sur cette région (appelé exemplaire régional) et une carte de probabilité moyenne est calculée à partir de l'ensemble des cartes de probabilités des morceaux d'atlas appartenant à ce sous-groupe. A l'issue de cette étape de fusion régionale hors-ligne, la segmentation d'un nouveau patient nécessite dans un premier temps sa normalisation dans le référentiel de l'image moyenne. Ensuite, pour chaque région, (1) le sous-groupe le plus approprié au patient est sélectionné grâce à des mesures de similarité basées intensités entre le patient normalisé et les exemplaires régionaux de cette région, et (2) la carte de probabilité moyenne du sous-groupe ainsi sélectionné est choisie pour cette région. Les cartes de probabilité moyennes sélectionnées sur chaque région sont ensuite juxtaposées et l'ensemble est lissé par un filtrage gaussien pour assurer une continuité entre régions. Enfin, l'atlas probabiliste adapté au patient ainsi créé est propagé sur l'image du patient par le biais d'un recalage non-linéaire. A noter que dans cet algorithme, seule une segmentation probabiliste adaptée au patient est construite, l'image moyenne utilisée pour la propagation finale est l'image moyenne initiale. Cependant, ce recalage final utilise un algorithme itératif de recalage-segmentation qui permet à la segmentation probabiliste de guider le recalage.

5.4.2.3 Algorithme proposé

Nous proposons de mettre en œuvre une méthode de construction hors-ligne de sous-atlas adaptée à la segmentation des structures ORL pour la radiothérapie. Une façon simple et objective de partitionner notre base d'atlas est d'utiliser le critère présenté en section 5.2.5.1, c'est-à-dire la localisation et le stade de la tumeur primitive. Ce critère est bien adapté à la construction hors-ligne de sous-atlas car la sélection du sous-atlas approprié pour un nouveau patient peut se faire de façon immédiate et objective. Afin de ne pas obtenir des sous-groupes trop petits, nous

choisissons de symétriser la base d'atlas avant de la partitionner. Ensuite, une image moyenne est construite pour chaque sous-groupe avec l'algorithme de Guimond et al. [Guimond 2000], et la segmentation moyenne correspondante est calculée avec l'algorithme de vote par majorité lissé présenté en section 4.5.2.2. Pour chaque nouveau patient, une segmentation par atlas classique est alors appliquée en utilisant l'atlas moyen correspondant à la localisation et au stade de tumeur du patient.

5.4.3 Création à la volée d'atlas adaptés au patient

5.4.3.1 Motivations

L'approche de type « fusion hors-ligne - propagation » présentée dans le paragraphe précédent suppose que le patient est suffisamment similaire à l'un des atlas représentatifs pré-calculés à partir de la base. Si le critère de sélection est de type méta-information (comme la localisation et le stade de la tumeur primitive, ou encore l'âge), cette hypothèse est vérifiée dès lors que nos atlas représentatifs couvrent la totalité des cas possibles (par exemple toutes les combinaisons localisation/stade possibles, ou toutes les tranches d'âges possibles). Cependant, si le critère de sélection est basé sur des informations image (intensité ou déformation), la validité de cette hypothèse n'est pas garantie, en particulier si le nombre d'atlas représentatifs auxquels on se restreint est faible par rapport à la variabilité anatomique. La création à la volée d'atlas adaptés au patient permet d'être plus flexible et de dépasser cette limitation. La fusion a cette fois-ci lieu à la volée et non plus hors-ligne. Alors que l'utilisation des sous-atlas construits hors-ligne restreint fortement le choix de l'atlas le plus adapté au patient, la fusion à la volée permet au contraire d'exploiter individuellement les atlas de la base initiale pour façonner un patient *virtuel* (ou *composite*) qui est régionalement ou localement adapté au patient à segmenter.

5.4.3.2 État de l'art sur la création à la volée d'atlas adaptés au patient

Dans la littérature, peu d'articles se sont intéressés à la création à la volée d'un atlas adapté au patient. En effet, la grande majorité des publications utilisent l'approche de « propagation - fusion » que nous présenterons en section 5.4.4. Nous présentons dans ce paragraphe l'algorithme proposé par Commowick et al. [Commowick 2009], et introduit comme une analogie à la créature de Frankenstein. Cet algorithme permet de construire un patient virtuel régionalement similaire au patient P à segmenter en suivant les étapes ci-dessous :

1. Recalage non-linéaire du patient P sur l'atlas moyen M pré-calculé,
2. Pour chaque région R_i définie sur M , sélection de l'atlas de la base le plus similaire à P selon un critère basé déformation (c'est-à-dire [Commowick 2007a]),
3. Construction d'un patient composite \tilde{M} en fusionnant à la volée les morceaux d'images sélectionnés,
4. Construction de la segmentation du patient composite à partir des segmentations des atlas sélectionnés.

L'algorithme de construction du patient composite est une généralisation de l'algorithme itératif de Guimond et al. [Guimond 2000] décrit en section 3.4.1.2. Alors que l'algorithme de Guimond donne des poids globaux (et égaux) à toutes les images dans la moyenne, cette généralisation permet de leur donner des poids variant spatialement. Afin d'assurer une transition lisse entre les régions R_l , une distance minimale est assurée entre deux régions adjacentes, et les poids donnés à chaque image dans les calculs de moyenne sont définis de façon continue en prenant en compte pour chaque voxel la distance à chacune des régions. A chaque itération, l'utilisation du cadre log-euclidien introduit dans [Arsigny 2006a] pour le calcul de la transformation moyenne permet de s'assurer que la transformation estimée est bien un difféomorphisme. En effet, ceci n'est pas garanti en utilisant la moyenne euclidienne classique à cause des poids variant spatialement. Enfin, les segmentations des atlas sélectionnés sont déformées sur le patient composite, puis moyennées avec les poids spatiaux pour créer des cartes de probabilité qui sont ensuite utilisées pour classifier les voxels. Ceci permet d'assurer des contours disjoints pour les différentes structures.

5.4.3.3 Algorithme proposé

L'algorithme présenté dans la section précédente permet de créer un patient composite régionalement similaire au patient à segmenter [Commowick 2009]. Cette approche est intéressante mais elle comporte des limitations, liées au fait qu'un seul atlas est sélectionné sur chaque région pour construire le patient composite. Cela rend en effet l'approche très sensible à l'étape de sélection, et peu robuste à d'éventuelles erreurs de sélection. De plus, comme une seule segmentation est sélectionnée pour chaque structure, l'approche est également très sensible à la variabilité de contourage intra- et inter-expert. Afin de dépasser ces limitations, nous proposons de généraliser cet algorithme à la sélection de plusieurs atlas par région. Les paragraphes suivants présentent les différentes étapes de construction du patient composite. A la fin, la segmentation automatique du patient est obtenue en utilisant le patient composite ainsi créé dans un schéma de segmentation par atlas classique.

Sélection et pondération des atlas On suppose ici qu'on dispose d'une mesure d permettant de définir pour chaque région R_l une distance $d_{R_l}(I_k, P)$ entre le patient P et chaque atlas I_k de la base. Cette distance peut être basée sur une mesure de similarité de type intensité ou déformation. Par exemple, on peut utiliser la distance basée sur les jacobiens proposée dans la section 5.2.5.2 (cf. équation 5.3).

Pour chaque région R_l , nous utilisons les distances $\{d_{R_l}(I_k, P)\}_{k \in [1 \dots N]}$ non seulement pour sélectionner les K_l atlas les plus proches du patient sur cette région (ceux dont la distance est la plus petite), mais aussi pour les pondérer. Les K_l atlas sélectionnés sur la région R_l sont notés $\{\hat{I}_{l,n}\}_{n \in [1 \dots K_l]}$, et nous définissons leurs poids respectifs dans la moyenne par un système de pondération gaussien basé sur les distances $\{d_{R_l}(\hat{I}_{l,n}, P)\}_{n \in [1 \dots K_l]}$, c'est-à-dire :

$$\alpha_{l,n} = G_{\mu,\sigma}(d_{R_l}(\hat{I}_{l,n}, P)) \quad (5.6)$$

Ces poids permettent de donner davantage d'importance dans la moyenne aux atlas sélectionnés qui sont très proches du patient, et au contraire de discriminer ceux qui bien que sélectionnés sont peu similaires au patient. Bien entendu, l'écart-type σ et la valeur centrale μ de la gaussienne doivent être choisis de façon appropriée. La gaussienne peut être centrée soit en zéro, soit sur la distance minimale parmi les distances des atlas sélectionnés. Nous avons choisi cette seconde solution afin de donner le poids le plus élevé possible à l'atlas le plus similaire au patient. Les poids sont ensuite renormalisés de telle sorte que pour chaque région R_l , $\sum_{n=1}^{K_l} \bar{\alpha}_{l,n} = 1$.

Construction du patient composite La construction du patient composite consiste à moyennner les atlas $\{\hat{I}_{l,n}\}_{n \in [1 \dots K_l]}$ sélectionnés sur chaque région R_l , tout en assurant une certaine continuité au niveau des frontières entre les régions lors de l'interpolation. Comme proposé dans [Commowick 2009], cette continuité est obtenue en assurant une distance minimale entre les régions, puis en définissant des poids variant spatialement. Plus précisément, le poids spatial donné à la région R_l au voxel x est défini par :

$$w_l(x) = 1/(1 + \beta \text{dist}(x, R_l)) \quad (5.7)$$

où $\text{dist}(x, R_l)$ représente la distance minimale à la région R_l au voxel x . Une renormalisation permet d'obtenir $\sum_{l=1}^L \bar{w}_l(x) = 1$.

La construction du patient composite régionalement similaire au patient P peut être vue comme une construction d'image moyenne classique (c'est-à-dire du type [Guimond 2000]) mais dans laquelle chaque image est associée à une carte de poids variant spatialement. Ces cartes de poids dépendent à la fois des poids de sélection $\bar{\alpha}_{l,n}$ et des poids spatiaux $\bar{w}_l(x)$ destinés à assurer la continuité aux frontières. Le patient composite \tilde{M} est initialisé avec l'atlas moyen M (construit selon [Guimond 2000] à partir de la totalité des atlas), c'est-à-dire $\tilde{M}_0 = M$, puis les étapes suivantes sont itérées :

1. Recaler les images $\hat{I}_{l,n}$ sur l'estimation courante du patient composite \tilde{M}_t . Cette étape fournit les transformations $T_{\hat{I}_{l,n} \leftarrow \tilde{M}_t} = T_{\hat{I}_{l,n} \leftarrow \tilde{M}_t}^{aff} \circ T_{\hat{I}_{l,n} \leftarrow \tilde{M}_t}^{nl}$.
2. Calculer l'image d'intensité moyenne M_{t+1} en interpolant les intensités des images $\hat{I}_{l,n}$ déformées sur \tilde{M}_t et en utilisant les deux systèmes de poids $\bar{w}_{l,t}(x)$ et $\bar{\alpha}_{l,n}$:

$$M_{t+1}(x) = \sum_{l=1}^L \left[\bar{w}_{l,t}(x) \left(\sum_{n=1}^{K_l} \bar{\alpha}_{l,n} \left(\hat{I}_{l,n} \circ T_{\hat{I}_{l,n} \leftarrow \tilde{M}_t} \right) (x) \right) \right] \quad (5.8)$$

3. Calculer la moyenne \bar{T}_t des déformations non-linéaires $T_{\hat{I}_{l,n} \leftarrow \tilde{M}_t}^{nl}$ en utilisant toujours les poids $\bar{w}_{l,t}(x)$ et $\bar{\alpha}_{l,n}$. L'utilisation du cadre log-euclidien [Arsigny 2006b] est nécessaire pour assurer que la transformation estimée est bien un difféomorphisme. Le calcul de \bar{T}_t se fait alors en intégrant entre le

temps 0 ($x(0) = x$) et le temps 1 ($x(1) = \bar{T}_t(x)$) l'équation différentielle suivante :

$$\dot{x} = \sum_{l=1}^L \left[\bar{w}_{l,t}(x) \left(\sum_{n=1}^{K_l} \bar{\alpha}_{l,n} \log \left(T_{\hat{I}_{l,n} \leftarrow \tilde{M}_t}^{nl} \right) (x) \right) \right] \quad (5.9)$$

4. Appliquer \bar{T}_t^{-1} à M_{t+1} pour obtenir la nouvelle estimation du patient composite $\tilde{M}_{t+1} = M_{t+1} \circ \bar{T}_t^{-1}$.
5. Mettre à jour les régions d'intérêt en appliquant \bar{T}_t^{-1} aux $R_{l,t}$: $R_{l,t+1} = R_{l,t} \circ \bar{T}_t^{-1}$, et recalculer les poids spatiaux $\bar{w}_{l,t+1}(x)$ correspondants.

Ce processus est proche de celui proposé par [Commowick 2009], mais il est plus général car il permet la fusion de plusieurs atlas par région : alors que l'algorithme de [Commowick 2009] construit un patient composite, nous construisons ici un patient composite moyen. Dans l'équation 5.8 (et de façon analogue dans l'équation 5.9), la somme interne (somme sur n) correspond à une moyenne pondérée des images sélectionnées pour la région R_l , tandis que la somme externe utilise les poids spatiaux $\bar{w}_{l,t}(x)$ pour combiner les contributions des atlas sélectionnés sur les différentes régions. Dans [Commowick 2009], la somme interne est absente puisque $K_l = 1$ pour chaque région R_l .

Construction de la segmentation associée Une fois que l'image du patient composite \tilde{M} a été construite, les segmentations manuelles des atlas sélectionnés sur les différentes régions peuvent alors être normalisées dans ce référentiel. Pour fusionner ces segmentations normalisées en une segmentation moyenne tout en assurant une continuité entre les différentes structures, nous appliquons l'algorithme de vote par majorité lissé présenté en section 4.5.2.2 en l'adaptant au cas présent. Lors de la construction de \tilde{M} , nous avons pris soin de définir une région R_l par structure, ce qui nous permet de calculer facilement et de façon indépendante la carte de probabilité de chacune des structures. Pour calculer la carte de probabilité de la structure s , nous nous restreignons aux segmentations des atlas $\{\hat{I}_{s,n}\}_{n \in [1 \dots K_s]}$ et nous les moyennons en utilisant les poids de sélection $\bar{\alpha}_{l,n}$. Il aurait également été possible de considérer les segmentations de la structure s de l'ensemble des atlas sélectionnés sur toutes les régions, et d'utiliser alors les poids spatiaux $\bar{w}_{l,t}(x)$ en supplément des poids $\bar{\alpha}_{l,n}$. Nous avons cependant choisi la première solution. A noter que lors de la sélection des K_s atlas les plus adaptés sur la région R_s , nous nous étions restreint aux atlas ayant la structure s segmentée. Cela permet d'éviter le cas où aucune segmentation n'est disponible pour le calcul de la carte de probabilité. Les cartes de probabilités ainsi obtenues pour les différentes structures sont ensuite lissées et seuillées. Enfin, les jonctions entre structures sont gérées par post-traitement de morphologie mathématique (SKIZ). Les détails sont présentés en section 4.5.2.2.

5.4.4 Fusion des segmentations dans le référentiel du patient

L'approche présentée dans le paragraphe précédent consiste à construire à la volée un atlas adapté au patient (image et segmentation) puis à propager la seg-

mentation de cet atlas sur le patient. Le principal inconvénient de cette approche est qu'elle est très coûteuse en temps de calcul, en particulier pour la construction de l'image du patient composite moyen.

Dans la littérature, la majorité des auteurs utilisent l'approche que nous avons appelée « propagation - fusion » : les segmentations des atlas sélectionnés sont propagées de façon non-linéaire dans le référentiel du patient à segmenter, puis fusionnées dans ce référentiel, ce qui fournit la segmentation automatique du patient. Les éléments qui divergent d'une méthode à l'autre sont les suivants :

- la façon dont se fait la propagation des segmentations des atlas sur le patient (directe ou indirecte en passant par un référentiel externe),
- la règle utilisée pour la fusion des segmentations dans le référentiel du patient.

Les différentes possibilités pour chacun de ces deux éléments sont présentées en sections 5.4.4.1 et 5.4.4.2, puis évaluées de façon approfondie dans le chapitre 6 (en section 6.3).

5.4.4.1 Propagation directe ou indirecte

Principe et état de l'art La propagation directe nécessite le recalage non-linéaire de l'ensemble des atlas sélectionnés sur le patient à segmenter. Son principe est illustré sur la figure 3.9. La propagation indirecte au contraire utilise un référentiel externe intermédiaire comme point de passage. Cela permet de réduire considérablement le temps de calcul puisqu'alors un seul recalage non-linéaire est nécessaire, celui du patient à segmenter avec le référentiel externe (les recalages des atlas de la base sur le référentiel externe pouvant être fait hors-ligne). Nous abordons dans la suite de ce paragraphe les avantages et inconvénients potentiels de chacune des approches en terme de précision de segmentation.

Dans la littérature, seuls Heckemann et al. ont comparé les deux types de propagation à notre connaissance [Heckemann 2006]. Ils ont reporté de meilleurs résultats avec la propagation directe qu'avec la propagation indirecte. Cependant, dans leur cas, le référentiel intermédiaire utilisé est une image particulière de la base d'atlas. Une façon d'optimiser les résultats de la fusion indirecte est d'utiliser comme référentiel intermédiaire une image moyenne construite à partir de l'ensemble des atlas de la base (par exemple avec [Guimond 2000]). Cela permet de réduire le biais lié à l'introduction d'une anatomie particulière, et donc d'optimiser en moyenne sur la base les recalages non-linéaires entre les atlas et le référentiel externe. Dans la suite, nous nous plaçons dans ce cas de figure optimal de propagation indirecte.

Avantages potentiels de la propagation directe Même dans le cas optimal où le référentiel intermédiaire est un atlas moyen, la propagation indirecte possède à première vue un certain nombre d'inconvénients par rapport à la propagation directe. Tout d'abord, elle introduit un biais dans la mise en correspondance des atlas avec le patient à segmenter. Dans le cas où le patient à segmenter à une anatomie atypique (c'est-à-dire très différente de l'anatomie moyenne) et où on suppose que le processus de sélection d'atlas a réussi à identifier les atlas de la base ayant cette même anatomie

atypique, les recalage non-linéaires du patient et des atlas sélectionnés avec l'atlas moyen ne seront pas de qualité optimale à cause des différences trop importantes d'anatomie. Au contraire, la propagation directe fournira des recalages de qualité bien meilleure. On peut donc imaginer que pour les patients ayant une anatomie atypique, la propagation directe sera significativement plus précise. En revanche, pour les patients ayant une anatomie proche de la moyenne, on peut s'attendre à ce que les différences entre les deux types de propagation soient relativement faibles.

Par ailleurs, avec la propagation indirecte, les labels des atlas propagés sur l'image du patient ne peuvent plus réellement être considérés comme étant indépendants à cause du biais introduit lors du recalage indirect. Ceci va à l'encontre du principe même de la segmentation multi-atlas puisque celle-ci a justement pour objectif d'obtenir plusieurs segmentations indépendantes pour que les erreurs locales introduites par chacune d'entre elles se compensent.

Enfin, la propagation indirecte a l'inconvénient de cumuler les erreurs d'estimation de transformation car elle nécessite deux recalages non-linéaires au lieu d'un.

Dans la littérature, la majorité des méthodes utilisent la propagation directe [Gorghi 2010, Wu 2007, Rohlfing 2004, Artaechevarria 2009, Isgum 2009, Langerak 2010, Teng 2007, van Rikxoort 2009, Aljabar 2009, Heckemann 2009, Teguh 2010, Han 2010]. Van Rikxoort et al. utilisent la propagation directe après sélection régionale, et seules les régions sélectionnées dans un atlas donné sont recalées ce qui permet d'accélérer le recalage [van Rikxoort 2008, van Rikxoort 2010].

Avantages potentiels de la propagation indirecte Malgré tout, la propagation indirecte peut potentiellement améliorer la précision de la mise en correspondance dans certaines configurations particulières. En effet, lorsque l'atlas sélectionné et l'image du patient ont des anatomies relativement éloignées de la moyenne mais également assez différentes entre elles, l'utilisation de l'atlas moyen comme référentiel intermédiaire peut potentiellement améliorer les résultats. Par exemple, si l'atlas représente un patient ayant la colonne très droite, tandis que le patient a la colonne très fléchie, on constate que le recalage direct de ces deux images est en général très mauvais. Au contraire, le recalage non-linéaire de chacune de ces images avec l'atlas moyen est de qualité correcte, et la propagation indirecte permettra donc d'atteindre une meilleure qualité que la propagation directe. On fait le même constat pour les images représentant des patients corpulents ou maigres. Cela s'explique par le fait que lors du recalage non-linéaire par blocs (cf. annexe A), le voisinage de recherche est limité pour assurer la robustesse de l'algorithme, et donc les déformations d'amplitude supérieure à ce voisinage de recherche ne pourront jamais être récupérées en un seul recalage.

En principe, les méthodes de sélection d'atlas permettent de sélectionner les atlas les plus similaires au patient, et donc le cas de figure mentionné précédemment (par exemple patient corpulent et atlas correspondant à un patient maigre) est peu probable, sous réserve que la méthode de sélection soit pertinente. Cela dit, il nous est difficile d'estimer la fréquence réelle d'occurrence de ce cas de figure tant que

nous n'avons pas évalué la pertinence des méthodes de sélection. Si les méthodes de sélection s'avèrent peu robustes sur notre base d'images, il n'est pas à exclure que ce cas de figure se produise plus souvent qu'on ne le pense.

Dans la littérature, la propagation indirecte dans le cadre de la fusion de segmentations dans le référentiel du patient a été utilisée par [Aljabar 2007] avec une image IRM artificielle comme référentiel intermédiaire (MNI Brainweb). D'autre part, Zhuang et al. ont proposé une approche originale qui apporte des arguments en faveur de la propagation indirecte [Zhuang 2010a]. Leur approche est un peu différente de celle généralement utilisée (sélection, propagation et fusion d'atlas) puisqu'elle consiste à sélectionner non plus des atlas (images + segmentations) mais des chemins de propagation indirecte permettant de déformer un atlas moyen sur le patient en passant par une image de la base. Chaque chemin de propagation correspond au recalage de l'atlas moyen sur une des images de la base, puis du recalage de cette image de la base sur le patient à segmenter. Cette approche suggère que dans certaines configurations (par exemple patient à segmenter trop différent de l'atlas moyen), la propagation indirecte via décomposition en deux recalages successifs peut permettre d'améliorer les résultats.

Enfin, le dernier avantage de la propagation indirecte est que son coût en temps de calcul est très faible. Son utilisation peut donc permettre d'exploiter une très grande base d'atlas et donc d'atteindre en principe de meilleures performances dès lors qu'une sélection d'atlas est effectuée.

Synthèse et discussion Outre l'avantage évident en terme de temps de calcul en faveur de la propagation indirecte, nous avons mis en évidence dans les deux paragraphes précédents que suivant la variabilité anatomique présente dans la base d'atlas et suivant la pertinence des méthodes de sélection d'atlas, il existe des configurations pour lesquelles la propagation directe aurait intuitivement tendance à fournir de meilleurs résultats, et inversement d'autres configurations qui seraient mieux gérées par la propagation indirecte. On s'attend à ce que la propagation directe soit plus précise que la propagation indirecte pour les atlas sélectionnés comme étant similaires au patient, sous réserve que la méthode de sélection d'atlas soit pertinente. Il est cependant difficile de prévoir lequel des deux types de propagation sera le plus précis en moyenne sur toute la base d'images et indépendamment de la méthode de sélection d'atlas.

Rohlfing et al. ont montré que dans leur base d'images (images de microscopie confocale de cerveau d'abeilles), la plupart des individus sont en moyenne plus proche de l'atlas moyen que de n'importe lequel des autres individus (ou en tout cas de la majorité des individus) (Annexe A, figure 15 de [Rohlfing 2004]). Dans notre cas, nous savons que la variabilité anatomique présente dans la région ORL est relativement élevée (corpulence, flexion de colonne). Nous sommes donc probablement dans une situation analogue à celle de [Rohlfing 2004], ce qui peut représenter un argument en faveur de la propagation indirecte. Néanmoins, aucune conclusion ne peut être tirée *a priori*.

Nous présentons dans le chapitre 6 (section 6.3) une évaluation comparative détaillée des deux types de propagation sur notre base d'images.

5.4.4.2 Algorithme utilisé pour la fusion de segmentations dans le référentiel du patient

Une fois que les segmentations des atlas sélectionnés ont été propagées dans le référentiel du patient, chacune d'entre elles représente alors un candidat potentiel pour la segmentation automatique du patient. Leur fusion permet de compenser les erreurs locales et par conséquent de fournir une meilleure estimation de la segmentation du patient.

Fusion dans le cadre d'une sélection globale ou régionale Lorsque les atlas ont été sélectionnés de façon globale ou régionale, la fusion de leurs segmentations correspond à la même problématique que celle traitée dans le chapitre 4 pour le calcul de la segmentation moyenne de l'atlas moyen.

Dans la littérature, la plupart des algorithmes utilisent un vote par majorité [Rohlfing 2004, van Rikxoort 2010, van Rikxoort 2008, Heckemann 2006, Aljabar 2007, Aljabar 2009, Heckemann 2009] (avec éventuellement un lissage de la carte de probabilité avant seuillage [Isgum 2009]), l'algorithme STAPLE [Han 2008, Leung 2010, Teguh 2010, Han 2010] ou d'autres algorithmes comme celui basé sur les cartes de distances [Rohlfing 2007] utilisé dans [Zhuang 2010a, Leung 2010]. Plusieurs études ont comparé les performances relatives de STAPLE et du vote par majorité pour la fusion de segmentations dans le référentiel du patient, et des conclusions différentes ont été tirées. Certaines ont obtenu de meilleurs résultats avec STAPLE [Leung 2010], mais d'autres ont obtenu de meilleurs résultats ou des différences non significatives avec le vote par majorité [Gorthi 2010, Klein 2008, Artaechevarria 2009, Zhuang 2010a]. Le choix de la méthode de fusion optimale semble donc être assez dépendant de la base d'images utilisée.

Dans la mise en œuvre de l'approche « propagation - fusion » après une sélection globale ou régionale, nous avons choisi d'utiliser l'algorithme de vote par majorité lissé présenté en section 4.5.2.2. En effet, cet algorithme s'est révélé être le plus performant pour le calcul de segmentation moyenne de l'atlas moyen parmi les algorithmes proposés et testés. Il s'est entre autre révélé plus performant que l'algorithme STAPLE et que le vote par majorité classique. La fusion de segmentations dans le cadre présent a lieu avec un plus faible nombre de données que dans le chapitre 4, donc les conclusions seraient peut être ici différentes. Malgré tout, nous choisissons cet algorithme pour deux raisons principales. Tout d'abord, il permet par nature de gérer les problèmes de continuité entre régions de sélection lorsque la sélection est régionale (grâce au SKIZ). De plus, le lissage gaussien des cartes de probabilité permet d'assurer des segmentations lisses, ce qui est d'autant plus nécessaire que le nombre de segmentations fusionnées est faible.

Fusion dans le cadre d'une pondération locale Comme mentionné dans la section 5.2.6.1, les approches de sélection globale et régionale ne permettent pas de prendre en compte des différences anatomiques à une échelle plus petite que celle de la structure. Pour dépasser cette limite, nous proposons dans ce paragraphe une méthode de pondération locale des segmentations des atlas dans le référentiel du patient. Cette méthode, proche de celle proposée par Artaechevarria et al. [Artaechevarria 2009], est décomposée selon les étapes suivantes :

1. Propagation (directe ou indirecte) des segmentations des atlas : on note \tilde{I}_k et \tilde{S}_k l'image et la segmentation de l'atlas k déformées sur le patient P .
2. Pour chaque structure s , et en notant K_s le nombre d'atlas ayant s segmentée :
 - Chaque atlas \tilde{I}_k est associé à une carte de poids locaux λ_k traduisant sa similarité locale en terme d'intensité avec l'image du patient P . Ces poids locaux sont définis voxel par voxel d'après les intensités dans le voisinage du voxel considéré :

$$\forall k \in [1 \dots K_s] \quad \lambda_k(x) = (\text{SIM}_{V(x)}(P, \tilde{I}_k))^p \quad (5.10)$$

où SIM désigne la mesure de similarité utilisée, $V(x)$ désigne le voisinage du voxel x , et p est un nombre entier permettant d'une part d'amplifier les différences entre les poids si nécessaire (c'est-à-dire si la mesure n'est pas suffisamment sensible), et d'autre part de prendre en compte le sens de variation de la mesure de similarité (c'est-à-dire $p > 0$ pour SIM = CC, et $p < 0$ pour SIM = SSD ou SIM = SAD).

- Calcul d'une carte de probabilité à partir des segmentations \tilde{S}_k et des cartes de poids λ_k associées :

$$p(x \in s) = \frac{1}{\sum_{k=1}^{K_s} \lambda_k(x)} \sum_{k=1}^{K_s} \lambda_k(x) \delta(\tilde{S}_k(x), s) \quad (5.11)$$

- Lissage éventuel de la carte de probabilité,
- Seuillage de la carte de probabilité,
- Fermeture morphologique éventuelle.

3. Gestion des chevauchements entre structures par SKIZ (cf. détails dans l'algorithme proposé en section 4.5.2.2).

Bien qu'un lissage intrinsèque soit introduit via la mesure de similarité calculée sur un certain voisinage, l'algorithme peut être malgré tout relativement sensible au bruit dans les images et au manque de contraste pour certaines structures. Cette sensibilité est accrue par la variabilité de contourage inter-expert qui fait que à même intensité une zone peut être incluse dans une segmentation mais pas dans une autre. Les étapes de lissage gaussien de la carte de probabilités et de fermeture morphologique sont optionnelles, mais elles permettent de limiter *a posteriori* l'influence de cette sensibilité sur la segmentation résultante.

5.5 Conclusion et synthèse des contributions

Dans ce chapitre, nous avons présenté un état de l'art approfondi des problématiques de sélection des atlas les plus adaptés au patient, et de leur fusion pour estimer la segmentation automatique du patient. Nous avons exposé différents schémas possibles pour chacune des phases (sélection et fusion) et évoqué les avantages et inconvénients de chacun d'entre eux.

Concernant la phase de sélection d'atlas, nous avons proposé :

- deux critères de sélection, l'un basé sur des méta-informations adaptées à la radiothérapie des tumeurs ORL (localisation et stade de tumeur), et l'autre basé sur des mesures de dilatation et de contraction de régions ;
- un protocole d'évaluation non biaisé des méthodes de sélection qui repose sur l'analyse des corrélations entre les classements d'atlas fournis par chacune d'entre elles.

Concernant la phase de fusion d'atlas, nous avons proposé :

- la construction hors-ligne de sous-atlas d'après la localisation et le stade de tumeur,
- un algorithme de construction à la volée d'un atlas composite moyen régionallement adapté au patient,
- un algorithme de pondération locale des segmentations des atlas dans le référentiel du patient.

Le chapitre 6 présente une évaluation détaillée de ces différentes contributions ainsi que de quelques unes des approches de la littérature évoquées dans ce chapitre.

Évaluation des algorithmes multi-atlas adaptés au patient pour la segmentation des structures ORL

Sommaire

6.1 Évaluation des sous-atlas par localisation et stade de tumeur	130
6.1.1 Partitionnement de la base d'images	130
6.1.2 Évaluation clinique	131
6.1.3 Évaluation sur notre base d'images	133
6.1.4 Synthèse et discussion	135
6.2 Évaluation des méthodes de sélection d'atlas basées image	136
6.2.1 Méthodes de sélection d'atlas étudiées	136
6.2.2 Évaluation d'après les classements d'atlas	138
6.2.3 Évaluation d'après les performances rang par rang	146
6.2.4 Synthèse et discussion	147
6.3 Évaluation des approches de fusion des segmentations dans le référentiel du patient	149
6.3.1 Évaluation dans le cadre d'une sélection globale ou régionale	151
6.3.2 Évaluation dans le cadre d'une pondération locale	156
6.3.3 Comparaison quantitative	158
6.3.4 Comparaison qualitative	160
6.4 Création à la volée d'un atlas composite adapté au patient	170
6.4.1 Paramètres	170
6.4.2 Résultats	170
6.5 Synthèse des contributions	175
6.6 Conclusion : limites et perspectives	179

Synthèse des contributions Dans ce chapitre, nous présentons une évaluation approfondie des algorithmes multi-atlas adaptés au patient présentés dans le chapitre précédent pour la segmentation des structures de la région ORL. Nous évaluons tout d'abord les sous-atlas par localisation et stade de la tumeur primitive en section 6.1. Dans la section 6.2, la phase de sélection d'atlas est évaluée selon les points suivants :

- Évaluation du critère de sélection proposé en section 5.2.5.2 (basé sur la contraction/dilatation des structures) ainsi que des principaux critères de sélection utilisés dans la littérature,
- Évaluation de l'ensemble des méthodes de sélection via le protocole de type partitionnement proposé en section 5.3,
- Choix de la méthode de sélection d'atlas optimale sur nos données.

Les sections 6.3 et 6.4 présentent l'évaluation de différentes approches présentées pour la phase de fusion d'atlas, à savoir :

- Évaluation des différentes approches de fusion des segmentations dans le référentiel du patient présentées en section 5.4.4,
- Évaluation de la méthode de création à la volée d'un atlas moyen composite régionalement adapté au patient proposée en section 5.4.3.3.

6.1 Évaluation des sous-atlas par localisation et stade de tumeur

Dans cette section, nous évaluons la construction hors-ligne et l'utilisation de sous-atlas caractéristiques de la localisation et du stade de la tumeur du patient à segmenter. Les détails de l'algorithme proposé sont présentés en section 5.4.2.3.

6.1.1 Partitionnement de la base d'images

La construction des sous-atlas nécessite dans un premier temps de partitionner notre base d'images selon la localisation et le stade de la tumeur primitive. Le tableau 6.1 illustre la répartition des images de notre base selon les différentes sous-catégories. Certaines sous-catégories sont clairement sous-représentées, comme la cavité orale N0-N1/N2, le nasopharynx N0-N1/N2, le larynx N2 et l'hypopharynx N2. Pour ces sous-catégories, la construction d'un sous-atlas est possible mais elle sera très sensible à la variabilité anatomique ainsi qu'à la variabilité de contourage en raison du faible nombre d'images et de segmentations moyennées. Pour cette raison, nous avons évalué la méthode sur un nombre restreint de sous-catégories pour lesquelles le nombre d'images est suffisant.

	N0-N1			N2		
	Gauche	Droit	Méd./Bil.	Gauche	Droit	Méd./Bil.
Cavité orale	2	3	0	1	2	0
Nasopharynx	0	2	1	0	0	0
Oropharynx	10	15	5	8	15	0
Hypopharynx	7	6	2	3	1	1
Larynx	8	5	4	1	1	2

FIG. 6.1 – Répartition des 105 images de notre base selon la localisation et le stade de la tumeur primitive. Méd./Bil. indiquent les tumeurs médianes ou bilatérales.

Les sous-catégories évaluées sont : oropharynx N0-N1 et oropharynx N2. Les images dont la tumeur est médiane ou bilatérale ne sont pas exploitées. De plus, comme le nombre d'images moyennées reste réduit, les images avec artefacts ou avec une anatomie très atypique (patient très corpulent ou très maigre, flexion de cou importante) ont un poids plus important dans la moyenne que pour la construction de l'atlas moyen global avec les 105 images. Afin de minimiser le biais causé par ces données atypiques, nous avons choisi de ne pas les prendre en compte. A l'issue de ce tri, puis de la symétrisation des images restantes, le nombre d'images disponibles est récapitulé pour chaque catégorie étudiée dans le tableau 6.2.

Sous-catégorie	Nombre d'images initial	Nombre d'images après suppression des images avec artefacts et anatomie atypique	Nombre d'images après symétrisation
Oropharynx N0-N1 gauche	10	9	21
Oropharynx N0-N1 droit	15	12	21
Oropharynx N2 gauche	8	6	17
Oropharynx N2 droit	15	11	17

FIG. 6.2 – Nombres d'images disponibles pour chaque sous-catégorie étudiée après tri et symétrisation.

6.1.2 Évaluation clinique

Des sous-atlas ont été construits pour les catégories oropharynx (droit et gauche) N0-N1 et N2 à partir des 21 et 17 images disponibles pour chacune de ces catégories. Pour la construction des segmentations moyennes, un seuil de 0.35 a été utilisé pour le seuillage des cartes de probabilité lissées (avec un filtrage gaussien d'écart-type 1). Cette valeur correspondait dans tous les cas au seuil optimal selon la mesure DICE sur notre base d'images.

Ces quatre sous-atlas ont ensuite fait l'objet d'une procédure d'évaluation clinique dans deux hôpitaux : le CHU de Tours et l'Institut Gustave Roussy (IGR).

L'étude menée par l'IGR à porté sur deux patients (un patient oropharynx N0-N1 et un patient oropharynx N2). Sur les deux patients étudiés, le radiothérapeute a rapporté des résultats qualitatifs très proches avec l'atlas moyen global et les sous-atlas adaptés. Certaines structures ont été mieux segmentées par l'un, et d'autres structures mieux segmentées par l'autre. Aucune conclusion en faveur de l'un ou l'autre des atlas n'a donc pu être tirée.

Concernant l'étude menée par le CHU de Tours, six patients ayant une tumeur primitive de l'oropharynx de stade N2 ont été sélectionnés pour l'évaluation. Pour chacun de ces six patients, les radiothérapeutes ont évalué qualitativement les

Chapitre 6. Évaluation des algorithmes multi-atlas adaptés au patient
132 **pour la segmentation des structures ORL**

contours automatiques obtenus avec l’atlas moyen global, ainsi que les contours automatiques obtenus avec le sous-atlas par localisation et stade de tumeur adéquat. En se basant sur cette évaluation qualitative, ils ont alors attribué pour chaque patient et pour chaque structure des scores allant de 1 (très mauvais) à 5 (excellent) à chacune des deux segmentations automatiques. De plus, les radiothérapeutes ont également contouré manuellement les images des six patients de façon à permettre une évaluation quantitative, par exemple avec l’indice de DICE. Le tableau 6.3 présente les résultats de cette étude clinique.

	Atlas moyen global		Sous-atlas par localisation et stade de tumeur	
	Score moyen	DICE moyen	Score moyen	DICE moyen
Patient 1	1.8 (médiocre)	0.62	3 (acceptable)	0.79
Patient 2	3.1 (acceptable)	0.74	3.6 (bon)	0.77
Patient 3	2.5 (moyen)	0.73	3 (acceptable)	0.78
Patient 4	2.85 (moyen)	0.49	3.3 (acceptable)	0.59
Patient 5	3.2 (acceptable)	0.79	3 (acceptable)	0.79
Patient 6	2.8 (moyen)	0.77	3.2 (acceptable)	0.80
Moyenne sur les 6 patients	2.7 ± 0.5	0.70 ± 0.15	3.2 ± 0.2	0.76 ± 0.09
Minimum et maximum sur les 6 patients	[1.8 ; 3.2]	[0.49 ; 0.79]	[3 ; 3.6]	[0.59 ; 0.80]

FIG. 6.3 – Résultats de l’évaluation clinique du CHU de Tours des sous-atlas par localisation et stade de tumeur. Les scores moyens et DICE moyens correspondent à une moyenne sur toutes les structures.

A l’exception du patient 5, les scores moyens et DICE moyens sur toutes les structures sont toujours meilleurs avec les sous-atlas par localisation et stade de tumeur qu’avec l’atlas moyen global. En moyenne sur tous les patients, l’utilisation des sous-atlas par localisation et stade de tumeur permet d’atteindre un score moyen de 3.2/5 et un DICE moyen de 0.76 (contre 2.7/5 et 0.70 respectivement avec l’atlas moyen global). De plus, on observe une certaine corrélation entre les scores moyens représentatifs de l’appréciation visuelle des radiothérapeutes et les DICE moyens. En effet, le patient pour lequel la différence entre les deux méthodes est la plus importante en terme de score (1.8 versus 3) correspond également à la différence la plus importante en terme de DICE (0.62 versus 0.79) (patient 1). De même, le patient pour lequel la différence entre les deux méthodes est la plus petite en terme de score (3.2 versus 3) correspond également à la différence la plus petite en terme de DICE (0.79 pour les deux) (patient 5).

6.1.3 Évaluation sur notre base d'images

Afin d'évaluer les sous-atlas par localisation et stade de tumeur sur davantage de patients, nous avons testé cette approche sur notre base d'images pour les catégories oropharynx N0-N1 et oropharynx N2.

6.1.3.1 Protocole d'évaluation

Étant donné le faible nombre d'images par catégorie, nous procédons à une évaluation par Leave-One-Out. Par exemple, pour chacune des 21 images originales de la sous-catégorie oropharynx N0-N1, nous excluons cette image de la base et construisons un atlas moyen à partir des 20 images restantes de la base symétrisée qui ont la tumeur du même côté. L'atlas moyen ainsi créé est alors utilisé dans un processus de segmentation par atlas classique pour segmenter l'image exclue. Les 21 images de la sous-catégorie oropharynx N0-N1 obtenues par symétrisation ne sont pas évaluées, elles servent uniquement à la construction d'atlas.

Pour chaque image exclue, le sous-atlas par localisation et stade de tumeur ainsi construit est évalué en comparant la segmentation automatique résultante à la segmentation manuelle, et ses performances sont comparées à celles obtenues avec des atlas moyens standards, à savoir :

- l'atlas moyen global construit à partir de la totalité de la base,
- des atlas moyens construits à partir du même nombre d'images que le sous-atlas par localisation et stade de tumeur, mais choisies aléatoirement dans la base, indépendamment de leur localisation et stade de tumeur. Cinq atlas moyens de ce type sont construits pour chaque image exclue, et leurs performances sont moyennées.

Dans tous les cas, l'image moyenne est calculée avec l'algorithme de Guimond et al. [Guimond 2000], et la segmentation moyenne est construite avec l'algorithme de vote par majorité lissé présenté en section 4.5.2.2, avec une valeur de seuil de 0.35 (valeur optimale sur notre base pour un filtrage gaussien d'écart-type 1).

6.1.3.2 Résultats

La figure 6.4 présente les résultats quantitatifs obtenus pour les patients de notre base ayant une tumeur de l'oropharynx de stade N0-N1 (colonne de gauche) ou N2 (colonne de droite). Les mesures de qualité utilisées sont l'indice de DICE (première ligne) et la distance de Hausdorff robuste $RHD_{95\%}$ (deuxième ligne). Dans chacun des quatre graphiques, nous avons présenté les moyennes sur tous les patients de la catégorie considérée des mesures suivantes :

- moyenne de DICE (ou $RHD_{95\%}$) sur toutes les structures,
- moyenne de DICE (ou $RHD_{95\%}$) sur les structures situées du même côté que la tumeur,
- moyenne de DICE (ou $RHD_{95\%}$) sur les structures situées du côté opposé à la tumeur.

Chapitre 6. Évaluation des algorithmes multi-atlas adaptés au patient 134 pour la segmentation des structures ORL

Les structures considérées sont les parotides, les sous-maxillaires et les aires ganglionnaires II, III et IV. La moelle épinière, le tronc cérébral ont été exclus de cette analyse car ce sont des structures relativement éloignées de la sphère ORL, et donc sur lesquelles la position et le stade de la tumeur ont un impact probablement négligeable. La mandibule n'est pas considérée non plus en raison du faible nombre de contours dans la base.

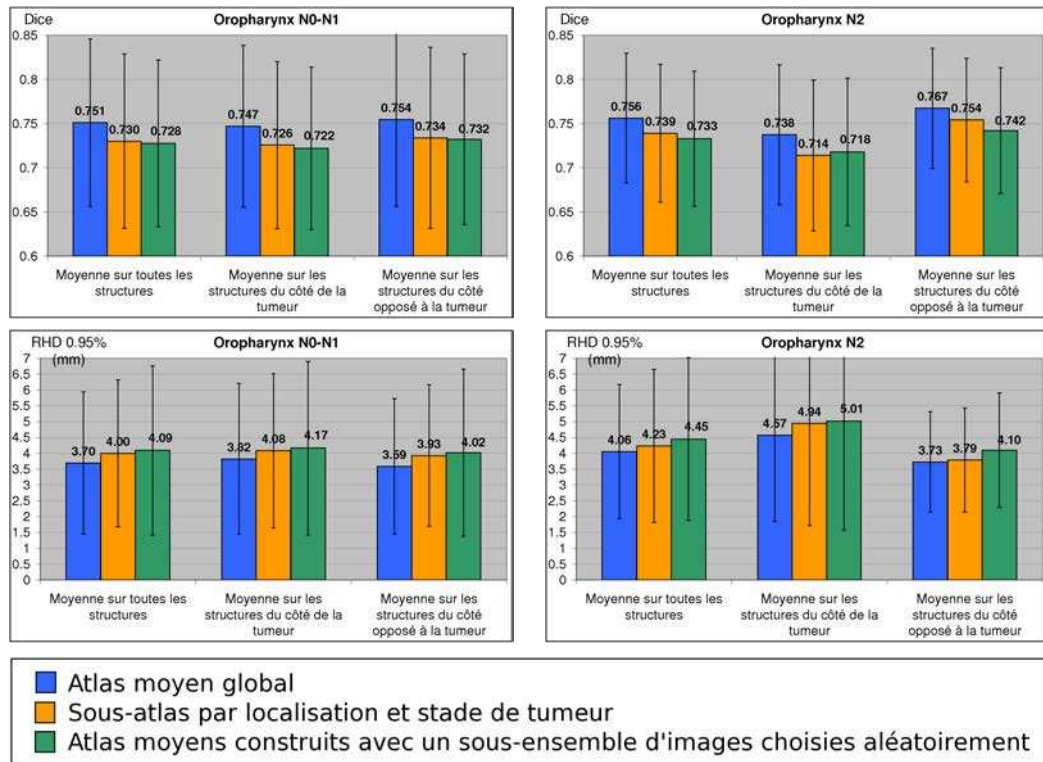


FIG. 6.4 – Résultats quantitatifs en terme de DICE (première ligne) et de distance de Hausdorff robuste (deuxième ligne) pour les patients de notre base ayant une tumeur de l'oropharynx de stade N0-N1 (colonne de gauche) et ceux ayant une tumeur de l'oropharynx de stade N2 (colonne de droite).

Tout d'abord, on remarque que dans tous les cas de figures (mesures/catégories/structures), l'atlas qui fournit les meilleurs résultats est l'atlas moyen global construit à partir de la totalité des patients (barres bleues). Les sous-atlas par localisation et stade de tumeur (barres oranges) sont moins performants, ce qui diffère des résultats obtenus en terme de DICE sur les 6 patients de l'évaluation clinique du CHU de Tours.

D'autre part, si l'on compare les sous-atlas par localisation et stade de tumeur (barres oranges) aux atlas moyens construits à partir du même nombre d'images mais choisies aléatoirement (barres vertes), les conclusions sont différentes suivant les structures considérées et suivant le stade (N0-N1 ou N2). En effet, pour les stades

N0-N1, les différences entre les deux approches ne sont pas significatives (p-values supérieures à 0.2 dans tous les cas). Pour le stade N2, les sous-atlas par localisation et stade fournissent des résultats significativement meilleurs pour les structures situées du côté opposé à la tumeur : DICE de 0.754 versus 0.742 (p-value= $7.8 \cdot 10^{-5}$), et $RHD_{95\%}$ de 3.79 mm versus 4.10 mm (p-value=0.001). En revanche, les différences ne sont pas significatives pour les structures situées du côté de la tumeur.

6.1.4 Synthèse et discussion

L'évaluation clinique et l'évaluation menée sur notre base d'images ont conduit à des conclusions différentes. Nous rappelons que ces deux évaluations portaient uniquement sur les patients ayant une tumeur de l'oropharynx. Tout d'abord, l'évaluation clinique menée par le CHU de Tours sur 6 patients de type oropharynx N2 a montré que les sous-atlas par localisation et stade de tumeur apportaient une amélioration significative par rapport à l'atlas moyen global. Cette amélioration a été quantifiée par des scores liés à l'appréciation visuelle des radiothérapeutes, ainsi que par les indices de DICE. L'évaluation clinique de l'IGR n'a quant à elle pas permis de mettre en avant l'une des deux méthodes par rapport à l'autre. Enfin, l'évaluation quantitative faite sur notre base a montré que les sous-atlas par localisation et stade de tumeur étaient significativement moins performants que l'atlas moyen global.

Plusieurs pistes peuvent permettre d'expliquer ces différences. Tout d'abord, les trois études ont été menées sur des patients différents, ce qui peut induire un biais surtout lorsque le nombre de patients évalués est faible (c'est le cas des deux évaluations cliniques). De plus, les deux évaluations cliniques ont été faites par des radiothérapeutes différents, ce qui induit un biais inter-expert dans le jugement qualitatif des contours. Concernant les conclusions divergentes en terme d'indice de DICE entre l'évaluation sur notre base et l'évaluation sur les 6 patients de Tours, une explication possible est un éventuel biais dans l'évaluation du CHU de Tours. En effet, il serait intéressant de savoir si les contours manuels ont été faits avec ou sans visualisation simultanée des contours automatiques. Comme les contours automatiques obtenus avec les sous-atlas ont reçu des meilleures appréciations visuelles que ceux obtenus avec l'atlas moyen global, un biais de contourage en faveur des sous-atlas a pu être introduit si les contours manuels ont été faits sans enlever de la visualisation les contours automatiques.

Quoi qu'il en soit, le critère de sélection selon la localisation et le stade de la tumeur primitive en tant que tel s'est malgré tout révélé pertinent pour la catégorie oropharynx N2. En effet, l'étude sur notre base d'images a permis de montrer que à nombre d'images moyennées égal, l'atlas moyen est plus performant si les images moyennées sont dans la même sous-catégorie que le patient à segmenter que si elles sont choisies aléatoirement indépendamment de leur localisation et stade de tumeur. Cela signifie que la localisation de la tumeur a effectivement un effet sur la forme et/ou la taille des structures ORL. Comme nous l'avons mis en évidence dans la section précédente, cet effet est cependant limité aux structures situées du côté opposé à la tumeur. Ainsi, le critère de sélection proposé ne permettrait pas de tirer béné-

face d'un éventuel impact prévisible de la tumeur sur les structures environnantes comme nous le pensions initialement, mais au contraire, de tirer profit de l'impact de l'absence de tumeur sur le côté opposé. Si l'absence de tumeur a un effet sur les structures, cela signifie indirectement que la présence de la tumeur a elle aussi un impact sur les structures environnantes. Le fait que nous n'ayons pas réussi à le mettre en évidence peut s'expliquer de deux façons : soit cet impact est imprévisible, soit il est trop variable d'un individu à l'autre et nous n'avons pas réussi à l'exploiter dans notre étude en raison du nombre restreint d'images dans chaque sous-catégorie. Par ailleurs, les structures manquantes dans notre base d'images sont plus souvent des structures situées du côté de la tumeur, ce qui contribue à augmenter artificiellement la variabilité de forme des structures du côté de la tumeur par rapport aux structures situées du côté sain.

Pour conclure, les sous-atlas par localisation et grade construits sur notre base ne permettent pas de dépasser les performances de l'atlas moyen global. Cependant, le critère de sélection par localisation et stade de tumeur nous semble tout de même pertinent pour la segmentation des structures ORL puisque les performances atteintes sont meilleures que celles fournies par la sélection aléatoire. Nous pensons par conséquent que la méthode proposée permettrait d'améliorer les résultats de l'atlas moyen global si elle était mise en œuvre avec une base d'images segmentées beaucoup plus grande que celle dont nous disposons. Une plus grande base permettrait d'une part d'estimer des sous-atlas plus précis. D'autre part, seules les tumeurs de l'oropharynx étaient suffisamment représentées dans notre base pour construire un sous-atlas représentatif, et une plus grande base serait indispensable pour construire un sous-atlas pour chacune des 20 sous-catégories.

Les perspectives liées à ce travail sont donc principalement liées à la constitution d'une plus grande base d'images et à la validation du schéma proposé sur l'ensemble des sous-catégories.

6.2 Évaluation des méthodes de sélection d'atlas basées image

Dans cette section, nous présentons l'évaluation des méthodes de sélection basées sur les informations image (c'est-à-dire sur les intensités ou sur les déformations). Les méthodes évaluées sont tout d'abord détaillées en section 6.2.1. Leur évaluation non biaisée selon le protocole proposé en section 5.3 est présentée en section 6.2.2. Enfin, le paragraphe 6.2.3 présente une autre approche d'évaluation des méthodes de sélection basée sur l'étude des performances rang par rang.

6.2.1 Méthodes de sélection d'atlas étudiées

6.2.1.1 Méthodes automatiques

Nous avons mis en œuvre et évalué des méthodes de sélection basées intensité et des méthodes de sélection basées déformation.

Concernant les méthodes de sélection basées intensité, nous avons testé les critères de similarité suivants :

- MI : information mutuelle calculée sur un masque \mathcal{M} ,
- NMI : information mutuelle normalisée calculée sur un masque \mathcal{M} ,
- SSD : « Sum of Squared Differences » calculée sur un masque \mathcal{M} ,
- CC : coefficient de corrélation calculé sur un masque \mathcal{M} ,
- CCLOC : moyenne sur un masque \mathcal{M} des coefficients de corrélation locaux (au carré) calculés sur des voisinages de taille $7 \times 7 \times 7$ voxels.

Pour chacun de ces critères de similarité, nous avons mis en œuvre les différentes approches de calcul résumées sur la figure 6.5 : mise en correspondance de type affine (notée AFF) ou non-linéaire (notée NL), mesure de similarité calculée dans le référentiel du patient à segmenter P (noté REFPAT) ou de l'atlas moyen M (noté REFAV), propagation directe (notée DIR) ou indirecte via M (notée IND) quand la mesure se fait dans le référentiel de P .

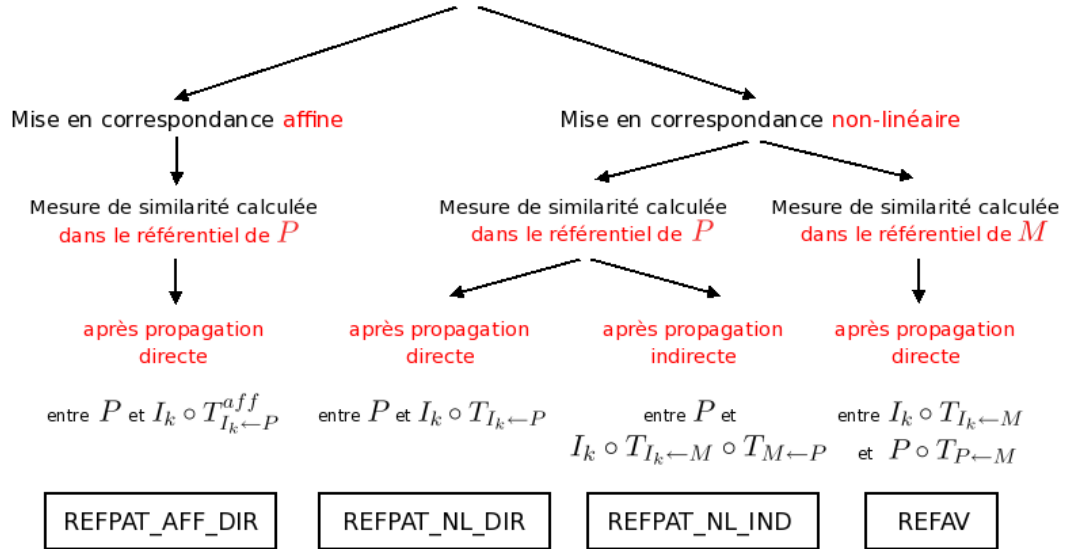


FIG. 6.5 – Méthodes de sélection basées sur le calcul d'une mesure de similarité de type intensité (CC, CCLOC, SSD, MI, NMI) sur un masque \mathcal{M} entre l'image P du patient à segmenter et l'image d'un atlas I_k (utilisant éventuellement l'atlas moyen M).

Pour chaque méthode de sélection, nous avons testé l'approche de sélection régionale (c'est-à-dire structure par structure) ainsi que l'approche de sélection globale (c'est-à-dire sur l'image entière), la seule différence entre ces deux approches étant le masque \mathcal{M} utilisé pour calculer la mesure de similarité. Nous avons choisi de définir les masques de calcul comme la dilatation morphologique des structures moyennes de l'atlas moyen M pour la sélection régionale, et comme la dilatation du contour externe de l'atlas moyen M pour la sélection globale. En pratique, trois dilatactions successives sont appliquées avec un élément structurant 3D de connexité 26.

Lorsque la sélection se fait dans le référentiel de M , ces masques sont utilisés tels quels. Lorsqu'elle se fait dans le référentiel du patient P à segmenter, ces masques sont déformés de façon non-linéaire sur P . L'utilisation de masques de calculs issus de l'atlas moyen plutôt que de l'union des segmentations de l'ensemble des atlas permet de s'affranchir de la très grande variabilité anatomique et de contourage présente dans notre base. En effet, avec cette seconde approche, la très grande variabilité conduit à des masques trop larges, et de plus l'échec d'un seul recalage lors de la mise en correspondance des atlas conduit à une augmentation considérable et non souhaitée du masque de calcul.

Nous avons également mis en œuvre deux méthodes de sélection basées sur les déformations. La première (notée DEF) est celle proposée par Commowick et al. [Commowick 2007a]. Elle consiste à quantifier l'amplitude de la déformation non-linéaire permettant de déformer chaque atlas sur le patient P en utilisant comme référentiel intermédiaire l'atlas moyen M . En notant \mathcal{M} un masque de calcul défini sur M , la distance entre chaque atlas I_k et le patient P sur ce masque est définie par :

$$d_{\mathcal{M}}(I_k, P) = \sum_{x \in \mathcal{M} \circ T_{P \leftarrow M}^{-1}} \|T_{I_k \leftarrow M}^{nl} \circ T_{P \leftarrow M}^{-1}(x) - Id(x)\| \quad (6.1)$$

La seconde méthode de sélection de type déformation que nous avons évaluée (notée JAC) est celle que nous avons proposée en section 5.2.5.2 et qui est basée sur le degré de dilatation ou de contraction des déformations non-linéaires des atlas vers l'atlas moyen, et du patient vers l'atlas moyen. Les masques de calcul sont ici définis dans le référentiel de l'atlas moyen M .

6.2.1.2 Méthodes de référence

Afin d'évaluer les méthodes de sélection automatiques détaillées précédemment, nous utilisons également des méthodes de sélection dites *de référence*. Ces méthodes de référence classent les atlas disponibles en fonction de leur aptitude à segmenter correctement le patient dans un processus de segmentation par atlas classique. Elles se caractérisent par des mesures de comparaison de segmentations calculées entre la segmentation manuelle du patient P et la segmentation automatique obtenue par propagation directe de la segmentation de chaque atlas sur P . Les mesures utilisées sont les mesures DICE, DIST, HD, RHD_{95%}, RHD_{90%} et VOL présentées dans l'annexe B. Pour être plus précis, VOL désigne ici la valeur absolue de la mesure décrite dans l'annexe. Lorsque la sélection est globale, la moyenne sur toutes les structures de chacune de ces mesures de référence est considérée.

6.2.2 Évaluation d'après les classements d'atlas

Dans cette section, nous présentons les résultats de l'évaluation non biaisée des méthodes de sélection d'atlas selon le protocole proposé en section 5.3. Les méthodes de sélection évaluées sont celles présentées dans le paragraphe précédent auxquelles

nous avons ajouté la sélection aléatoire (notée `RANDOM`) à titre de comparaison. Les notations utilisées sont celles décrites dans le paragraphe précédent et sur la figure 6.5. Nous détaillons d'abord les résultats de l'évaluation pour la sélection globale puis pour la sélection régionale.

6.2.2.1 Évaluation dans le cadre d'une sélection globale

Dans l'algorithme de partitionnement par affinité-propagation proposé par Frey et al. [Frey 2007], nous donnons une valeur d'auto-similarité commune pour toutes les méthodes car aucune des méthodes n'est *a priori* plus propice qu'une autre à être un exemplaire (c'est-à-dire une méthode représentative du sous-groupe auquel elle appartient). La valeur d'auto-similarité utilisée détermine indirectement le nombre de sous-groupes : si cette valeur est très faible par rapport aux similarités données en entrée, l'algorithme convergera vers un groupe unique, tandis que si cette valeur vaut 1, chaque méthode aura tendance à créer son propre sous-groupe et à être son propre exemplaire. L'étude du partitionnement résultant en fonction de la valeur d'auto-similarité permet d'observer les regroupements de méthodes à différentes échelles.

La figure 6.6 montre l'évolution des sous-groupes en fonction de la valeur d'auto-similarité donnée en entrée. L'algorithme d'affinité-propagation ne garantit en théorie pas une organisation hiérarchique des sous-groupes telle que nous l'avons représentée sur le schéma, mais dans notre cas l'évolution observée en est relativement proche. Par mesure de clarté, la figure illustre schématiquement les principales lignes de cette évolution sans tenir compte des exceptions qui ne corroboraient pas cette vision hiérarchique. Pour des valeurs d'auto-similarité $p < 0.8$, la représentation hiérarchique est vérifiée expérimentalement. Pour $0.8 < p < 1$, l'algorithme est très sensible et les sous-groupes ne s'organisent plus de façon hiérarchique ; la représentation hiérarchique choisie pour ces valeurs-là est une simplification. La figure 6.7 illustre les résultats de l'algorithme de partitionnement pour les différentes valeurs d'auto-similarité annotées sur le schéma 6.6. Sur cette figure, les couleurs représentent les valeurs des similarités données en entrée de l'algorithme (c'est-à-dire les coefficients de Spearman moyens pour chaque couple de méthode). Les sous-groupes résultants représentés par des carrés blancs permettent de mieux visualiser la maximisation des similarités intra-groupes (blocs diagonaux) et la minimisation des similarités inter-groupes.

Tout d'abord, nous observons sans surprise une distinction assez marquée entre les méthodes de référence d'un côté et les méthodes automatiques de l'autre. En effet, comme illustré sur le schéma 6.6, ces deux classes de méthodes ne sont réunies dans un même sous-groupe qu'au stade final où un seul groupe existe ($p < -2$). Concernant les méthodes de référence, on observe également un regroupement entre les mesures qui quantifient le chevauchement de segmentations (`DICE` et `DIST`) et celles qui caractérisent la distance entre contours (distances de Hausdorff et généralisations robustes).

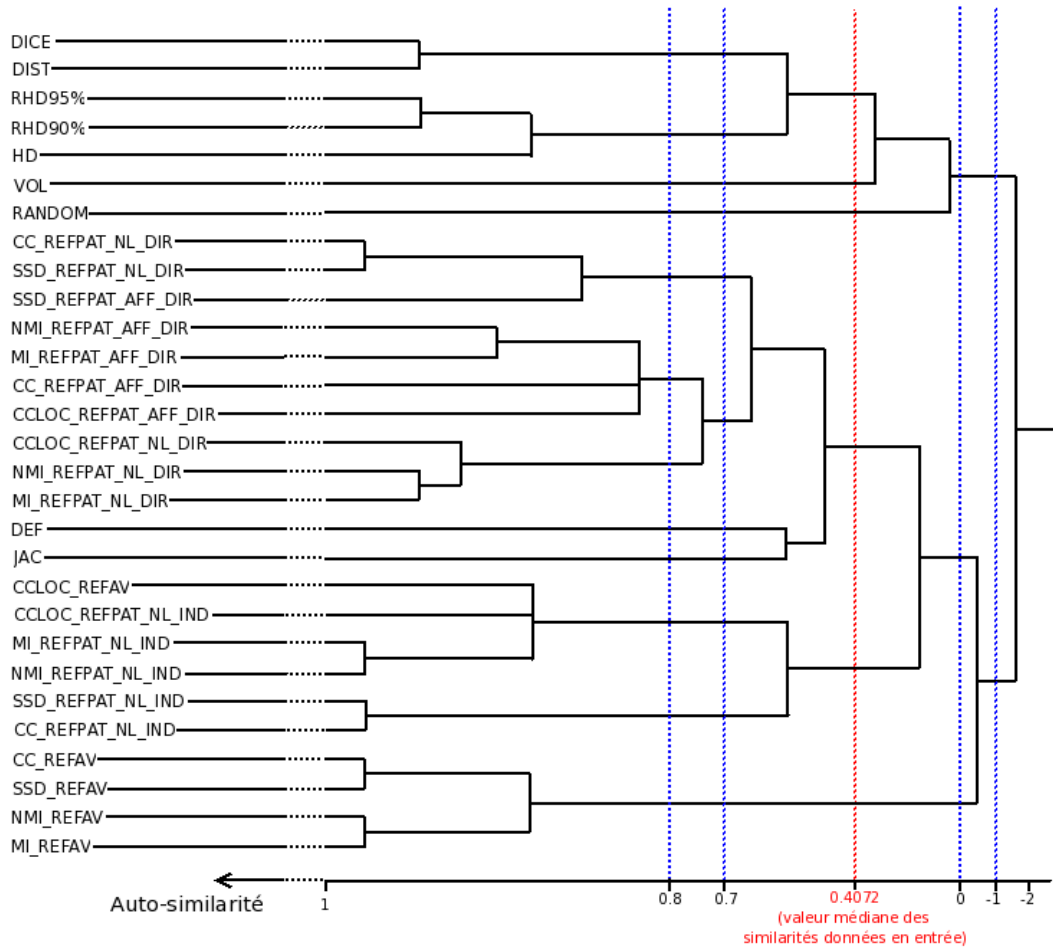


FIG. 6.6 – Évolution des sous-groupes de méthodes de sélection d’atlas en fonction de la valeur d’auto-similarité donnée en entrée de l’algorithme d’affinité propagation.

Concernant les méthodes automatiques, on observe principalement trois groupes sur le schéma 6.6 :

1. les méthodes basées intensité qui calculent la mesure de similarité dans le référentiel de l'atlas moyen (notées REFAV),
2. les méthodes basées intensité qui calculent la mesure de similarité dans le référentiel du patient mais qui utilisent l'atlas moyen comme référentiel intermédiaire pour la propagation (notées REFPAT_NL_IND),
3. les méthodes basées intensité qui calculent la mesure de similarité dans le référentiel du patient après une propagation directe affine ou non-linéaire (notées REFPAT_AFF_DIR et REFPAT_NL_DIR), ainsi que les méthodes basées sur les déformations (DEF et JAC).

Le premier groupe (REFAV) semble relativement isolé des deux autres groupes comme le met en évidence le code de couleurs sur la figure 6.7 pour $p = 0.4072$. Ceci est confirmé par le schéma 6.6 qui montre un regroupement tardif du premier groupe avec les deux autres (pour $p = -1$ seulement). Pour le second groupe (REFPAT_NL_IND), on observe globalement une distinction entre propagation affine et propagation non-linéaire.

Comme détaillé en section 5.3.3, le calcul d'une corrélation inter-groupe permet de déterminer quel sous-groupe de méthodes automatiques se rapproche le plus d'un sous-groupe de référence. Pour ce calcul, nous utilisons la valeur d'auto-similarité conseillée par Frey et al., à savoir la valeur médiane des similarités données en entrée (0.4092 dans notre cas). Comme illustré sur les figures 6.6 et 6.7, les méthodes se répartissent en six sous-groupes pour cette valeur d'auto-similarité, la méthode de sélection aléatoire et la méthode VOL étant seules dans leurs sous-groupes respectifs. Parmi les méthodes de référence, la mesure VOL est isolée, donc nous calculons les corrélations inter-groupes entre chaque sous-groupe de méthodes automatiques et le sous-groupe réunissant les cinq autres méthodes de référence. Les valeurs obtenues pour les sous-groupes REFAV, REFPAT_DIR/DEF/JAC et REFPAT_NL_IND sont respectivement 0.23, 0.32, 0.22. On en conclut donc que le sous-groupe le plus proche du sous-groupe de référence est le sous-groupe REFPAT_DIR/DEF/JAC, c'est-à-dire les méthodes mesurant la similarité dans le référentiel du patient après propagation directe de l'atlas (affine ou non-linéaire) ainsi que les méthodes basées déformation. A titre indicatif, le sous-groupe de référence obtient une corrélation de 0.40 et 0.00 avec les méthodes VOL et RANDOM respectivement.

D'autre part, lorsque toutes les méthodes se réunissent en un seul groupe ($p = -2$), la méthode qui joue le rôle d'exemplaire (c'est-à-dire méthode la plus représentative du groupe) est la méthode CCLOC_REFPAT_NL_DIR. Ceci semblerait indiquer que la méthode CCLOC_REFPAT_NL_DIR est la plus proche des méthodes de référence. Ceci sera confirmé dans le paragraphe suivant.

6.2.2.2 Évaluation dans le cadre d'une sélection régionale

Dans le paragraphe précédent, nous avons comparé et évalué les méthodes automatiques de sélection d'atlas lorsque la sélection est effectuée de façon globale sur l'image. Nous procédons maintenant à une étude similaire, mais lorsque la sélection se fait de façon régionale, c'est-à-dire structure par structure.

La figure 6.8 présente à titre d'exemple le résultat de l'algorithme de partitionnement appliqué à une sélection régionale sur les parotides. La valeur d'auto-similarité utilisée est la valeur médiane des similarités données en entrée (0.1737). Sur cette figure, on remarque tout d'abord que les similarités inter-groupes (en dehors des blocs diagonaux) sont beaucoup plus faibles que pour la sélection régionale. Cela traduit des différences plus significatives entre les différentes méthodes de sélection. D'autre part, pour le cas particulier des parotides, on remarque que contrairement au cas de la sélection globale, la mesure de similarité utilisée (CC, CCLOC, SSD, MI ou NMI) est un critère plus important que la façon dont est mise en œuvre le calcul de similarité (REFAV, REFPAT_AFF_DIR, REFPAT_NL_DIR ou REFPAT_NL_IND). En effet, les méthodes basées sur les mesures MI et NMI semblent très peu corrélées aux méthodes de référence quelle que soit la façon dont elles sont mises en œuvre.

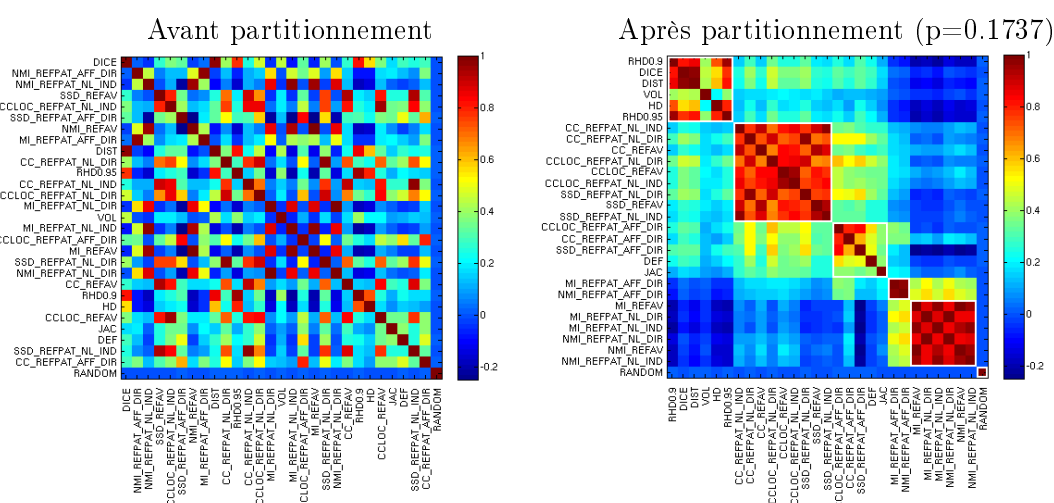


FIG. 6.8 – Coefficient de Spearman moyen pour chaque couple de méthodes de sélection d'atlas (méthodes automatiques ou de référence appliquées ici de façon régionale sur les parotides) avant partitionnement (à gauche), et après partitionnement pour une valeur d'auto-similarité égale à la médiane des valeurs de similarité données en entrée (à droite).

Les résultats de l'algorithme de partitionnement sont relativement différents selon les structures. Par concision, nous ne les présentons pas pour toutes les structures. Afin d'avoir tout de même une vue d'ensemble des résultats pour la sélection régionale, la figure 6.9 présente les coefficients de Spearman moyens de chaque méthode de sélection avec la méthode de référence basée sur l'indice de DICE. Par me-

sure de clarté, les écarts-types ne sont pas représentés. Les résultats sont présentés pour la sélection régionale sur chaque structure (AG. II, AG. III, AG. IV, Paros., Sous-max, Moelle, Tronc cérébral, Mandibule) ainsi que pour la sélection globale (en haut, Global) à titre de comparaison. Les différents groupes de méthodes sont représentés par des nuances de couleurs.

Pour la sélection régionale, à l'exception du tronc cérébral et de la mandibule, le graphique permet de tirer les conclusions suivantes :

1. Quelle que soit la manière dont est calculée la mesure de similarité (REFPAT_AFF_DIR, REFPAT_NL_DIR, REFPAT_NL_IND ou REFAV), la mesure CCLOC est très souvent celle qui donne le meilleur résultat.
2. Quelle que soit la mesure utilisée (CCLOC, CC, SSD, MI ou NMI) :
 - la méthode REFPAT_NL_DIR (nuances de gris) donne souvent le meilleur résultat, tandis que la méthode REFPAT_AFF_DIR (nuances de bleu) donne souvent le moins bon résultat ;
 - les méthodes REFAV (nuances de vert) et REFPAT_NL_IND (nuances de rose) donnent des résultats similaires.

On remarque que la méthode CCLOC_REFPAT_NL_DIR (en blanc) est la méthode automatique qui fournit la meilleure corrélation de classement avec l'indice de DICE, et ce pour toutes les structures à l'exception du tronc cérébral (pour lequel elle figure tout de même parmi les trois meilleures méthodes). Après cette méthode, les deux autres méthodes qui semblent les mieux corrélées aux classements selon DICE pour la sélection régionale sont les méthodes CCLOC_REFPAT_NL_IND (en rose vif) et CCLOC_REFAV (en vert foncé). Ainsi, l'utilisation du coefficient de corrélation local après mise en correspondance non-linéaire semble être la combinaison la plus pertinente pour la sélection régionale.

Pour la sélection régionale, le critère de sélection basé sur le jacobien que nous avons proposé semble fournir de moins bons résultats que la méthode DEF basée sur l'amplitude des déformations (effet moins marqué pour la sélection globale). Par ailleurs, comme mis en évidence pour les parotides sur la figure 6.8, l'utilisation de l'information mutuelle (normalisée ou non) pour la sélection régionale donne globalement de moins bons résultats que les autres mesures, en particulier pour les parotides, les sous-maxillaires et les aires ganglionnaires IV (coefficient de Spearman moyen négatif). Au niveau régional pour ces structures, la relation entre les intensités des deux images comparées semble ainsi correspondre suffisamment bien aux hypothèses faites sur leur nature par les mesures SSD, CC et CCLOC, tandis que l'information mutuelle a probablement tendance à être sous-contrainte.

Enfin, les données présentées permettent également de conclure que la sélection régionale est plus pertinente que la sélection globale. Cette conclusion n'est pas générale à la totalité des méthodes de sélection, mais toutes méthodes confondues la sélection régionale permet d'atteindre de bien meilleures corrélations avec la méthode de référence (DICE) que la sélection globale.

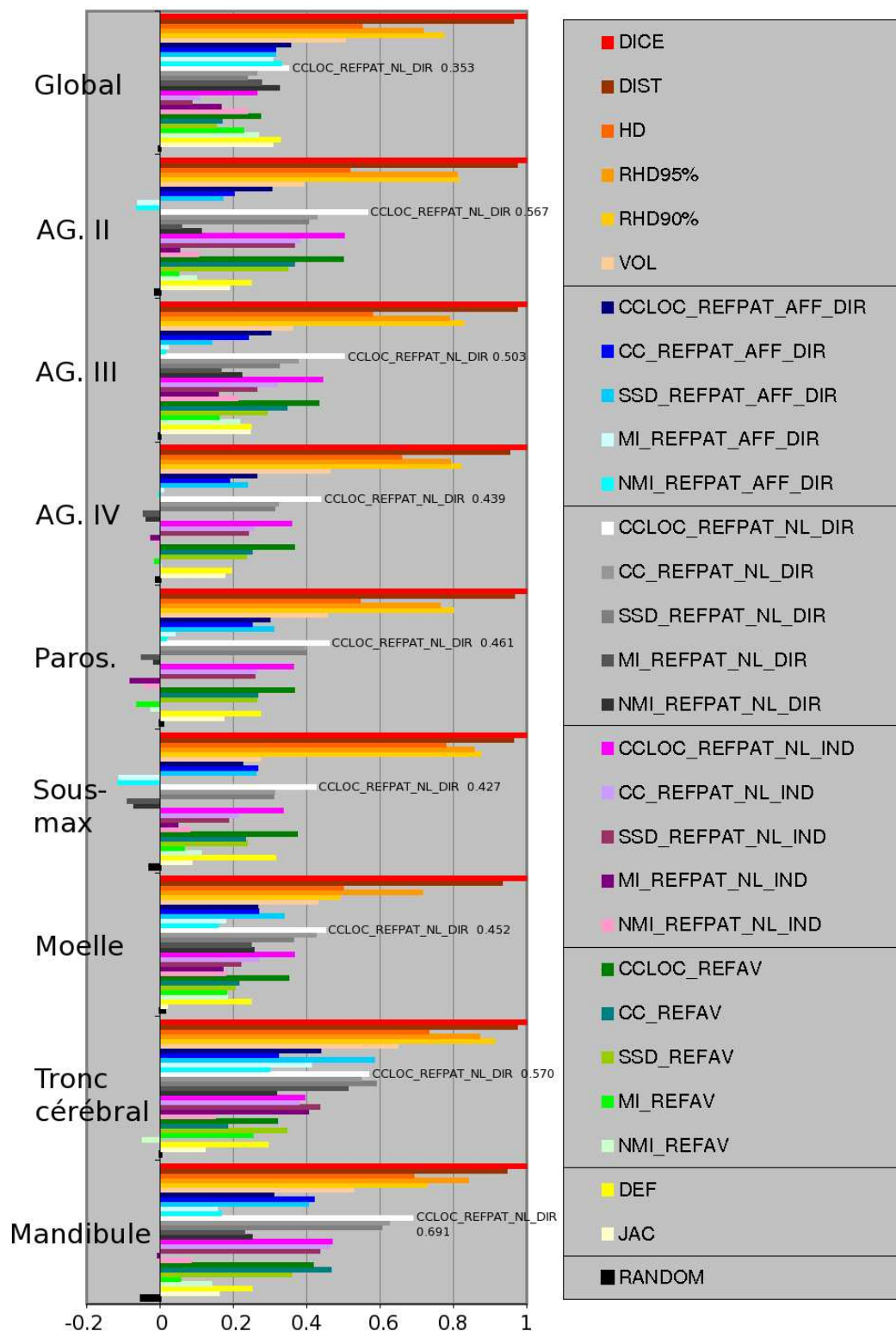


FIG. 6.9 – Coefficients de Spearman moyens de chaque méthode de sélection avec la méthode de référence basée sur l'indice de DICE, et ce pour la sélection globale (en haut, Global) ainsi que pour la sélection régionale pour les différentes structures AG. II, AG. III, AG. IV, Paros., Sous-max, Moelle, Tronc cérébral et Mandibule. Les différents groupes de méthodes sont représentés par des nuances de couleurs.

6.2.3 Évaluation d'après les performances rang par rang

L'évaluation présentée dans le paragraphe précédent n'exploitait que les classements d'atlas à proprement parler, mais pas les performances de segmentation (en terme d'indice de DICE par exemple) obtenues avec les atlas selon leur rang dans le classement. Nous présentons dans ce paragraphe une étude des performances des atlas rang par rang (analyse de type Leave-One-Out sur la base).

La mesure de performance utilisée est l'indice de DICE après propagation directe des contours de chaque atlas sur le patient. Par soucis de concision, nous n'étudions que les méthodes de sélection basées sur les coefficients de corrélation locaux (CCLOC), les méthodes basées sur les déformations (DEF et JAC) ainsi que la méthode de sélection aléatoire (RANDOM). Nous présentons seulement les résultats correspondant à une sélection globale. La figure 6.10 présente l'indice de DICE moyen par rang pour ces différentes méthodes de classement d'atlas. La courbe rouge fait office de référence, elle correspond à l'indice de DICE moyen par rang lorsque les atlas sont justement classés en fonction de l'indice de DICE qu'ils permettent d'atteindre après propagation directe. Ici encore, les écarts-types ne sont pas représentés pour plus de clarté.

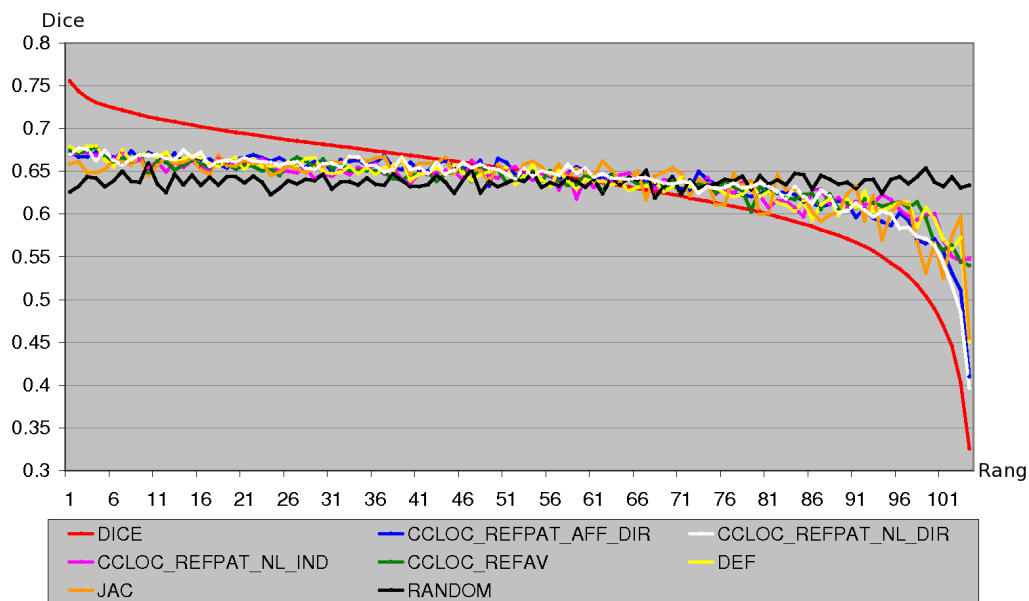


FIG. 6.10 – Indice de DICE moyen par rang pour différentes méthodes de classement d'atlas dans le cadre d'une sélection globale. La courbe rouge correspond au meilleur résultat atteignable (DICE moyen par rang lorsque les atlas sont classés selon DICE).

Tout d'abord, on remarque que la sélection aléatoire fournit en moyenne un DICE identique à quelques oscillations près quel que soit le rang de l'atlas (0.64 en moyenne sur tous les rangs), ce qui est cohérent. Concernant les méthodes automatiques étudiées, leurs courbes se rapprochent de la courbe rouge surtout pour les

derniers rangs (approximativement à partir du rang 70 sur 105). Pour les premiers rangs en revanche, elles permettent en moyenne d'atteindre une performance un peu supérieure à celle de la sélection aléatoire (autour de 0.67 pour le premier rang), mais largement inférieure à la performance de référence (0.76 pour le premier rang). Cela signifie que les méthodes automatiques de sélection d'atlas semblent beaucoup plus efficaces pour discriminer les atlas les moins similaires au patient à segmenter plutôt que pour détecter les atlas les plus similaires. Cela s'explique par le fait que les mauvais atlas correspondent souvent à des atlas pour lesquels le recalage avec le patient a échoué, donc leur identification est relativement simple. En revanche, la sensibilité des mesures de similarité rend plus difficile de dissocier parmi les atlas restants ceux pour lesquels le recalage est de très bonne qualité de ceux pour lesquels le recalage est de qualité correcte. Il se peut cependant que les profils relativement plats des courbes pour les rangs inférieurs à 70 soient dus à un effet de moyenne. En d'autres termes, il est possible que la sélection des atlas les plus similaires au patient soit très efficace pour certains patients (par exemple ceux ayant une anatomie atypique) et peu efficace pour d'autres. Si l'on compare maintenant plus finement les courbes des méthodes automatiques entre elles, on remarque plusieurs choses. Les méthodes les plus efficaces pour discriminer les plus mauvais atlas (fin de courbe) sont `CCLOC_REFPAT_AFF_DIR` (en bleu) et `CCLOC_REFPAT_NL_DIR` (en blanc). Cela vient confirmer nos conclusions basées sur le classement (cf. section 6.2.2.1), à savoir que pour la sélection globale, la propagation directe vers le référentiel du patient est plus efficace que les autres approches qui utilisent l'atlas moyen d'une façon ou d'une autre. Les méthodes `CCLOC_REFPAT_NL_IND` (en rose) et `CCLOC_REFAV` (en vert) semblent ici donner des résultats similaires. Enfin, la courbe correspondant à la méthode `JAC` présente davantage d'oscillations que les autres pour les rangs élevés, ce qui suggère une plus grande sensibilité. La comparaison des courbes pour les premiers rangs ne met pas en évidence de différences significatives.

6.2.4 Synthèse et discussion

Dans cette partie, nous avons comparé et évalué différents critères de sélection d'atlas basés sur les intensités ou les déformations, ainsi que différentes façons de les mettre en œuvre : de façon globale ou régionale, et dans différents référentiels (patient ou atlas moyen) avec différents degrés de mise en correspondance (affine ou non-linéaire). Nous avons mené deux types d'évaluations : la première est basée sur le partitionnement des méthodes de sélection d'après les corrélations entre les classements d'atlas qu'elles fournissent, la seconde est basée sur l'analyse des performances des atlas rang par rang. Ces deux types d'évaluations fournissent des informations complémentaires. En effet, la première est une évaluation conjointe de l'ensemble des méthodes de sélection et permet de mettre en évidence des sous-groupes de méthodes de sélection approximativement équivalentes. Dans la seconde approche au contraire, les méthodes de sélection sont évaluées les unes indépendamment des autres selon les performances qu'elles permettent d'atteindre en terme de qualité de segmentation rang par rang.

Les deux types d'évaluation menées (selon les classements et selon les performances) ont conduit à des conclusions similaires :

1. Pour la sélection globale, les méthodes basées intensité les plus efficaces sont celles qui utilisent une propagation directe des atlas vers le patient (affine ou non-linéaire) ; celles qui utilisent l'atlas moyen comme référentiel de calcul ou comme référentiel intermédiaire pour la propagation des atlas vers le patient se sont révélées moins pertinentes.
2. Pour la sélection régionale, le coefficient de corrélation local `CCLOC` est la mesure de similarité la plus pertinente, et la mise en correspondance de type non-linéaire (de préférence avec une propagation directe) donne de meilleurs résultats que celle de type affine.
3. A la fois pour les sélection globale et régionale, la méthode de sélection la plus efficace est `CCLOC_REFPAT_NL_DIR`.
4. La méthode de sélection basée sur les jacobiens que nous avons proposée donne dans l'ensemble de moins bons résultats que la méthode basée sur l'amplitude des déformations (différences plus marquées pour la sélection régionale que pour la sélection globale).

Enfin, l'évaluation des performances rang par rang a permis de mettre en évidence qu'avec notre base d'images, les méthodes de sélection étudiées semblaient plus efficaces (du moins plus robustes) pour discriminer les plus mauvais atlas que pour détecter les meilleurs atlas pour un patient donné. Cela peut cependant être dû à un effet de moyenne.

La figure 6.11 montre un exemple de sélection globale d'atlas avec la méthode optimale `CCLOC_REFPAT_NL_DIR` pour un patient donné. On remarque que les trois atlas sélectionnés comme étant les plus similaires au patient ont une anatomie ressemblante à celle du patient, notamment au niveau de la flexion de la colonne. Les trois atlas les moins similaires au patient sont principalement des patients corpulents.

La figure 6.12 présente les valeurs de l'indice de DICE (après propagation directe) en fonction de la mesure de similarité `CCLOC_REFPAT_NL_DIR`, pour la sélection globale (premier graphique) et pour la sélection régionale structure par structure (graphiques suivants). La totalité des cas est ici reportée : chaque point correspond aux mesures entre un des 105 patients, et un des 104 atlas parmi les patients restants. Une courbe de tendance linéaire est ajoutée sur chaque graphique (en jaune). La dispersion des points autour de cette courbe de tendance est variable selon les structures. Cependant, des conclusions comparatives entre les différentes structures ne peuvent pas être tirées car le nombre de points affichés est différent pour chaque structure. Les faibles valeurs de la mesure `CCLOC_REFPAT_NL_DIR` proviennent du fait que ce sont des coefficients de corrélation locaux au carré qui sont moyennés.

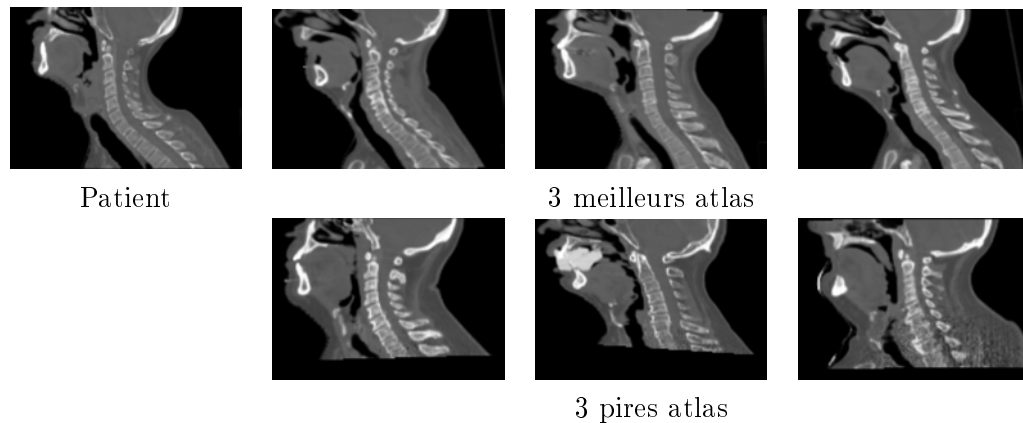


FIG. 6.11 – Exemples d’atlas sélectionnés de façon globale avec la méthode `CCLOC_REFPAT_NL_DIR` pour un patient donné : les trois atlas les plus similaires (première ligne) et les moins similaires (deuxième ligne) au patient ont été déformés linéairement sur le patient pour la visualisation.

6.3 Évaluation des approches de fusion des segmentations dans le référentiel du patient

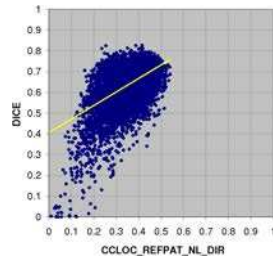
Dans la section précédente, nous avons étudié différentes méthodes de sélection d’atlas indépendamment de la phase de fusion. Nous avons mis en évidence que la méthode la plus pertinente pour nos images est celle basée sur le coefficient de corrélation local mesuré dans le référentiel du patient entre l’image du patient et les atlas propagés de façon directe sur le patient (notée `CCLOC_REFPAT_NL_DIR`). Nous avons également montré que la sélection régionale permet d’obtenir de meilleurs résultats que la sélection globale, point que nous allons confirmer dans cette partie.

Dans cette section, nous comparons maintenant en utilisant la méthode de sélection choisie (`CCLOC_REFPAT_NL_DIR`) les performances atteintes avec les différentes approches de fusion de segmentation dans le référentiel du patient présentées en section 5.4.4.2, à savoir :

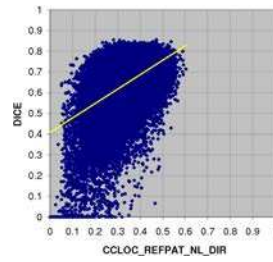
- la fusion des segmentations après sélection globale ou régionale,
- la fusion des segmentations avec une pondération locale.

Nous présentons dans les sections 6.3.1 et 6.3.2 pour chaque méthode l’influence des paramètres principaux, notamment le type de propagation utilisé pour normaliser les segmentations dans le référentiel du patient (directe ou indirecte via l’atlas moyen). Pour la sélection régionale ou globale, le fil conducteur de l’étude est l’évaluation de l’évolution de l’indice de DICE en fonction du nombre d’atlas fusionnés.

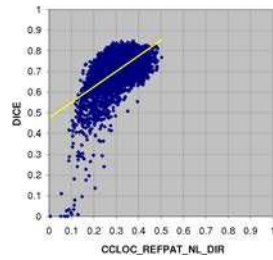
Les différentes approches présentées (sélection globale, sélection régionale ou pondération locale, ainsi que propagation directe ou indirecte) sont ensuite comparées entre elles et comparées à l’utilisation de l’atlas moyen global, de façon quantitative (cf. section 6.3.3) puis qualitative (cf. section 6.3.4). Par soucis de concision, les résultats quantitatifs présentés se restreignent à l’indice de DICE.



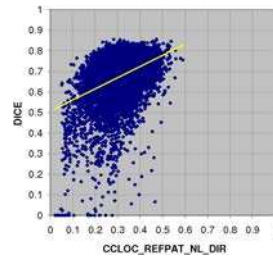
Sél. globale



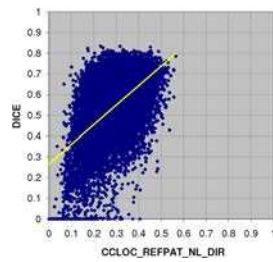
Sél. régionale (toutes structures)



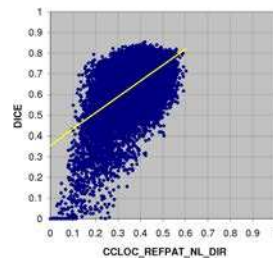
Sél. régionale sur les AG. II



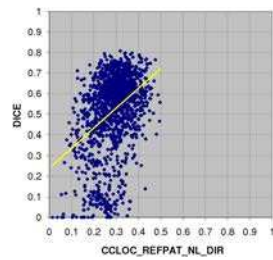
Sél. régionale sur les AG. III



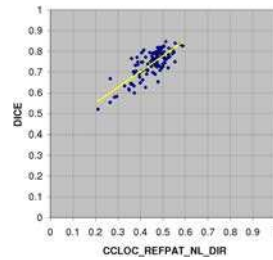
Sél. régionale sur les AG. IV



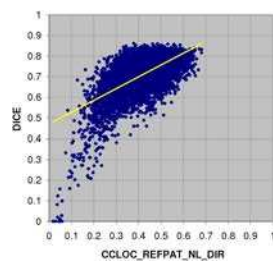
Sél. régionale sur les parotides



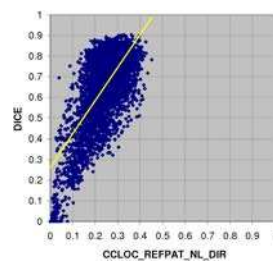
Sél. régionale sur les sous-maxillaires



Sél. régionale sur la mandibule



Sél. régionale sur la moelle épinière



Sél. régionale sur le tronc cérébral

FIG. 6.12 – Valeur de la mesure DICE_NL_DIR en fonction de la mesure CCLOC_REFPAT_NL_DIR, pour la sélection globale (premier graphique) et pour la sélection régionale (graphiques suivants).

6.3.1 Évaluation dans le cadre d'une sélection globale ou régionale

Pour un patient P donné, l'algorithme utilisé comporte les étapes suivantes :

- Pour chaque structure ou bien de façon globale, sélection des K atlas les plus similaires à P avec la méthode `CCLOC_REFPAT_NL_DIR` (lorsque la sélection se fait de façon régionale sur une structure donnée, seuls les atlas ayant cette structure segmentée sont considérés),
- Propagation directe ou indirecte (c'est-à-dire via l'atlas moyen) des atlas sélectionnés sur P ,
- Fusion des segmentations dans le référentiel du patient avec l'algorithme de vote par majorité lissé présenté en section 4.5.2.2,
- Évaluation de la segmentation automatique résultante par comparaison à la segmentation manuelle de P (avec l'indice de DICE).

Afin de confirmer notre choix de méthode de sélection, nous avons également appliqué l'algorithme avec quelques autres des méthodes de sélection : les différentes mises en œuvre de la sélection avec `CCLOC` ainsi que les méthodes de sélection basées sur les déformations (`DEF` et `JAC`). Comme borne supérieure et borne inférieure de performances, nous utilisons les méthodes de sélection `DICE` et `RANDOM` respectivement. Une étude exhaustive non présentée ici a permis de montrer que pour l'ensemble de ces méthodes de sélection et dans tous les cas de figure (sélection globale ou régionale, propagation directe ou indirecte), la valeur optimale pour le seuil des cartes de probabilités lissées est 0.35 (avec un filtrage gaussien d'écart-type 1).

La figure 6.13 présente les performances atteintes (en terme de DICE moyen sur toutes les structures, moyenné sur l'ensemble des patients) pour ces différentes méthodes de sélection en fonction du nombre de segmentations d'atlas fusionnées dans le référentiel du patient, et ce pour la sélection globale ou régionale, et pour la propagation directe ou indirecte des segmentations des atlas dans le référentiel du patient. Sur cette figure, nous avons également ajouté à l'abscisse 104 la performance obtenue avec l'atlas moyen global construit à partir des 105 patients à titre de comparaison (carré bleu). Dans les deux paragraphes suivants, nous détaillons l'analyse des différents graphes de cette figure. Dans le dernier paragraphe, nous présentons brièvement des résultats similaires mais cette fois-ci structure par structure.

6.3.1.1 Propagation directe *versus* indirecte

Convergence Tout d'abord, nous remarquons que les courbes de performances correspondant à la propagation indirecte (b et d) convergent vers la performance de l'atlas moyen global, tandis que les courbes de performances correspondant à la propagation directe (a et c) convergent vers une valeur qui est inférieure. D'une part, le fait que les courbes des graphiques (b) et (d) convergent vers l'atlas moyen est cohérent puisque les seules différences entre le cas $K = 104$ et l'utilisation de l'atlas moyen global sont minimales : (1) la fusion de segmentations se fait dans le référentiel du patient alors que pour l'atlas moyen, elle s'est faite dans le référentiel de l'atlas moyen, et (2) l'atlas moyen utilisé n'a pas été construit en `Leave-One-Out`. D'autre part, le fait que les courbes des graphiques (a) et (c) convergent vers

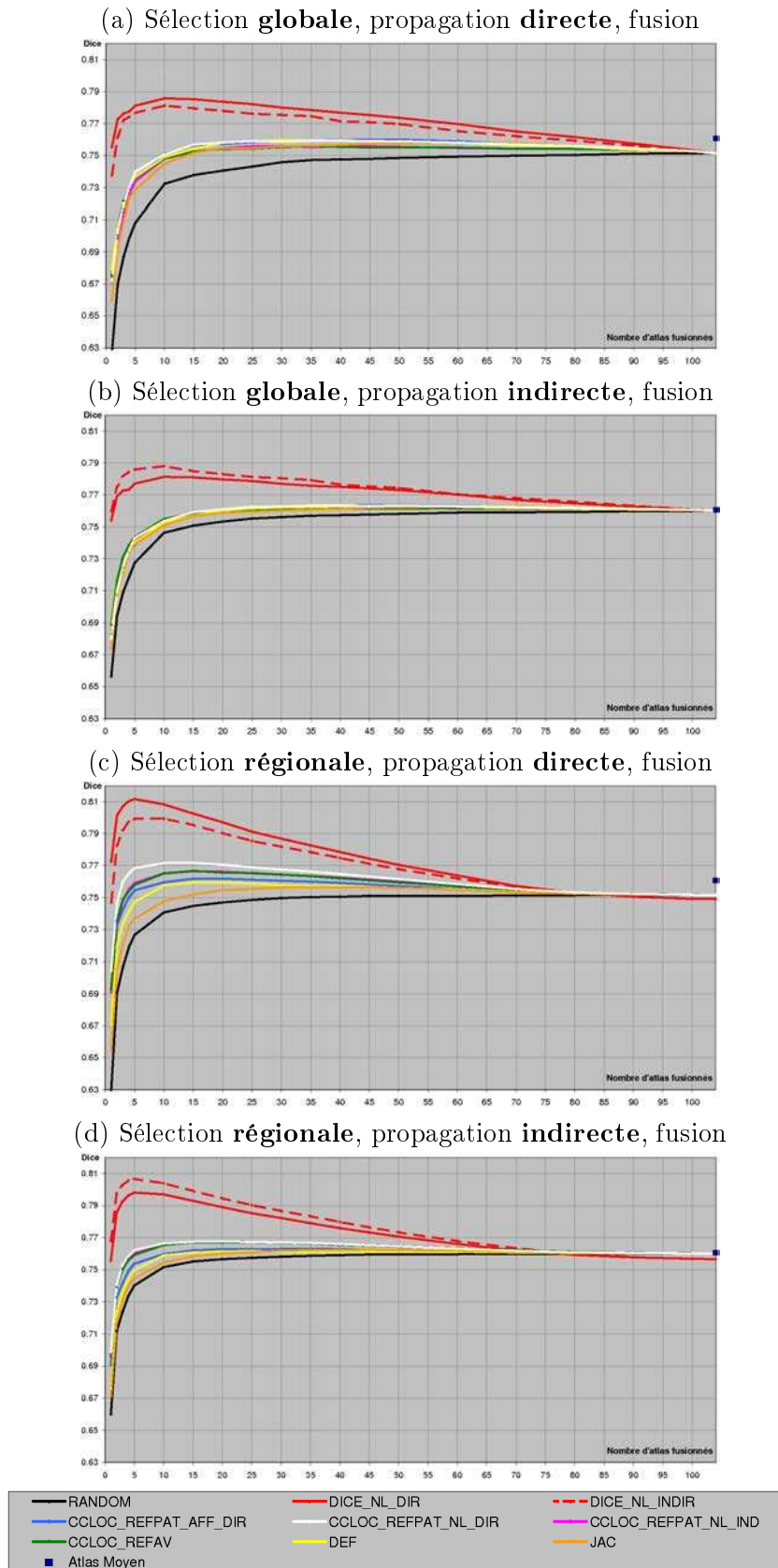


FIG. 6.13 – Moyenne sur tous les patients de l'indice de DICE moyen sur toutes les structures en fonction du nombre de segmentations d'atlas fusionnées dans le référentiel du patient après sélection globale ou régionale (selon diverses méthodes de sélections) et propagation directe ou indirecte.

une valeur inférieure signifie que en moyenne sur tous les patients, la propagation directe est moins performante que la propagation indirecte. Cette observation est à corrélérer avec la grande variabilité anatomique de notre base d'images, qui fait que en moyenne, chaque patient est plus proche de l'atlas moyen que des autres patients (et donc dans une majorité de cas, l'utilisation de l'atlas moyen comme référentiel intermédiaire permet de décomposer en deux recalages successifs les grandes déformations difficilement récupérables par un simple recalage non-linéaire).

Meilleures performances Si l'on compare maintenant les propagations directe et indirecte lorsqu'un faible nombre d'atlas sont fusionnés, on constate que la propagation directe permet d'atteindre de meilleures performances que la propagation indirecte. Ceci n'est pas vérifié pour la sélection globale, pour laquelle les deux types de propagation conduisent à des performances similaires (graphes (a) et (b)). Pour la sélection régionale en revanche (graphes (c) et (d)), cela est vérifié de façon significative pour la courbe blanche (`CCLOC_REFPAT_NL_DIR`). En fait, dès lors que la sélection des meilleurs atlas est de bonne qualité, la propagation directe des segmentations des atlas sélectionnés dans le référentiel du patient donne de meilleurs résultats que la propagation indirecte. Si au contraire la sélection n'est pas satisfaisante (sélection globale, ou sélection `RANDOM` par exemple), les atlas sélectionnés seront en moyenne assez différents du patient et le recalage indirect donnera en moyenne de meilleurs résultats. La propagation directe a ainsi une plus grande plage de variations : elle recalc mieux les atlas similaires au patient, mais moins bien les atlas trop différents du patient.

Synthèse Pour conclure cette partie, nous avons mis en évidence que lorsque les segmentations des atlas sont fusionnées dans le référentiel du patient, la propagation indirecte donne de meilleurs résultats en moyenne sur tous les atlas, mais si la sélection d'atlas est satisfaisante (c'est-à-dire si les atlas sélectionnés sont effectivement les plus similaires au patient), alors la propagation directe permet d'atteindre de meilleures performances par des recalages plus précis. En d'autres termes, les atlas réellement similaires au patient sont mieux recalés par recalage direct, tandis que les atlas les moins similaires au patient sont mieux recalés par recalage indirect via l'atlas moyen. En conséquence, le type de propagation à utiliser (direct ou indirect) dans le cadre de la fusion des atlas sélectionnés comme étant les plus similaires au patient dépend avant tout de la pertinence de la méthode de sélection utilisée. Si celle-ci est efficace, alors la propagation directe sera préférable. Si au contraire elle n'est pas efficace ou pas suffisamment robuste, alors la propagation indirecte aura tendance à donner de meilleurs résultats.

6.3.1.2 Sélection globale *versus* régionale

Tout d'abord, les graphes présentés pour les deux types de propagations (directe et indirecte) confirment les conclusions tirées lors de l'évaluation des méthodes de sélection d'atlas, à savoir :

- Les méthodes automatiques sont relativement équivalentes les unes des autres pour la sélection globale (graphes (a) et (b)), mais ce n'est pas le cas pour la sélection régionale (graphes (c) et (d)) pour laquelle on observe des différences significatives entre les différentes méthodes de sélection. Ceci est à corrélérer avec les résultats présentés sur la figure 6.9.
- Pour la sélection régionale (graphes (c) et (d)), les méthodes de sélection de la plus performante à la moins performante sont : CCLOC_REFPAT_NL_DIR, CCLOC_REFAV et CCLOC_REFPAT_NL_IND qui sont équivalentes, CCLOC_REFPAT_AFF_DIR, DEF puis JAC. Encore une fois, ceci est cohérent avec les résultats présentés sur la figure 6.9 et les conclusions que nous en avons tirées.
- La sélection régionale permet d'atteindre de meilleures performances que la sélection globale, ce qui corrobore également nos précédentes conclusions.

De plus, les graphes permettent de comparer plus finement les sélections globale et régionale. La sélection globale ne permet pas de dépasser les performances de l'atlas moyen, que ce soit avec une propagation directe (graphe (a)) ou indirecte (graphe (b)). Dans les deux cas, les performances de l'atlas moyen sont atteintes pour un nombre d'atlas fusionnés environ égal à 40, puis un pallier est atteint. La sélection régionale permet d'atteindre de meilleures performances que la sélection globale et que l'atlas moyen, et le nombre optimal de segmentations à fusionner par structure est compris entre 10 et 20 pour la plupart des méthodes. Ainsi, la sélection régionale est plus pertinente que la sélection globale car plus ciblée, et elle fournit de meilleures performances avec un plus faible nombre de segmentations sélectionnées par structure. Pour la méthode CCLOC_REFPAT_NL_DIR (courbe blanche), le nombre optimal est situé entre 10 et 15 pour la propagation directe (graphe (c)) et entre 15 et 20 pour la propagation indirecte (graphe (d)). Au passage, cela vient confirmer que pour la sélection régionale et avec ce critère de sélection, la propagation directe est plus pertinente que la propagation indirecte puisqu'un plus faible nombre d'atlas est nécessaire pour atteindre le maximum. Par ailleurs, les courbes obtenues avec la sélection régionale (graphes (c) et (d)) ont un profil plus avantageux que celles de la sélection globale (graphes (a) et (b)) : elles ont en effet une plus forte pente en début de courbe, ce qui caractérise une sélection plus robuste. De même, les courbes rouges correspondant à nos bornes supérieures ont un maximum beaucoup plus marqué dans les graphes (c) et (d) que dans les graphes (a) et (b).

Enfin, on remarque que les courbes de sélection régionale et globale convergent vers les mêmes valeurs pour chaque type de propagation (c'est-à-dire mêmes valeurs de convergence pour les courbes (a) et (c) d'un côté, et (b) et (d) de l'autre). On remarque cependant que pour la sélection régionale (graphes (c) et (d)), les deux courbes rouges de référence (DICE_NL_DIR et DICE_NL_IND) convergent vers une valeur un peu inférieure à la valeur de convergence des courbes correspondant aux méthodes de sélection automatique. Ceci s'explique par le fait que la sélection régionale d'après l'indice de DICE ne permet pas de calculer une segmentation automatique pour les structures non segmentées manuellement chez le patient. De ce fait, les structures voisines de structures manquantes chez le patient ne seront pas corrigées

au niveau de la jonction lors du SKIZ. Au contraire, la sélection régionale d'après les méthodes automatiques fournit dans tous les cas une segmentation complète, et toutes les jonctions entre structures sont affectées par le SKIZ. On voit d'ailleurs que cette correction de chevauchement a un effet positif puisque la valeur de convergence des méthodes automatiques est plus élevée. Pour les patients ayant toutes les structures segmentées, les courbes de sélection globale et régionale convergent vers la même valeur pour un type donné de propagation (non présenté ici).

Pour finir, on remarque que pour la sélection régionale, seules les méthodes de sélection basées sur la mesure CCLOC et sur une mise en correspondance non-linéaire (c'est-à-dire REFPAT_NL_DIR, REFPAT_NL_IND et REFAV) permettent de dépasser les performances de l'atlas moyen global (courbes blanche, rose, et verte). Pour les autres méthodes de sélection (CCLOC après mise en correspondance affine et méthodes basées déformation), la performance maximale est approximativement égale à celle de l'atlas moyen.

6.3.1.3 Évaluation structure par structure

Dans le paragraphe précédent, nous avons étudié l'influence de la sélection (globale ou régionale et nombre d'atlas sélectionnés) et de la propagation (directe ou indirecte) sur les performances atteintes en terme d'indice de DICE moyen sur toutes les structures. D'après cette analyse, la combinaison optimale est la suivante :

- sélection régionale des atlas les plus similaires au patient de façon régionale (structure par structure) avec la méthode CCLOC_REFPAT_NL_DIR (nombre optimal situé ente 10 et 15),
- propagation directe des segmentations des atlas sélectionnés sur le patient.

Cette combinaison optimale permet en moyenne sur toutes les structures de dépasser les performances de l'atlas moyen de façon significative. Cependant, l'amélioration par rapport à l'atlas moyen n'a pas la même ampleur selon les structures. Cela est dû au fait que certaines structures sont plus contrastées que d'autres et la sélection n'est donc pas aussi robuste pour toutes les structures.

Nous présentons sur la figure 6.14 les résultats obtenus structure par structure pour la combinaison optimale citée plus haut. Pour les structures symétriques, seule une des deux est présentée. Cette figure montre bien que la dynamique n'est pas la même pour toutes les structures. L'amélioration par rapport à l'atlas moyen est davantage prononcée pour la parotide, la moelle et le tronc cérébral (le maximum des courbes est peu marqué car nous avons choisi une échelle commune à toutes les structures pour l'axe des ordonnées). Pour les aires ganglionnaires en revanche l'amélioration est moins importante. Cela peut s'expliquer par le fait que les aires ganglionnaires présentent une variabilité de contourage manuel plus importante que les autres structures, et donc un atlas similaire au patient en terme d'intensité sur cette zone peut malgré tout présenter un contourage assez différent de celui du patient. Par ailleurs, le nombre optimal de segmentations à fusionner n'est pas le même pour toutes les structures : 15/46 pour l'AG. II droite, 10/71 pour l'AG. III gauche, 15/76 pour l'AG. IV droite, 10/102 pour la parotide gauche, 4/26 pour

la sous-maxillaire gauche, 4/10 pour la mandibule, 10/88 pour la moelle et 10/68 pour le tronc cérébral. A noter que ces nombres optimaux sont en partie dépendants du nombre total de segmentations disponibles pour chaque structure, puisque l'on se restreint lors de la sélection pour une structure donnée aux atlas qui ont cette structure segmentée. Enfin, on remarque que l'impact des structures manquantes pour les sélections de type DICE sur la correction de chevauchement (valeur de convergence des courbes rouges inférieure aux autres courbes) est davantage marqué pour les aires ganglionnaires que pour les autres structures, ce qui est cohérent car leurs jonctions représentent une surface non négligeable.

Nous avons également comparé structure par structure les performances présentées sur ces graphiques (sélection régionale) à celles obtenues avec une sélection globale (résultats non présentés ici). Pour l'ensemble des structures à l'exception de la moelle, les résultats sont meilleurs avec une sélection régionale. Pour la moelle, la sélection globale permet d'atteindre une meilleure performance que la sélection régionale (avec une propagation directe pour les deux). Cela peut s'expliquer par le fait que les différences d'anatomie au niveau de la moelle sont souvent à l'origine d'échec de recalage. Ces échecs de recalage ne sont pas localisés au niveau de la moelle mais se répercutent sur toute l'image, ce qui peut expliquer que la sélection globale soit relativement pertinente pour la moelle.

6.3.2 Évaluation dans le cadre d'une pondération locale

Dans le paragraphe précédent, nous avons montré que la sélection régionale permettait de fournir de meilleurs résultats que la sélection globale. Cependant, la sélection régionale ne permet pas de gérer le fait qu'un atlas puisse être similaire au patient sur une partie de la structure considérée mais pas sur le reste. Les structures d'intérêt de la sphère ORL ne sont pas de toutes petites structures, donc la prise en compte de ce genre de considérations est importante. Pour cela, nous avons mis en œuvre une approche de pondération locale des atlas lors de la fusion des segmentations dans le référentiel du patient (cf. détails en section 5.4.4.2).

Nous avons étudié l'effet des paramètres suivants :

- propagation : directe ou indirecte,
- mesure de similarité utilisée : CC, SAD, ou SSD,
- voisinage de calcul de la mesure de similarité : $3 \times 3 \times 3$, $5 \times 5 \times 5$, ou $7 \times 7 \times 7$,
- exposant utilisé pour amplifier les différences entre les poids : 1, 2, 3, 4, 5, 6 pour CC ; -1, -2, -3, -4, -5, -6 pour SSD et SAD,
- lissage de la carte de probabilité ou pas,
- seuil utilisé pour seuiller les cartes de probabilités.

Que ce soit pour la propagation directe ou indirecte, la meilleure configuration au regard de ces paramètres s'est révélée être l'utilisation de la SSD (calculée sur un voisinage de taille $3 \times 3 \times 3$ et élevée à l'exposant -5), avec un lissage additionnel des cartes de probabilités (écart-type du filtre gaussien égal à 1), un seuillage à 0.35 et une fermeture morphologique appliquée à chaque segmentation binaire avant gestion des chevauchements. Les meilleurs résultats sont obtenus avec la propagation directe.

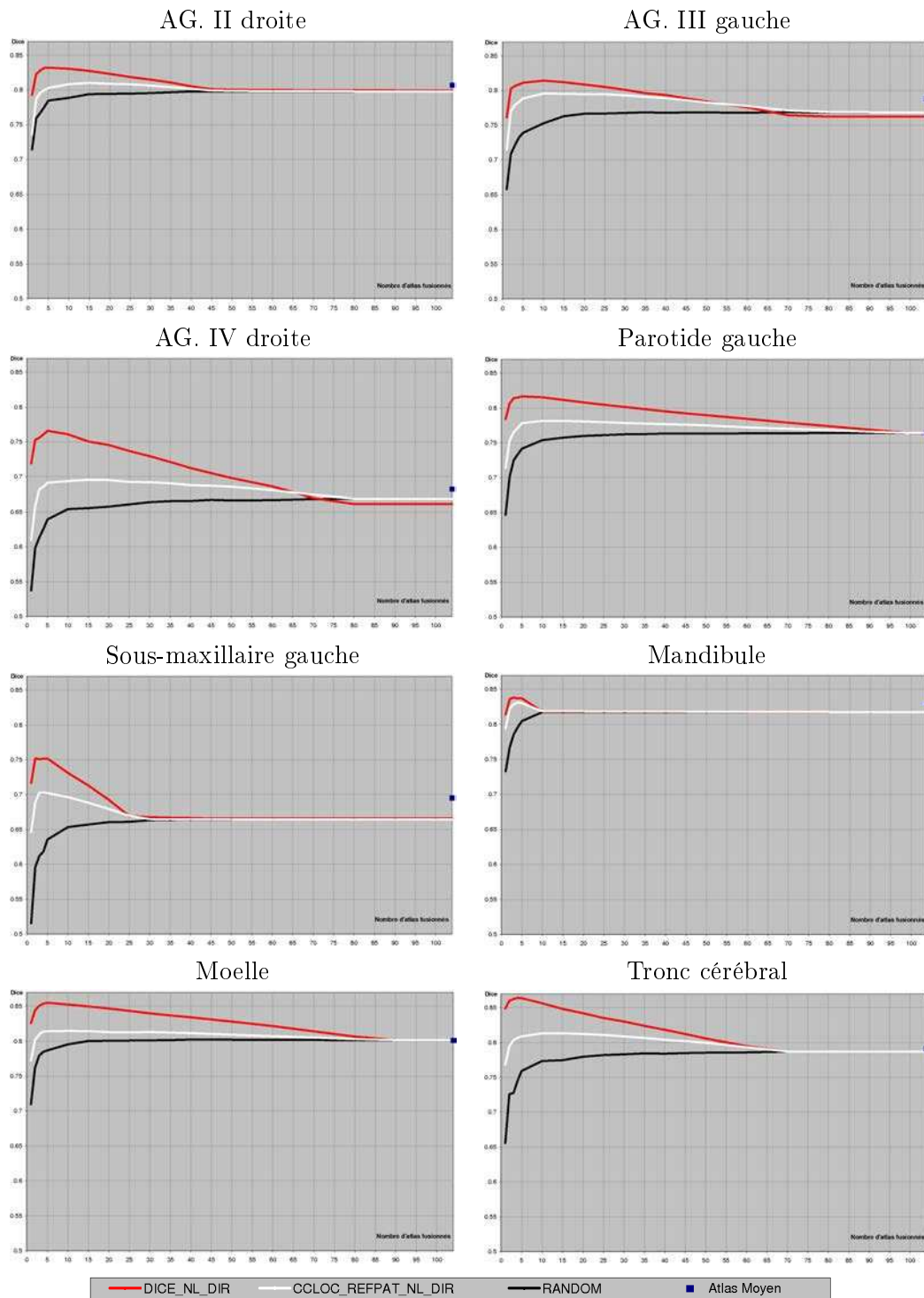


FIG. 6.14 – Pour chaque structure, indice de DICE moyen sur tous les patients en fonction du nombre de segmentations d'atlas fusionnées dans le référentiel du patient après sélection régionale selon CCLOC_REFPAT_NL_DIR et propagation directe des segmentations des atlas.

L'algorithme a été utilisé pour la segmentation des parotides dans le challenge de segmentation tenu en conjonction de la conférence MICCAI'2010. Pour cette application, nous avons utilisé une propagation indirecte (pour des raisons de temps de calcul), et nous avons mis en évidence une amélioration significative par rapport à l'atlas moyen (non présenté ici).

6.3.3 Comparaison quantitative

Les sections précédentes ont permis de mieux appréhender les effets des différents paramètres et de déterminer pour chacune des approches de fusion des segmentations dans le référentiel du patient (globale, régionale et locale) les paramètres optimaux donnant les meilleurs résultats (entre autres, critère de sélection/pondération d'atlas et nombre d'atlas à sélectionner pour les approches globale et régionale). A partir de ces configurations optimales pour chaque approche, notre objectif ici est double.

Notre premier objectif est de quantifier l'effet de l'échelle de la sélection (globale/régionale/locale) sur la qualité de segmentation résultante. Notre second objectif est de quantifier l'amélioration apportée par la propagation directe, tant pour la phase de sélection que pour la phase de fusion (toutes deux étant faites dans le référentiel du patient). En effet, nous avons vu qu'elle permettait en général d'améliorer les résultats par rapport à la propagation indirecte, mais comme elle représente un surcoût important en temps de calcul, il est important de regarder dans quelle mesure l'amélioration est significative ou ne l'est pas. Pour cela, en notant SIM la mesure de similarité utilisée pour la sélection, on compare trois cas de figures :

1. sélection dans le référentiel du patient après propagation directe (c'est-à-dire SIM_REFPAT_NL_DIR), puis fusion des segmentations dans le référentiel du patient après propagation directe : ce cas de figure, abrégé DIR_DIR dans la suite, correspond d'après nos premières analyses au cas optimal pour la sélection régionale et pour la pondération locale ;
2. sélection dans le référentiel du patient après propagation indirecte (c'est-à-dire SIM_REFPAT_NL_IND), puis fusion des segmentations dans le référentiel du patient après propagation directe : dans ce cas de figure, abrégé IND_DIR dans la suite, une approximation est faite pour éviter de recalculer tous les atlas lors de la phase de sélection mais chaque atlas sélectionné doit ensuite être recalculé directement avec le patient pour la phase de fusion ;
3. sélection dans le référentiel du patient après propagation indirecte (c'est-à-dire SIM_REFPAT_NL_IND), puis fusion des segmentations dans le référentiel du patient après propagation indirecte : ce cas de figure, abrégé IND_IND dans la suite, est le plus rapide puisqu'il ne nécessite qu'un seul recalage non-linéaire (celui du patient avec l'atlas moyen).

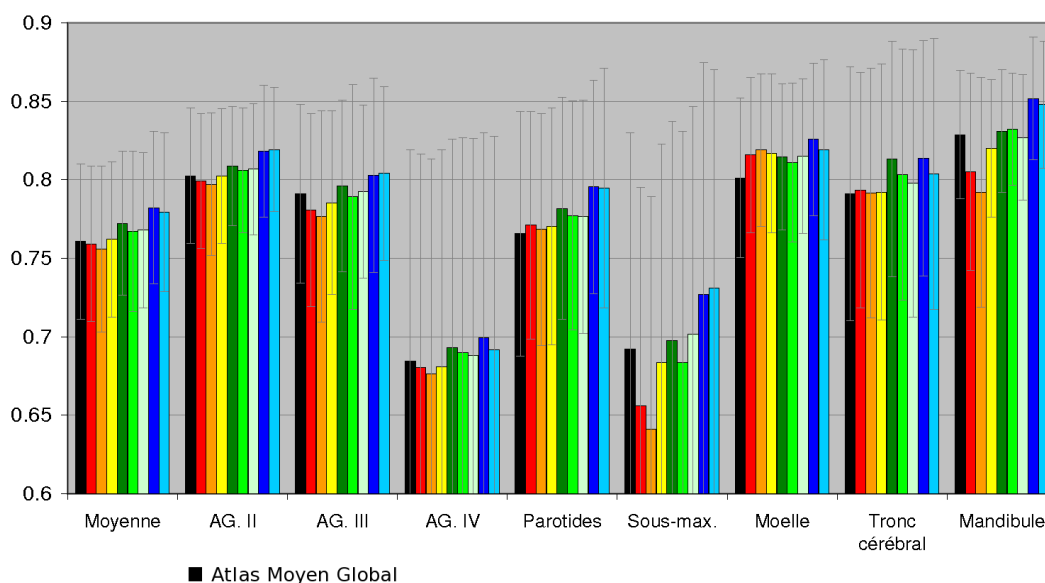
Le cas de figure DIR_IND (propagation directe pour la sélection et indirecte pour la fusion) n'est pas étudié car non pertinent : il donne de moins bons résultats que le cas de figure 1 pour un même temps de calcul. La pondération locale ne s'inscrit

que dans les cas de figures 1 et 3 (puisque les deux étapes (sélection/fusion) sont combinées en une seule (pondération)).

Les résultats obtenus en terme d'indice de DICE moyen sur tous les patients avec ces différentes approches sont comparés sur la figure 6.15. Pour chaque approche, le jeu de paramètre optimal est utilisé :

- Sélection globale : $SIM=CLOC$, et nombres optimaux d'atlas sélectionnés pour chaque cas de figure (40, 30, 45 pour les cas de figures 1, 2 et 3 respectivement) ;
- Sélection régionale : $SIM=CLOC$, et nombres optimaux de segmentations sélectionnées pour chaque structure et pour chaque cas de figure (en autorisant cette fois-ci un nombre différent pour chaque structure) ;
- Pondération locale : $SIM=SSD$, voisinage= $3 \times 3 \times 3$, exposant=5.

Dans tous les cas, un seuil de 0.35 est utilisé pour seuiller les cartes de probabilités (avec un filtre gaussien d'écart-type 1). Dans la suite, toutes les valeurs de p-values mentionnées concernent l'indice de DICE moyen sur toutes les structures.



Différentes approches pour la sélection/fusion des segmentations dans le référentiel du patient		Type de propagation pour la sélection / pondération dans le référentiel du patient	
		Directe	Indirecte
Type de propagation pour la fusion des segmentations dans le référentiel du patient	Directe	(DIR_DIR) ■ Sélection Globale ■ Sélection Régionale ■ Pondération Locale	(IND_DIR) ■ Sélection Globale ■ Sélection Régionale
	Indirecte	X	(IND_IND) ■ Sélection Globale ■ Sélection Régionale ■ Pondération Locale

FIG. 6.15 – Indice de DICE moyen sur tous les patients pour les différentes approches de fusion des segmentations dans le référentiel du patient (les valeurs moyennées sur les 105 patients sont les DICE moyens sur toutes les structures (première série) ainsi que les DICE pour chaque structure indépendamment).

Tout d’abord, si l’on compare les approches globale/régionale/locale, on se rend compte que quel que soit le cas de figure étudié pour la propagation, les résultats sont d’autant meilleurs que la sélection/pondération est ciblée (globale (tons feu) < régionale (tons verts) < locale (tons bleus)), et les différences sont significatives (toutes les p-values sont inférieures à 10^{-15}). La seule exception est la moelle pour laquelle la sélection globale donne de meilleurs résultats que la sélection régionale, comme nous l’avions déjà mentionné précédemment. En moyenne sur toutes les structures, les différences entre l’approche globale et l’atlas moyen sont soit non significatives, soit en faveur de l’atlas moyen. L’approche globale n’est donc pas suffisante, comme nous l’avions déjà remarqué. Pour obtenir une amélioration significative par rapport à l’atlas moyen, il faut au moins cibler la sélection d’atlas au niveau structure par structure (p-values toutes inférieures à 10^{-7}), et les résultats sont encore améliorés avec l’approche locale (p-values toutes inférieures à 10^{-30}).

Si l’on compare maintenant les trois cas de figure en terme de propagation directe/indirecte pour les approches régionale et locale, le meilleur cas de figure est le cas DIR_DIR, comme nous l’avions déjà remarqué pour la sélection régionale, et les différences sont significatives par rapport aux deux autres cas. Ainsi, avec les approches régionale et locale, toute économie en terme de temps de calcul se fera au détriment de la qualité de segmentation puisque les différences selon DICE sont significatives. Pour la sélection régionale, si l’on souhaite tout de même réduire le temps de calcul, le cas de figure le plus rapide (IND_IND) est préconisé puisqu’il donne des résultats similaires au cas IND_DIR qui est plus long. Bien que significatives (p-value = 0.0075) les différences entre propagation indirecte et directe pour la pondération locale sont relativement faibles.

Pour conclure, l’approche de fusion des segmentations dans le référentiel du patient permet d’améliorer significativement les résultats de l’atlas moyen. L’approche la plus efficace est la pondération locale des segmentations après propagation directe dans le référentiel du patient. Cette approche nécessite cependant le recalage non-linéaire de chaque image de la base sur le patient, ce qui peut vite devenir rédhibitoire en temps de calcul dès que la base atteint une certaine taille. Dans l’objectif de réduire le temps de calcul, une propagation indirecte via l’atlas moyen ne nécessite plus qu’un seul recalage non-linéaire à faire en ligne, et les performances ne sont que faiblement réduites. En conclusion, l’approche représentant le meilleur compromis entre qualité de segmentation et temps de calcul est la pondération locale des segmentations après propagation indirecte dans le référentiel du patient. Cette approche a donc été intégrée dans le logiciel ISOgray de DOSIsoft, et également choisie pour le challenge de segmentation au workshop MICCAI.

6.3.4 Comparaison qualitative

Des résultats visuels de segmentations sont présentés pour la moelle épinière et le tronc cérébral (cf. figures 6.16 et 6.17), pour les parotides (cf. figures 6.18 et 6.19), pour les aires ganglionnaires II (cf. figure 6.20) et IV (cf. figure 6.21), et pour les sous-maxillaires et la mandibule (cf. figure 6.22). Sur ces figures, les segmentations

manuelles sont représentées en rouge, et les segmentations automatiques obtenues avec les trois méthodes suivantes sont comparées :

- l'atlas moyen : en bleu, première ligne de chaque figure ;
- la fusion des segmentations dans le référentiel du patient après sélection régionale (avec une propagation directe pour les phases de sélection et de fusion c'est-à-dire DIR_DIR) : en blanc, deuxième ligne de chaque figure ;
- la pondération locale des segmentations dans le référentiel du patient après propagation directe : en jaune, troisième ligne de chaque figure.

Pour chacune de ces trois méthodes automatiques, nous avons utilisé le jeu de paramètres optimal précédemment détaillé. La dernière ligne de chaque figure regroupe la totalité des contours pour comparaison, et des flèches indiquent les différences principales. Enfin, la figure 6.23 regroupe les résultats visuels des différentes structures.

Tout d'abord, on constate que pour la moelle épinière et le tronc cérébral (figures 6.16 et 6.17), les approches de sélection régionale et de pondération locale permettent de segmenter la moelle (respectivement le tronc cérébral) jusqu'en bas (respectivement jusqu'en haut), alors qu'un certain nombre de coupes sont manquantes pour la segmentation automatique obtenue avec l'atlas moyen. Cela vient du fait que les images de notre base ont des champs d'acquisition différents. Lors de la construction de l'atlas moyen, toutes les segmentations de la base sont moyennées quel que soit leur champ de vue, et la segmentation moyenne résultante a des coupes manquantes. Au contraire, la sélection/pondération des atlas d'après les intensités permet de pénaliser pour la moelle et le tronc cérébral les images de la base qui ont un champ de vue différent de celui du patient, ce qui permet d'estimer des segmentations complètes pour ces deux structures.

Pour les parotides (figures 6.18 et 6.19), la sélection régionale et la pondération locale des atlas permettent d'améliorer la segmentation de façon significative, en particulier au niveau des lobes accessoires et des lobes profonds ainsi que sur l'extrémité postérieure des structures (zones indiquées par les flèches).

Pour les aires ganglionnaires II (figure 6.20), les différences entre les trois jeux de contours sont principalement localisées sur la partie antérieure latérale des structures (pointées par les flèches). Pour les aires ganglionnaires IV (figure 6.21), les différences sont surtout situées sur les trois extrémités repérées par les flèches.

Enfin, le cas de la mandibule présenté en figure 6.22 montre que le bloc d'immobilisation a fortement perturbé le recalage unique entre l'image du patient et celle de l'atlas moyen. Avec les deux autres méthodes, la propagation directe a été utilisée, et la fusion dans le référentiel du patient de plusieurs segmentations issues de recalages indépendants a ici permis d'être beaucoup plus robuste au bloc d'immobilisation.

L'ensemble des figures montrent que la pondération locale fournit des contours plus proches des contours manuels que la sélection régionale, qui fournit elle-même des contours plus précis que l'atlas moyen global. De plus, pour ne pas surcharger, nous n'avons pas représenté les contours obtenus avec la méthode de pondération locale des segmentations dans le référentiel du patient après propagation indirecte. Cependant, les différences avec la méthode de pondération locale après propagation directe (contours jaunes) sont assez marginales, et ce sur la plupart des patients.

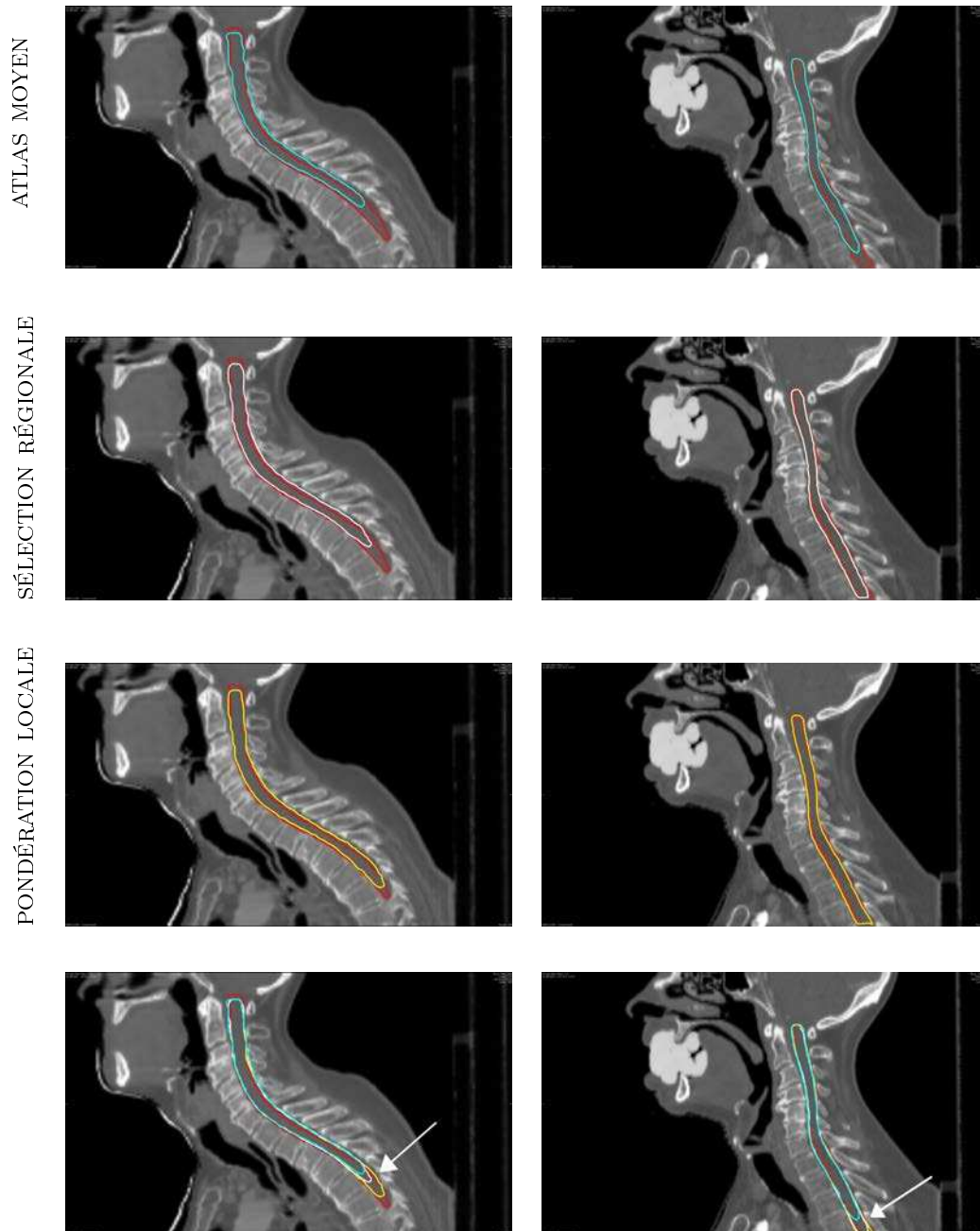


FIG. 6.16 – Résultats visuels de segmentations sur la moelle épinière pour deux patients différents : les contours automatiques obtenus avec l’atlas moyen (en bleu, première ligne), la sélection régionale (en blanc, deuxième ligne) et la pondération locale (en jaune, troisième ligne) sont comparés aux contours manuels (en rouge). La dernière ligne montre l’ensemble des contours.

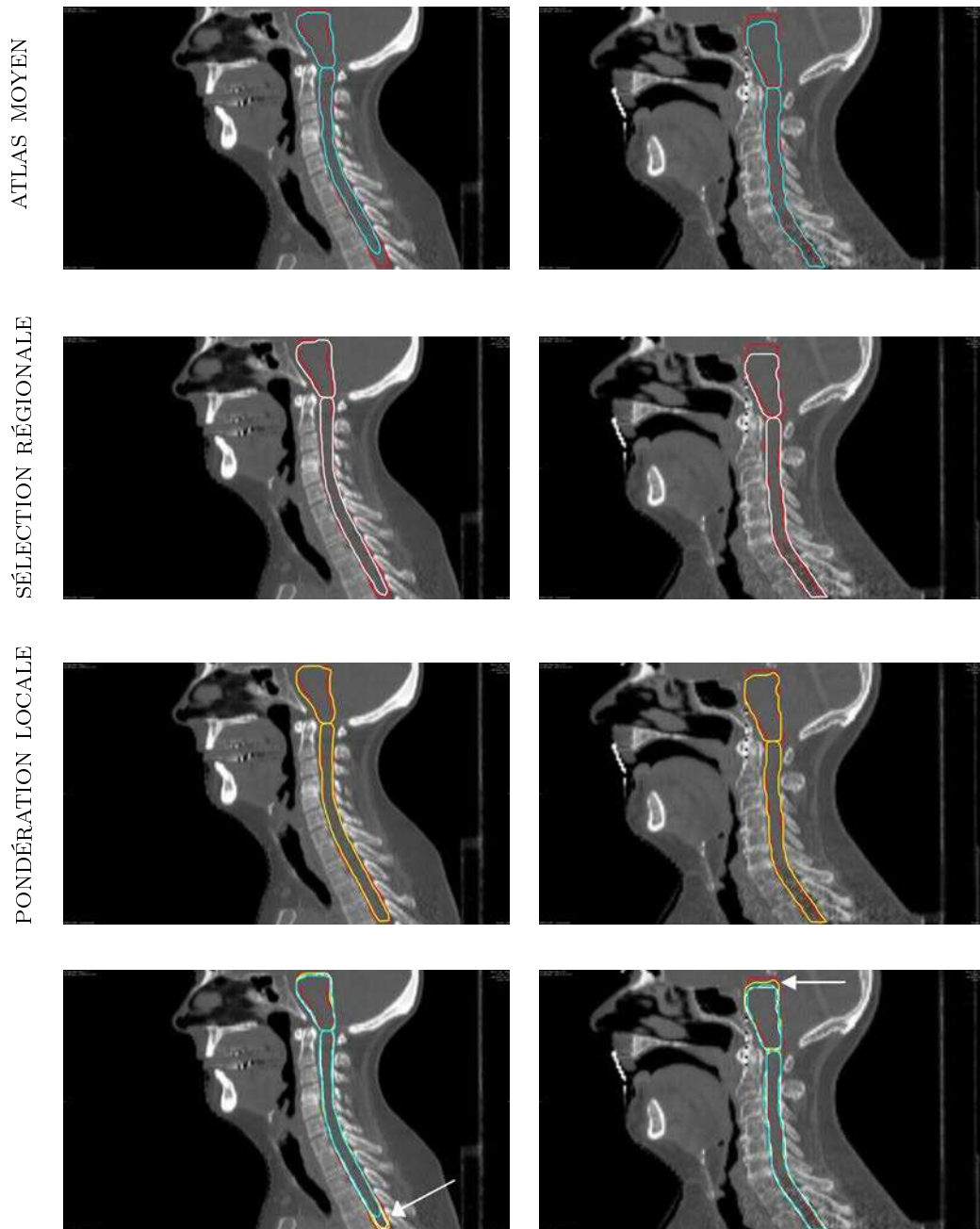


FIG. 6.17 – Résultats visuels de segmentations sur la moelle épinière et le tronc cérébral pour deux patients différents : les contours automatiques obtenus avec l’atlas moyen (en bleu, première ligne), la sélection régionale (en blanc, deuxième ligne) et la pondération locale (en jaune, troisième ligne) sont comparés aux contours manuels (en rouge). La dernière ligne montre l’ensemble des contours.

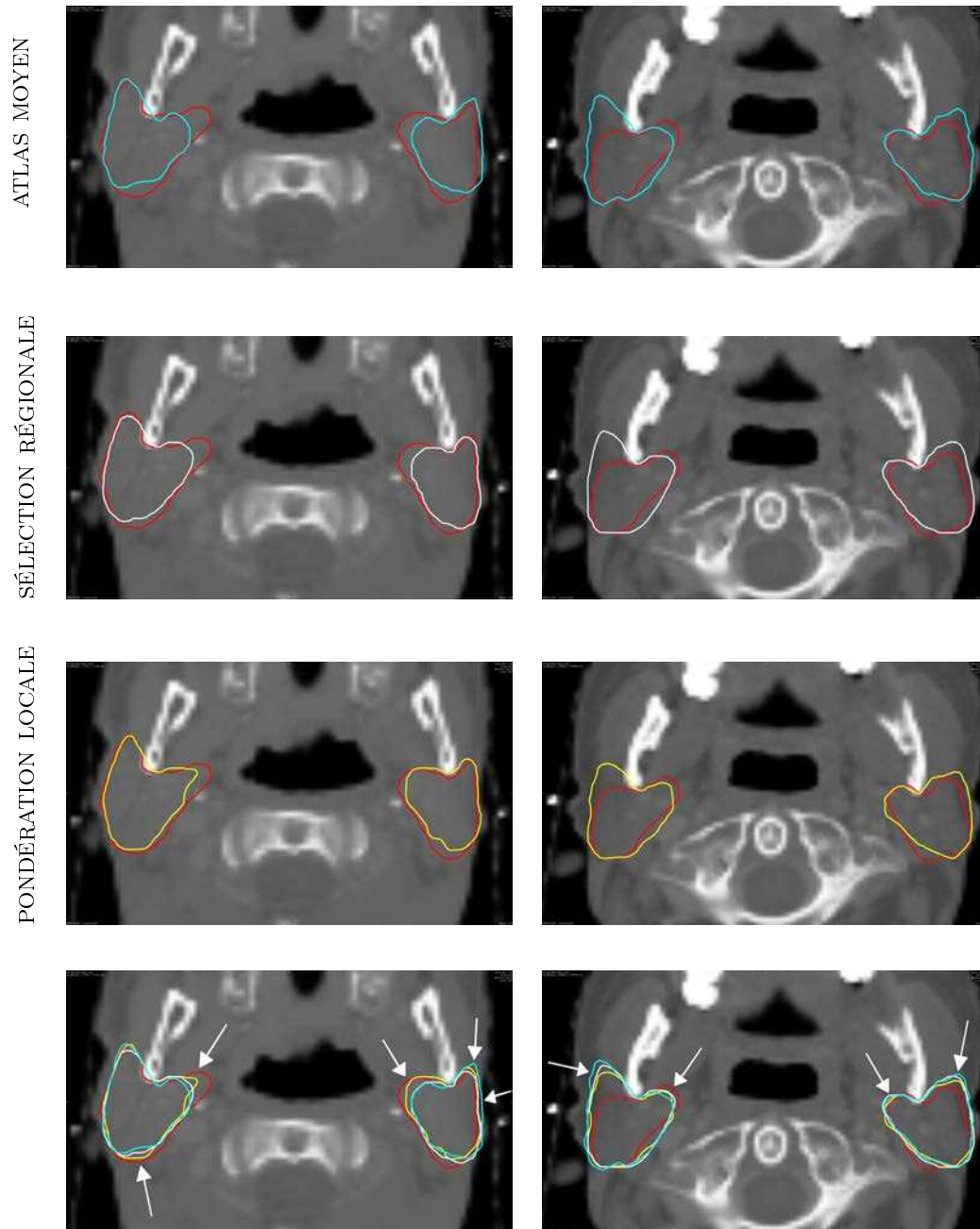


FIG. 6.18 – Résultats visuels de segmentations sur les parotides pour deux patients différents : les contours automatiques obtenus avec l’atlas moyen (en bleu, première ligne), la sélection régionale (en blanc, deuxième ligne) et la pondération locale (en jaune, troisième ligne) sont comparés aux contours manuels (en rouge). La dernière ligne montre l’ensemble des contours.

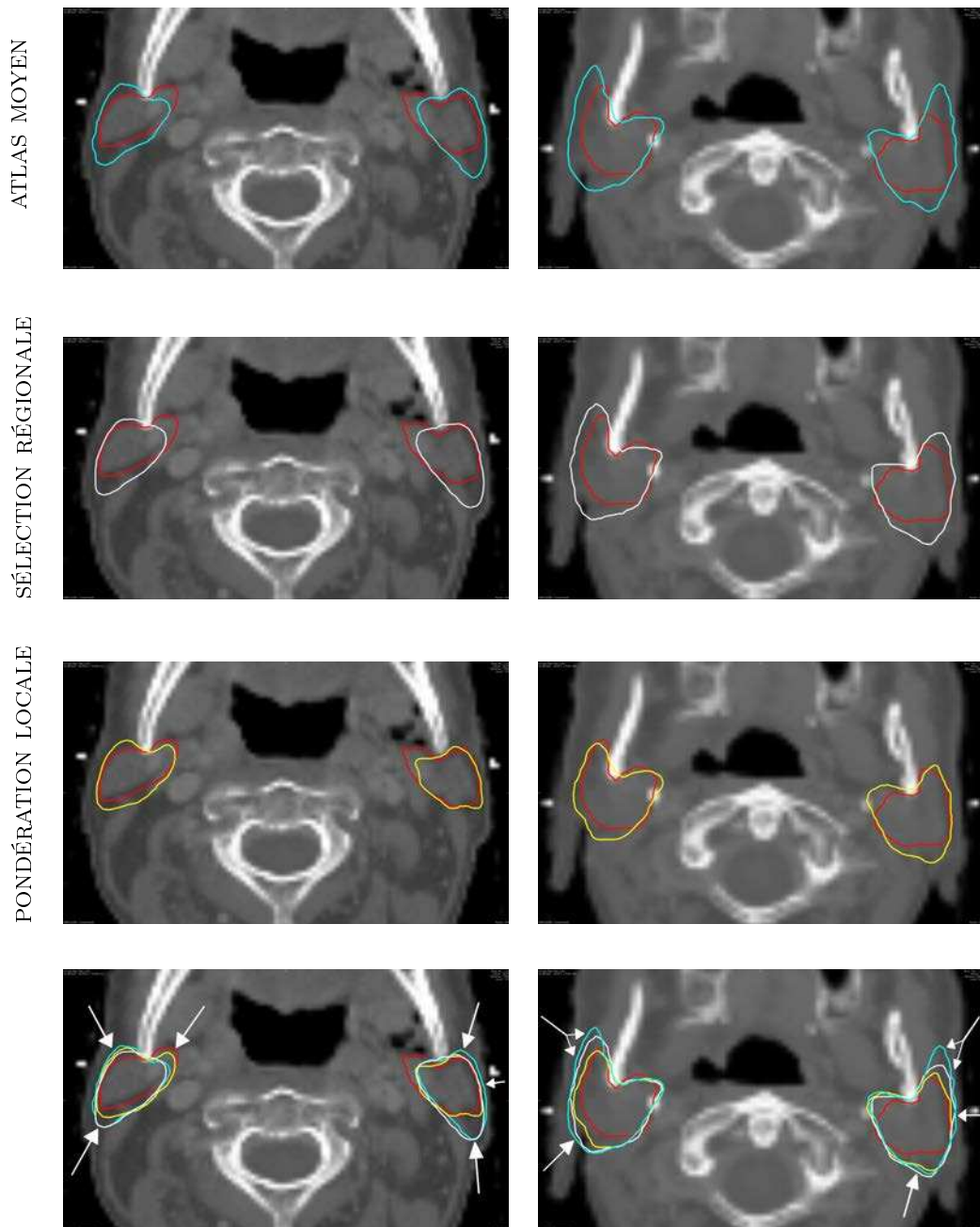


FIG. 6.19 – Résultats visuels de segmentations sur les parotides pour deux patients différents : les contours automatiques obtenus avec l’atlas moyen (en bleu, première ligne), la sélection régionale (en blanc, deuxième ligne) et la pondération locale (en jaune, troisième ligne) sont comparés aux contours manuels (en rouge). La dernière ligne montre l’ensemble des contours.

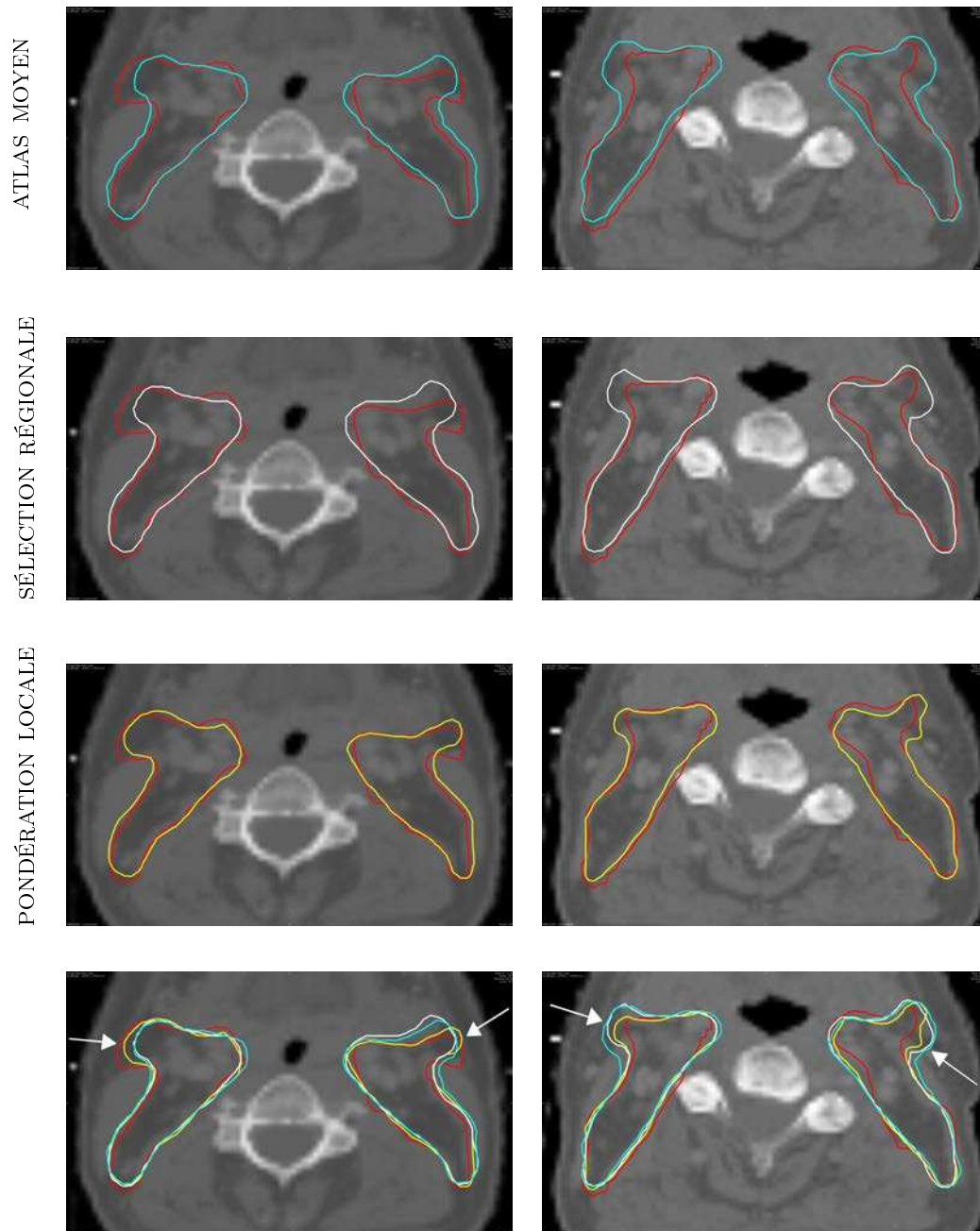


FIG. 6.20 – Résultats visuels de segmentations sur les aires ganglionnaires II pour deux patients différents : les contours automatiques obtenus avec l’atlas moyen (en bleu, première ligne), la sélection régionale (en blanc, deuxième ligne) et la pondération locale (en jaune, troisième ligne) sont comparés aux contours manuels (en rouge). La dernière ligne montre l’ensemble des contours.

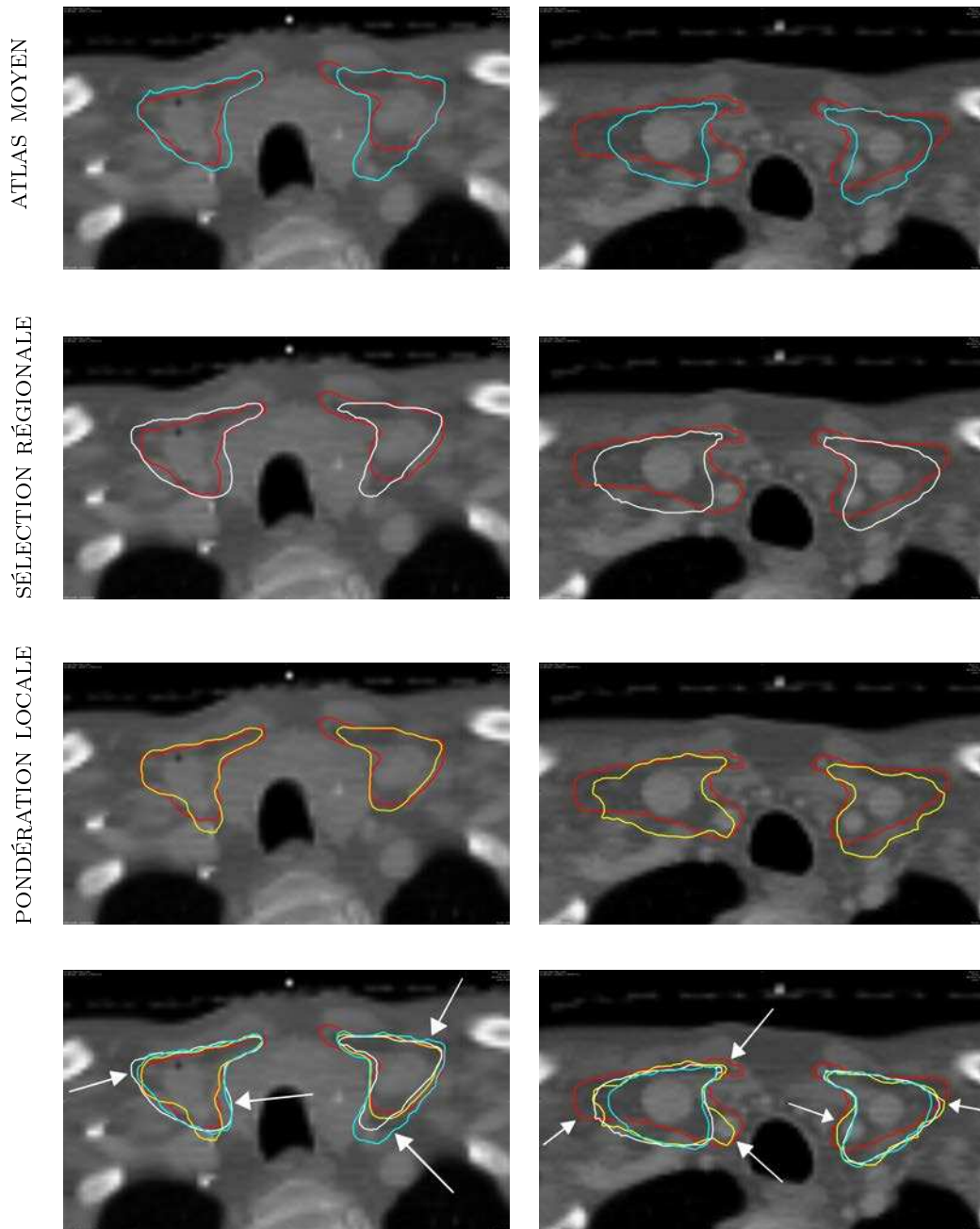


FIG. 6.21 – Résultats visuels de segmentations sur les aires ganglionnaires IV pour deux patients différents : les contours automatiques obtenus avec l’atlas moyen (en bleu, première ligne), la sélection régionale (en blanc, deuxième ligne) et la pondération locale (en jaune, troisième ligne) sont comparés aux contours manuels (en rouge). La dernière ligne montre l’ensemble des contours.

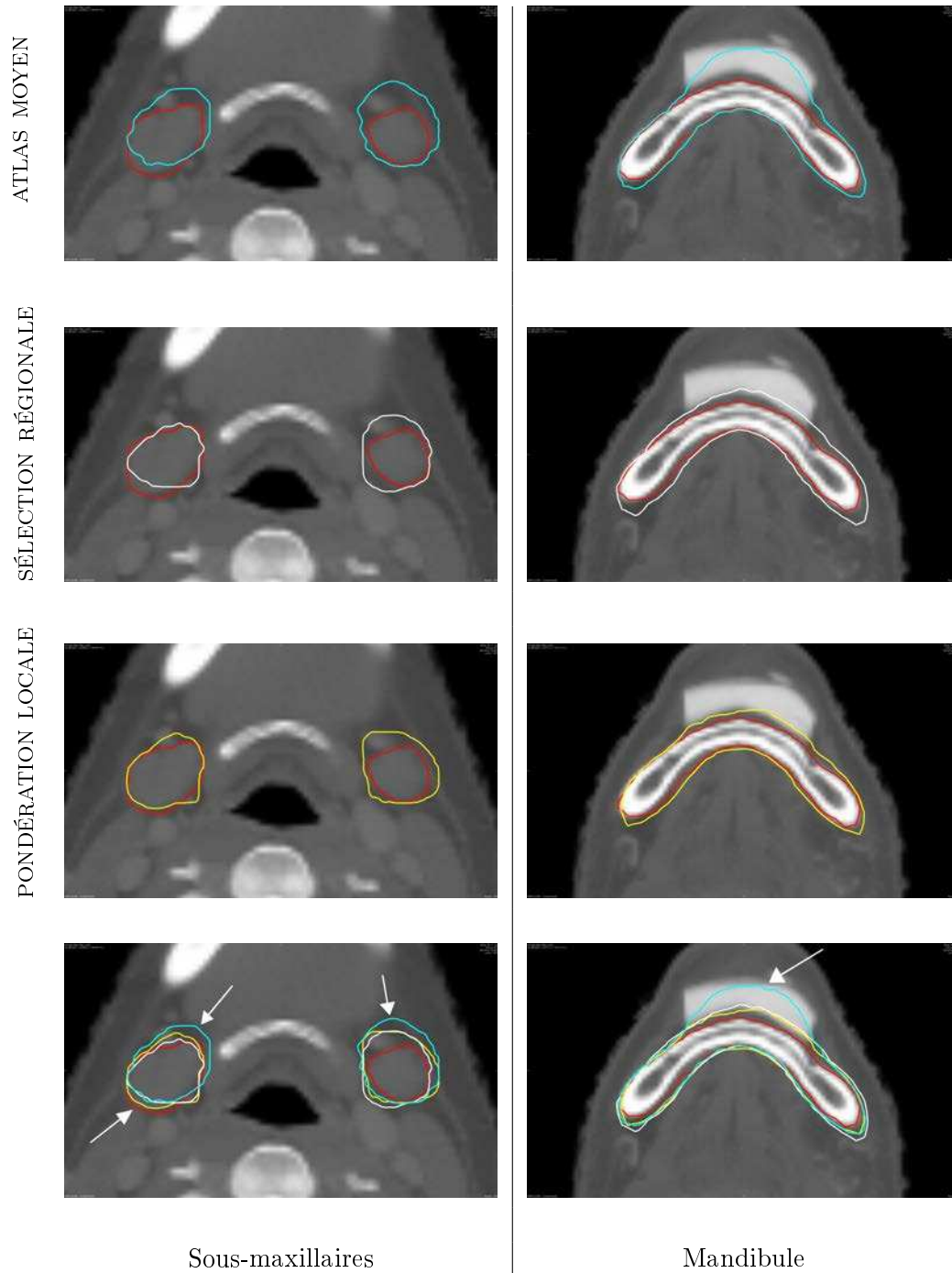


FIG. 6.22 – Résultats visuels de segmentations sur les sous-maxillaires (première colonne) et la mandibule (deuxième colonne) : les contours automatiques obtenus avec l’atlas moyen (en bleu, première ligne), la sélection régionale (en blanc, deuxième ligne) et la pondération locale (en jaune, troisième ligne) sont comparés aux contours manuels (en rouge). La dernière ligne montre l’ensemble des contours.

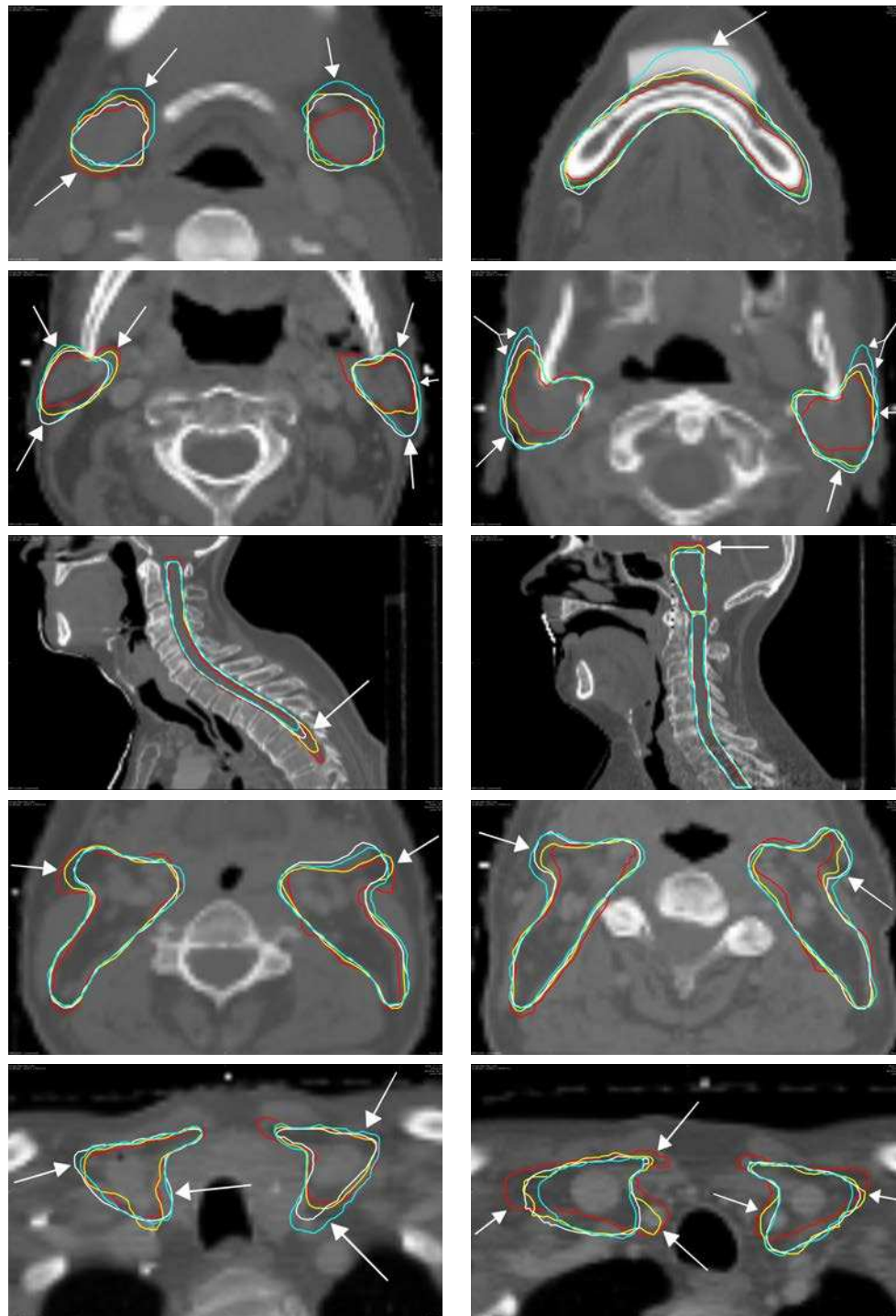


FIG. 6.23 – Résultats visuels de segmentations automatiques obtenues avec l'atlas moyen (en bleu), avec la sélection régionale (en blanc) et avec la pondération locale (en jaune). Les contours manuels sont représentés en rouge.

6.4 Création à la volée d'un atlas composite adapté au patient

Dans cette section, nous exposons les résultats obtenus avec la création à la volée d'un atlas moyen composite adapté au patient proposée en section 5.4.3.3.

6.4.1 Paramètres

L'algorithme proposé est basé sur une sélection régionale des atlas les plus similaires au patient. Pour choisir les paramètres relatifs à cette première étape, nous nous basons sur les conclusions de la section précédente faites dans le cas de la fusion des segmentations dans le référentiel du patient. Nous avons alors montré que dans le cadre d'une sélection régionale, la méthode de sélection qui est la plus efficace est `CCLOC_REFPAT_NL_DIR`. Notre objectif est donc de regarder dans quelle mesure, l'algorithme de création à la volée d'un atlas moyen composite permet, avec cette méthode de sélection, de dépasser les performances atteintes avec la fusion des segmentations dans le référentiel du patient.

Nous nous basons également sur nos conclusions précédentes pour déterminer le nombre d'atlas à sélectionner par région (figure 6.13). La valeur retenue est de 15 atlas sélectionnés pour chaque région. Concernant la pondération gaussienne, nous fixons l'écart-type de la gaussienne à 0.1. Cette valeur a été déterminée sur un sous-ensemble de trois patients de façon à ce que, quelle que soit la région considérée, le ratio entre le poids de l'atlas le plus similaire au patient parmi les 15 sélectionnés et le poids de l'atlas le moins similaire au patient parmi les 15 sélectionnés soit au maximum de l'ordre de grandeur 10% - 5% (c'est-à-dire l'atlas le plus similaire parmi les 15 compte deux fois plus que l'atlas le moins similaire parmi les 15).

Enfin, nous définissons une région par structure sauf pour les sous-maxillaires. En effet, les sous-maxillaires sont des structures relativement petites et elles sont de plus relativement proches d'un certain nombre d'autres structures (mandibule, aires ganglionnaires II et parotides). Ces deux éléments combinés posent problème pour assurer une distance minimale entre les régions. Nous choisissons donc de ne pas définir de région sur les sous-maxillaires.

Nous utilisons le même nombre d'itérations que pour la construction de l'atlas moyen global selon l'algorithme de Guimond et al., c'est-à-dire 4. Pour le calcul de la segmentation moyenne, nous utilisons un filtre gaussien d'écart-type 1 et un seuil de 0.35. Ces valeurs sont les mêmes que celles utilisées pour l'ensemble des méthodes, et ce sont également celles qui donnent les meilleurs résultats.

6.4.2 Résultats

Pour l'évaluation, nous excluons les sous-maxillaires puisqu'aucune région n'y est associée. De plus, en raison du faible nombre de segmentations de mandibule dans notre base (11), nous avons préféré ne pas nous restreindre aux atlas ayant cette structure segmentée lors de la sélection pour cette région, ceci afin de ne pas

sélectionner des atlas trop peu similaires au patient qui pourraient dégrader les résultats artificiellement lors du calcul de l'image composite. De ce fait, nous disposons souvent d'aucune segmentation manuelle pour la mandibule, ou en tout cas d'un nombre réduit qui conduirait à une estimation très peu robuste de la segmentation. Pour cette raison, nous excluons également la mandibule de l'évaluation.

6.4.2.1 Résultats quantitatifs

Afin d'évaluer notre algorithme de création d'atlas moyen composite adapté au patient, nous le comparons à deux méthodes de fusion des segmentations dans le référentiel du patient :

- l'approche de sélection régionale avec propagation directe pour la sélection et la fusion (mode DIR_DIR, cf. section 6.3.3 pour les détails),
- l'approche de pondération locale, en propagation directe ou indirecte.

Les résultats en terme d'indice de DICE sont présentés sur la figure 6.24.

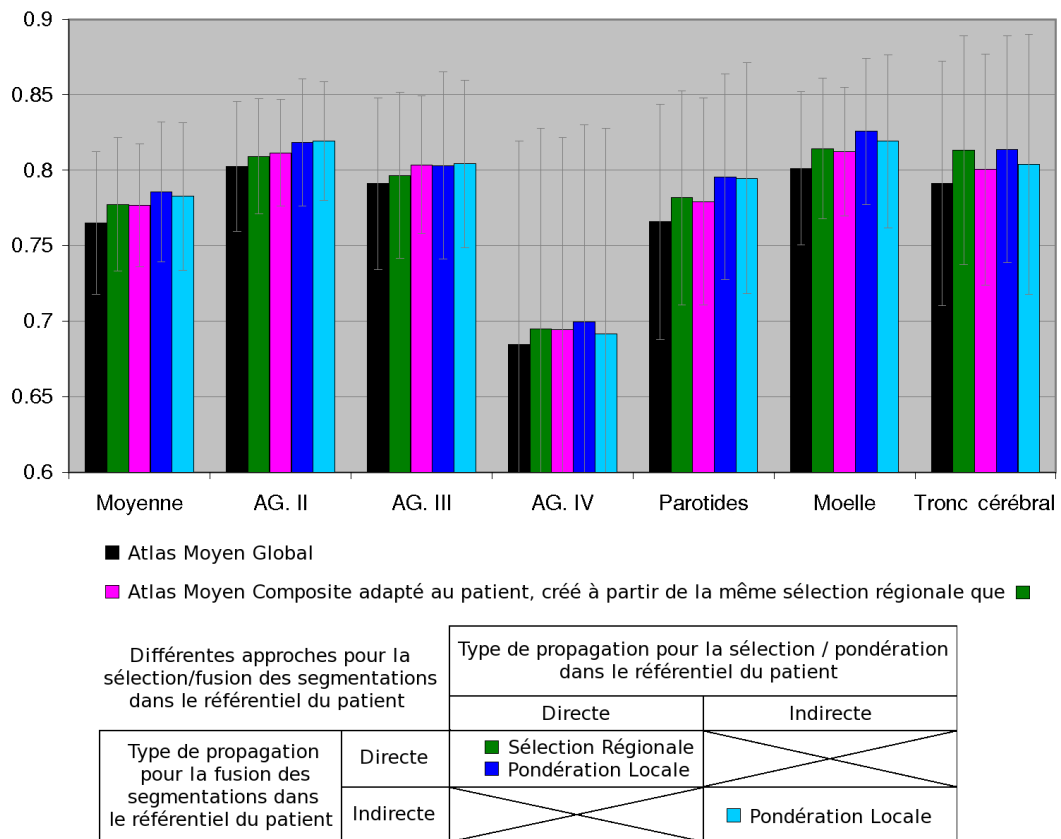


FIG. 6.24 – Indice de DICE moyen sur tous les patients obtenu avec l'algorithme de création d'atlas moyen composite adapté au patient (en rose), avec l'atlas moyen global (en noir) et avec deux approches de fusion des segmentations dans le référentiel du patient (sélection régionale en vert, pondération locale en bleu).

Tout d’abord, on remarque que notre algorithme permet d’améliorer significativement les résultats par rapport à l’atlas moyen global (p-value de $2 \cdot 10^{-13}$). Cela est l’effet combiné de la sélection des images des atlas (qui permet de construire une image moyenne composite plus similaire au patient que l’atlas moyen global, et donc d’optimiser le recalage non-linéaire avec le patient) et de la sélection des segmentations correspondantes. D’autre part, à critère de sélection identique (CCLOC_REFPAT_NL_DIR), les différences entre notre algorithme (rose) et la méthode de fusion des segmentations dans le référentiel du patient après sélection régionale et propagation directe (vert) sont non significatives (p-value 0.31). Cela signifie que l’image moyenne composite créée est approximativement aussi proche du patient sur une région donnée que l’ensemble des atlas sélectionnés sur cette région. Cela peut s’expliquer par le fait que la moyenne en intensité des atlas les plus similaires au patient faite lors de la construction du patient composite moyen a un double effet : (1) un effet positif puisque qu’elle gomme les détails anatomiques présents sur une minorité des atlas sélectionnés, et permet ainsi d’estimer de façon robuste un patient virtuel en moyenne plus proche du patient que l’ensemble des atlas, et (2) un effet négatif puisqu’elle introduit malgré tout un flou sur l’image composite qui peut potentiellement dégrader la précision du recalage. Enfin, les approches de pondération locale des segmentations (bleu) restent en conclusion les plus performantes.

Contrairement aux méthodes de fusion des segmentations dans le référentiel du patient, nous n’avons pas fait une étude exhaustive des jeux de paramètres optimaux, notamment : la méthode de sélection optimale, le nombre optimal d’atlas fusionnés pour chaque région, qui peut être ici différent de 15 car le contexte n’est pas le même (fusion d’images et plus seulement de segmentations), l’écart-type de la gaussienne utilisée pour la pondération des atlas sélectionnés, etc. Une étude exhaustive permettrait peut être d’optimiser les résultats présentés ici.

Par ailleurs, comme l’algorithme de calcul de l’image moyenne composite est relativement long, il serait intéressant d’accélérer au maximum la phase de sélection, et donc d’évaluer les performances obtenues pour une sélection dans le référentiel de l’atlas moyen par exemple (CCLOC_REFAV), ou encore dans le référentiel du patient mais avec une propagation indirecte (CCLOC_REFPAT_NL_IND).

Enfin, l’évaluation de l’algorithme avec une méthode de sélection de référence (DICE) permettrait de déterminer les bornes supérieures du schéma de fusion proposé en terme de performance, et de les comparer aux bornes supérieures des autres approches de fusion.

6.4.2.2 Résultats qualitatifs

La figure 6.25 compare pour quatre patients différents l’atlas moyen composite créé à la volée et adapté au patient à l’atlas moyen global. On remarque que après normalisation affine, l’atlas moyen global est assez différent de l’anatomie du patient, ce qui peut dégrader la qualité du recalage. Cela est particulièrement vrai pour les patients corpulents (patient 3) ou maigres (patients 1 et 2), et pour les patients ayant

une flexion de colonne particulière (patient 4). Les atlas moyens composites créés à la volée pour chaque patient sont quant à eux plus similaires aux patients, comme le mettent en évidence les curseurs rouges, ce qui permet d'optimiser les recalages non-linéaires avec les patients, et par conséquent la qualité des segmentations.

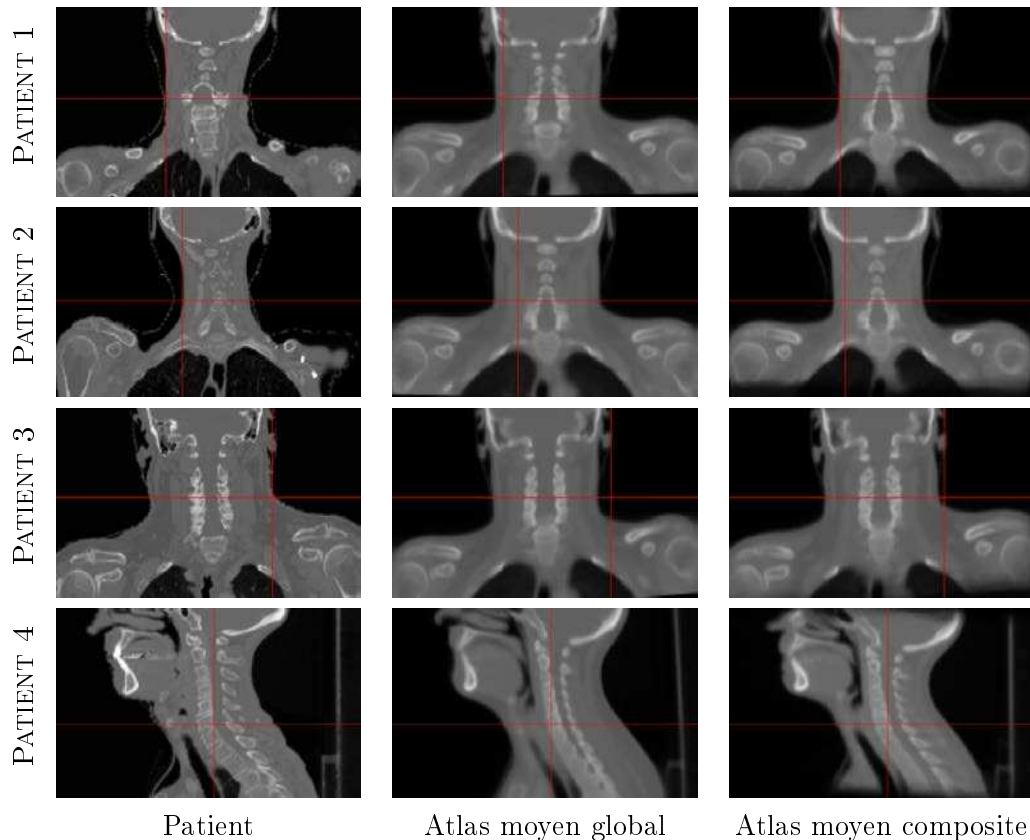


FIG. 6.25 – Comparaison entre l'atlas moyen global (deuxième colonne) et les atlas moyens composites créés à la volée (troisième colonne) pour quatre patients différents (première colonne). Pour chaque patient, l'atlas moyen et l'atlas composite ont été déformés linéairement sur le patient pour la visualisation.

6.4.2.3 Analyse patient par patient

En moyenne sur tous les patients, les résultats quantitatifs obtenus avec l'atlas moyen composite n'ont pas montré de différences significatives par rapport à l'approche de fusion dans le référentiel du patient après sélection régionale et propagation directe, et ils ont mis en évidence des résultats moins bons par rapport à la pondération locale. Cependant, si l'on regarde les résultats patient par patient, on se rend compte que pour un petit nombre de patients, les résultats sont bien meilleurs en utilisant l'atlas moyen composite. Ceci est mis en évidence sur la figure 6.26. Sur cette figure, les différences entre les indices de DICE moyens sur toutes les

structures obtenus avec différentes méthodes sont reportées :

- La première barre représente la boîte à moustache des différences $DICE(\text{AtMoyComp}) - DICE(\text{SelReg})$ entre les DICE moyens sur toutes les structures obtenus avec l'atlas moyen composite ($DICE(\text{AtMoyComp})$) et les DICE moyens sur toutes les structures obtenus avec l'approche de fusion dans le référentiel du patient après sélection régionale et propagation directe ($DICE(\text{SelReg})$).
- La seconde barre représente la boîte à moustache des différences $DICE(\text{AtMoyComp}) - DICE(\text{PondLoc})$ entre les DICE moyens sur toutes les structures obtenus avec l'atlas moyen composite ($DICE(\text{AtMoyComp})$) et les DICE moyens sur toutes les structures obtenus avec l'approche de pondération locale dans le référentiel du patient après propagation directe ($DICE(\text{PondLoc})$).

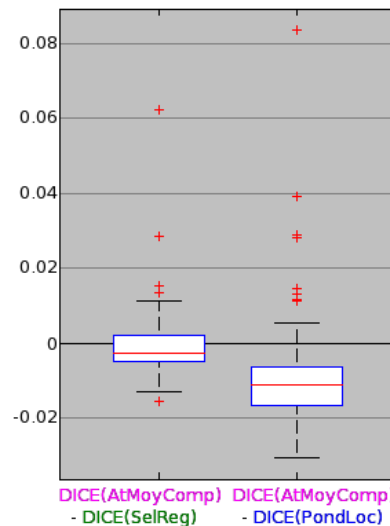


FIG. 6.26 – Boîtes à moustaches correspondant aux différences entre deux méthodes en terme de DICE moyen sur toutes les structures pour l'ensemble des 105 patients. Pour la première barre, les valeurs obtenues avec la sélection régionale et propagation directe sont soustraites aux valeurs obtenues avec l'atlas moyen composite. Pour la seconde barre, les valeurs obtenues avec la pondération locale après propagation directe sont soustraites aux valeurs obtenues avec l'atlas moyen composite.

Cette figure confirme qu'en moyenne l'utilisation de l'atlas moyen composite donne de moins bons résultats que la pondération locale (différences significativement négatives), et des résultats comparables par rapport à la sélection régionale (boîte à moustache approximativement centrée en zéro). Cependant, pour un faible nombre de patients, on constate que l'utilisation de l'atlas moyen composite donne de biens meilleurs résultats que les deux autres approches. Pour tous les patients qui sont mieux segmentés par les autres approches, une telle différence en terme de DICE n'est jamais obtenue.

En regardant plus en détail les patients particulièrement mieux segmentés avec l'atlas moyen composite qu'avec les deux autres approches, on se rend compte que ces patients correspondent à des patients ayant une anatomie très atypique. Trois de ces patients sont représentés sur la figure 6.27. Le premier est très corpulent (c'est le plus corpulent de notre base) et il n'a quasiment pas de cou, ce qui représente un réel défi pour la segmentation des aires ganglionnaires III et IV. Les deux autres patients sont très maigres (ce sont les plus maigres de notre base). La figure 6.28 montre pour le patient le plus corpulent (première colonne) et pour le patient le plus maigre (deuxième colonne) les segmentations automatiques obtenues avec l'atlas moyen (en bleu), avec les approches régionale (en blanc) et locale (en jaune) de fusion dans le référentiel du patient, et avec l'atlas moyen composite (en rose). Pour le patient le plus corpulent, on remarque que l'atlas moyen composite est la méthode qui arrive le mieux à gérer la position très enfoncée du cou.

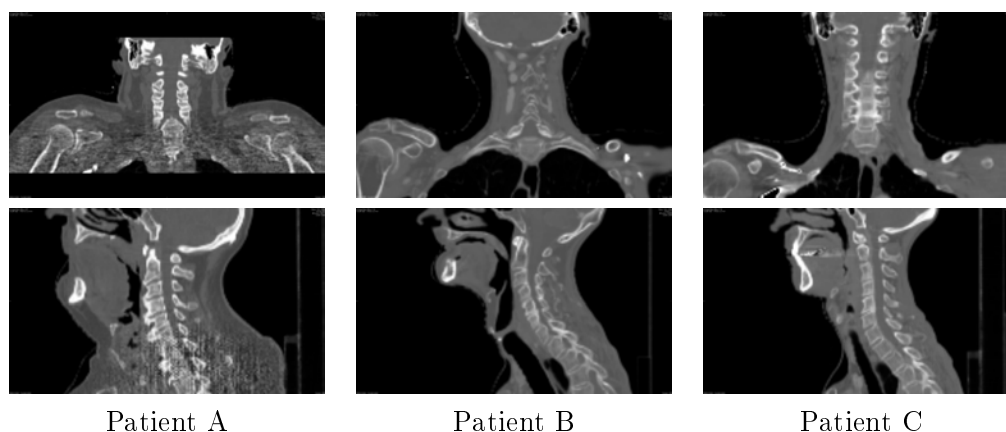


FIG. 6.27 – Trois patients pour lesquels l'utilisation de l'atlas moyen composite donne des résultats significativement meilleurs par rapport à toutes les autres approches. Pour chaque patient, la première ligne montre une coupe coronale, la seconde une coupe sagittale.

6.5 Synthèse des contributions

Dans ce chapitre, nous avons présenté une évaluation approfondie des différents algorithmes multi-atlas proposés dans le chapitre 5.

Sous-atlas par localisation et stade de tumeur Tout d'abord, nous avons montré en section 6.1 que l'utilisation des sous-atlas par localisation et stade de tumeur ne permettait pas de dépasser les performances de l'atlas moyen global. Cependant, le critère de type méta-information proposé (c'est-à-dire la localisation et le stade de la tumeur primitive) s'est révélé plus pertinent qu'une sélection aléatoire pour la catégorie oropharynx N2, ce qui laisse à penser que les sous-atlas construits

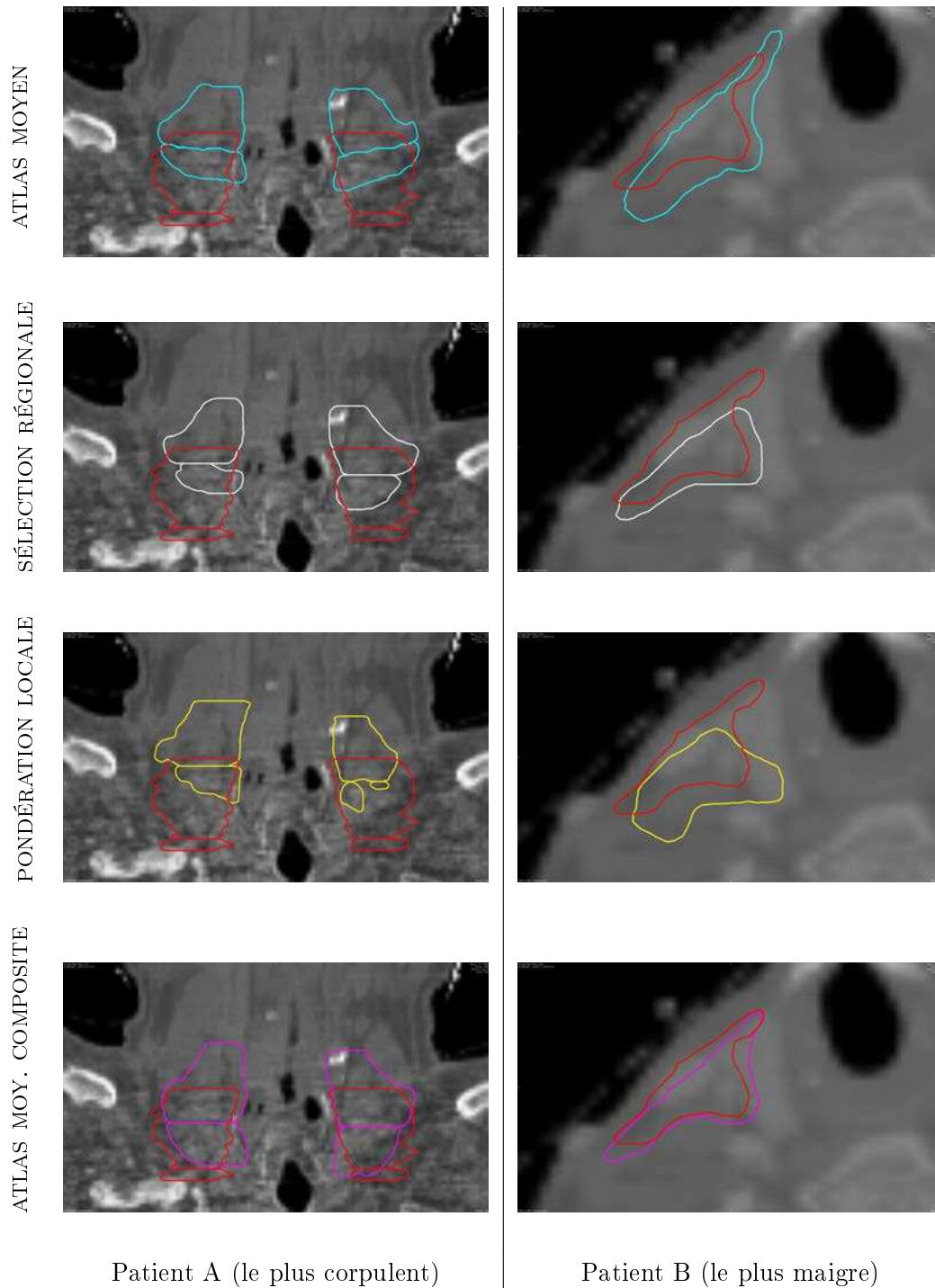


FIG. 6.28 – Résultats visuels de segmentations pour le patient le plus corpulent (première colonne) et pour le patient le plus maigre de notre base (deuxième colonne) sur les aires ganglionnaires III et IV : les contours automatiques obtenus avec l’atlas moyen (en bleu, première ligne), la sélection régionale (en blanc, deuxième ligne), la pondération locale (en jaune, troisième ligne) et l’atlas moyen composite construit à la volée (en rose, quatrième ligne) sont comparés aux contours manuels (en rouge).

d'après ce critère pourraient donner de meilleurs résultats que l'atlas moyen global sur une base d'images plus grande que celle dont nous disposons. Pour la catégorie oropharynx N0-N1, les différences n'étaient pas significatives par rapport à la sélection aléatoire, ce qui suggère que les tumeurs de stade moins avancé ont un impact moindre sur les structures environnantes.

Sélection des atlas les plus similaires au patient Concernant les différentes méthodes de sélection d'atlas, nous les avons dans un premier temps évaluées indépendamment de la phase de fusion et du nombre d'atlas sélectionnés (cf. section 6.2). Cette évaluation était basée d'une part sur les comparaisons de classements d'atlas à travers le protocole proposé dans le chapitre 5, et d'autre part sur les performances des atlas en terme de qualité de segmentation selon leur rang dans le classement. Les deux types d'évaluation ont conduit aux conclusions suivantes :

- La sélection régionale permet d'atteindre de meilleures performances que la sélection globale.
- Pour la sélection globale comme pour la sélection régionale, la méthode de sélection optimale est la mesure de coefficients de corrélation locaux calculés dans le référentiel du patient après propagation non-linéaire directe des atlas.
- Pour la sélection globale, les différences entre les méthodes de sélection ne sont pas très prononcées ; pour la sélection régionale au contraire, on constate clairement deux points : (1) l'utilisation des coefficients de corrélation locaux est préférable à toute autre mesure de similarité, et (2) la mise en correspondance de type non-linéaire (de préférence directe plutôt que indirecte via l'atlas moyen) est plus pertinente que la mise en correspondance de type affine.
- Les critères de sélection basés sur les déformations (dont celui proposé utilisant les jacobiens) sont moins pertinents que les critères basés intensité, en particulier pour la sélection régionale.

Dans un deuxième temps, nous avons évalué un nombre restreint de méthodes de sélection en comparant leurs performances en terme de qualité de segmentation à l'issue de la phase de fusion (cf. section 6.3.1). Pour cette analyse, la méthode de fusion utilisée est la fusion des segmentations des atlas dans le référentiel du patient, après propagation directe ou indirecte. Les résultats de cette étude ont corroboré nos précédentes conclusions, que ce soit dans le cadre d'une propagation directe ou indirecte, et d'une sélection globale ou régionale.

Fusion des segmentations dans le référentiel du patient Concernant les méthodes de fusion d'atlas, nous avons dans un premier temps comparé les différentes approches de fusion des segmentations des atlas après propagation dans le référentiel du patient (cf. section 6.3). Pour cela, nous avons étudié l'effet de l'échelle à laquelle est faite la sélection ou pondération des atlas (globale, régionale ou locale) ainsi que l'effet du type de propagation utilisé (directe ou indirecte). Nous avons étudié l'influence du type de propagation à la fois pour la phase de sélection et pour la phase de fusion.

Concernant l'échelle à laquelle est appliquée la sélection d'atlas, notre étude a montré que les performances étaient d'autant meilleures que la sélection (ou pondération) est ciblée, la sélection globale ne permettant pas de dépasser l'atlas moyen global (c'est-à-dire $DICE(\text{locale}) > DICE(\text{régionale}) > DICE(\text{globale}) \approx DICE(\text{atlas moyen global})$).

Concernant l'influence de la propagation sur la phase de fusion, nous avons tout d'abord remarqué que quand la totalité des atlas est fusionnée (pas de sélection ni de pondération), la propagation directe donne de moins bons résultats que la propagation indirecte (qui donne quant à elle des résultats similaires à l'atlas moyen global). Cela est à corrélérer avec la très grande variabilité anatomique dans notre base, et signifie que les patients de la base sont en moyenne plus proches de l'atlas moyen que des autres patients. Cependant, lorsqu'un nombre restreint d'atlas est sélectionné puis fusionné, l'utilisation d'une propagation non-linéaire directe pour les deux phases (sélection et fusion) est la combinaison la plus efficace dès lors que la méthode de sélection d'atlas est pertinente.

Pour la sélection régionale, nous avons étudié les performances atteintes par chaque approche en fonction du nombre d'atlas fusionnés. Avec la méthode de sélection optimale, les performances maximales sont atteintes en fusionnant environ 15 segmentations d'atlas par structure (en moyenne sur toutes les structures).

Pour la pondération locale des segmentations des atlas :

- Les différences entre propagation non-linéaire directe et indirecte sont significatives, mais elles restent faibles.
- La propagation directe nécessite autant de recalages non-linéaires que d'atlas dans la base, tandis que la propagation indirecte en nécessite un seul.

L'utilisation de la pondération locale après propagation indirecte semble donc le meilleur compromis entre qualité de segmentation et temps de calcul. Une inspection visuelle des contours montre que les différences entre les deux jeux de contours sont marginales, ce qui justifie notre choix. La figure 6.29 résume les performances atteintes avec cette méthode comparativement à l'atlas moyen global.

Enfin, des résultats visuels ont permis de comparer les performances respectives des approches régionale et locale et de l'atlas moyen global, et de confirmer les améliorations mises en évidence de façon quantitative.

Atlas moyens composites adaptés aux patients En section 6.4, nous avons montré que l'utilisation de l'atlas moyen composite permet d'améliorer significativement la qualité de la segmentation par rapport à l'atlas moyen global. Nous avons montré par des exemples visuels que les atlas moyens composites ainsi construits sont davantage similaires aux patients que l'atlas moyen global.

Comparativement aux approches de fusion des segmentations dans le référentiel du patient, l'atlas moyen composite donne des résultats similaires à la sélection régionale et moins bons que la pondération locale. Cependant, nous avons mis en évidence que l'algorithme proposé est le plus efficace de tous pour les patients ayant une anatomie très atypique (en particulier les patients corpulents ou maigres).

	Atlas moyen global	Pondération locale après propagation indirecte
AG. II	0.803	0.819
AG. III	0.791	0.804
AG. IV	0.685	0.692
Parotides	0.766	0.795
Sous-maxillaires	0.692	0.731
Moelle	0.801	0.819
Tronc cérébral	0.791	0.804
Mandibule	0.829	0.848
Moyenne	0.761	0.779

FIG. 6.29 – Performances atteintes en terme d’indice de DICE moyen sur les 105 patients avec l’atlas moyen global et avec l’algorithme multi-atlas représentant le meilleur compromis entre temps de calcul et qualité de segmentation (c’est-à-dire la pondération locale des segmentations des atlas après propagation indirecte dans le référentiel du patient).

6.6 Conclusion : limites et perspectives

L’utilisation des algorithmes multi-atlas permet d’améliorer significativement les performances obtenues avec l’atlas moyen. Leur principale limite est liée à la variabilité de contourage inter/intra expert des structures ORL (décrite dans l’annexe C), cette variabilité étant entre autre liée à la prise en compte d’informations sur le diagnostic clinique lors du contourage. Plusieurs solutions peuvent être envisagées pour s’affranchir de cette limite.

Tout d’abord, une inspection approfondie de l’ensemble de la base peut permettre de repérer et d’exclure les contours atypiques. Par exemple, Teguh et al. excluent de la base d’atlas ainsi que de l’évaluation les contours d’aires ganglionnaires pour lesquels les muscles sont envahis [Teguh 2010]. Comme nous avons mis en évidence que le stade de la tumeur avait une influence sur les contours, une autre piste consisterait à dissocier dans la base les patients dont la tumeur est au stade N0 des patients dont la tumeur est à un stade supérieur (N+), et à appliquer nos algorithmes multi-atlas en utilisant pour chaque patient la sous-base adéquate uniquement. Par ailleurs, la symétrisation de la base permettrait de doubler le nombre d’atlas disponibles, et donc d’améliorer potentiellement nos résultats.

Enfin, des algorithmes de segmentation non-basés sur les atlas pourraient être utilisés à la suite des algorithmes multi-atlas pour affiner la segmentation automatique. Par exemple, Han et al. ont mis en évidence une amélioration significative pour les parotides en raffinant la segmentation à l’aide d’un modèle déformable [Han 2010]. De même, des approches de classification peuvent être utilisées en combinaison des stratégies multi-atlas [Sdika 2010, Qazi 2010, Akhondi-Asl 2010]. Nous détaillerons davantage ces perspectives dans la conclusion de la thèse.

Segmentation automatique des dents pour l'évaluation des risques liés à une chirurgie dentaire post-irradiation

Sommaire

7.1	Introduction	182
7.1.1	Contexte clinique des soins dentaires post-irradiation	182
7.1.2	Comment estimer la dose reçue par chaque dent ?	182
7.1.3	Organisation du chapitre	184
7.1.4	Dénomination des dents	184
7.2	État de l'art sur la segmentation des dents	184
7.2.1	Difficultés principales	184
7.2.2	État de l'art	185
7.3	Description de l'algorithme	186
7.3.1	Base de données	186
7.3.2	Algorithme proposé	188
7.4	Résultats	190
7.4.1	Protocole d'évaluation	190
7.4.2	Évaluation en terme de qualité de segmentation	191
7.4.3	Évaluation en terme d'estimation de dose	195
7.4.4	Évaluation dent par dent	200
7.5	Discussion	202
7.5.1	Synthèse des résultats	202
7.5.2	Comparaison avec la dose sur la mandibule ou le maxillaire	203
7.5.3	Applications cliniques	204
7.5.4	Perspectives	204
7.6	Conclusion et synthèse des contributions	205

Synthèse des contributions Les principales contributions présentées dans ce chapitre sont les suivantes :

- Proposition d'une méthode de segmentation automatique des dents,
- Analyse de la pertinence de l'utilisation des contours automatiques obtenus pour l'estimation de dose dans le contexte de l'évaluation des risques liés à une chirurgie dentaire post-irradiation,
- Comparaison de la méthode proposée à l'utilisation d'un atlas moyen.

Publications associées Les travaux présentés dans ce chapitre ont donné lieu à un article de journal en co-première auteur [Thariat 2011]. Ils ont également fait l'objet de publications en tant que deuxième auteur dans des conférences nationale et internationale de radiothérapie [Thariat 2010d, Thariat 2010c].

7.1 Introduction

7.1.1 Contexte clinique des soins dentaires post-irradiation

Lors du traitement radiothérapeutique d'un cancer ORL, l'irradiation des glandes salivaires et des muqueuses buccales entraîne un assèchement de la bouche (xérostomie) qui augmente le risque de complications dentaires telles que les caries, et donc la nécessité de soins dentaires à la suite du traitement radiothérapeutique. Ces soins dentaires post-irradiations présentent des risques accrus de complications parce que la mâchoire et les racines des dents ont également été fragilisées par l'irradiation. Les complications classiques sont l'ostéoradionécrose (nécrose osseuse) à la suite d'une extraction dentaire [Epstein 1987, Epstein 1997], ainsi que l'échec de la pose d'un implant [Ozen 2005, Thariat 2010a, Thariat 2010b]. Avant toute chirurgie dentaire de ce type sur un patient ayant subi une radiothérapie ORL, le dentiste doit donc contacter le radiothérapeute pour que celui-ci estime le niveau de risque associé : risque de nécrose pour une extraction ou risque d'échec d'implant. Ce niveau de risque est lié à la dose reçue par la dent concernée (et en particulier par sa racine et l'os sous-jacent) et cette dose doit être estimée à partir de la dosimétrie du traitement.

7.1.2 Comment estimer la dose reçue par chaque dent ?

7.1.2.1 Estimation exacte, à partir de contours manuels

L'estimation la plus exacte possible de la dose reçue par une dent donnée nécessite de contourner manuellement cette dent sur l'image scanner ayant servi à la planification. Les TPS (Treatment Planning System) standards permettent alors, à partir des contours et de la dosimétrie, de calculer automatiquement l'histogramme dose-volume (HDV) correspondant ainsi que des caractéristiques comme les doses minimum, maximum et moyenne. La dose reçue par la dent est alors considérée comme représentative de la dose reçue par sa racine et l'os sous-jacent. Cependant, cette approche n'est pas celle utilisée en routine clinique, et ce pour plusieurs

raisons. D'une part, elle nécessite de récupérer le dossier électronique de la planification de traitement, ce qui peut prendre parfois du temps dans le cas où le dossier a été archivé (les complications dentaires peuvent survenir plusieurs années après l'irradiation). D'autre part, une fois le dossier électronique récupéré, le contournage manuel coupe par coupe est une tâche longue et fastidieuse (environ 2 heures par patient).

7.1.2.2 Estimation en routine clinique

En routine clinique, le radiothérapeute estime généralement la dose visuellement à partir de planches papier de la dosimétrie du traitement. Cependant, cette méthode est très imprécise car les planches papier ne sont pas imprimées avec le fenêtrage en intensité adéquat pour la visualisation des dents, et de plus les contours des dents ne sont pas présents (dans les cas les plus favorables, les contours des mâchoires -mandibule et maxillaire- sont présents). Quand le désarchivage du dossier électronique est possible, l'estimation de la dose se fait en visualisant la dosimétrie et l'image scanner sur le TPS, mais dans tous les cas sans contours. D'autre part, bien que le niveau de risque associé à une extraction dentaire ou à une pose d'implant dépende essentiellement de la dose reçue par la racine de la dent (et l'os sous-jacent), la dose estimée visuellement est généralement la dose maximum reçue sur l'ensemble de la dent, de façon à ne pas sous-estimer le risque de complications.

Par ailleurs, l'évolution des techniques de radiothérapie a permis de mieux conformer la distribution de dose à la tumeur, mais cela a pour conséquence des distributions de dose complexes et des gradients de dose abrupts (cf. figure 7.1) qui rendent plus difficile et plus imprécise l'estimation visuelle de la dose telle qu'elle est pratiquée en routine clinique (c'est-à-dire sans les contours et souvent à partir des versions papier seulement). L'estimation visuelle de la dose est donc imprécise et de plus peu reproductible car sujette à un fort biais lié à l'observateur.

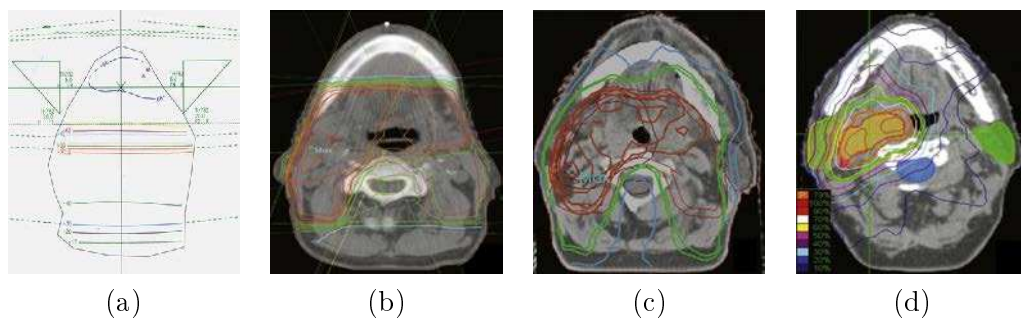


FIG. 7.1 – Mise en évidence des distributions de doses obtenues avec les techniques de radiothérapie : (a) radiothérapie conventionnelle, (b) radiothérapie conformationnelle 3D, (c) radiothérapie conformationnelle avec modulation d'intensité (RCMI), et (d) radiothérapie stéréotaxique CyberKnife (source de (b), (c), (d) : [Thariat 2010a]).

7.1.2.3 Estimation à partir de contours automatiques

L'estimation de dose telle qu'elle est faite en routine clinique (cf. section 7.1.2.2) est imprécise et non reproductible. L'estimation exacte de la dose reçue par chaque dent (cf. section 7.1.2.1) est précise et reproductible mais impossible à mettre en œuvre en routine clinique à cause du temps de contourage manuel des dents. Dans ce chapitre, nous proposons de segmenter automatiquement les dents sur l'image scanner, et d'utiliser ces contours automatiques pour déterminer automatiquement et de façon reproductible la dose reçue par chaque dent (dose maximum et histogrammes dose-volume). Un autre avantage de cette approche est que les doses reçues par chaque dent peuvent être calculées d'emblée au moment de la planification du traitement, et stockées dans le dossier papier du patient de façon à être directement accessibles par le dentiste sans que celui-ci ait besoin de contacter le radiothérapeute.

7.1.3 Organisation du chapitre

Nous présentons l'algorithme proposé pour segmenter automatiquement les dents en section 7.3, puis son évaluation sur des patients traités en RCMI en section 7.4. L'évaluation compare les performances de notre méthode en terme de qualité de contours par rapport aux contours manuels, et également en terme de précision de la dose estimée par rapport à la dose estimée avec les contours manuels (c'est-à-dire la méthode exacte décrite dans la section 7.1.2.1). Nous comparons aussi la précision de notre méthode en terme de dose à celle de la méthode utilisée en routine clinique (c'est-à-dire l'estimation visuelle à partir de l'image scanner et de la dosimétrie mais sans contours, décrite dans la section 7.1.2.2).

7.1.4 Dénomination des dents

La figure 7.2 montre la numérotation standard des dents. Les dents 11 à 18 désignent le quart supérieur droit, les dents 21 à 28 le quart supérieur gauche, les dents 31 à 38 le quart inférieur gauche et les dents 41 à 48 le quart inférieur droit. Les dents X1 et X2 correspondent aux incisives, les dents X3 aux canines, les dents X4 et X5 aux pré-molaires, les dents X6 et X7 aux molaires, et les dents X8 aux dents de sagesse.

7.2 État de l'art sur la segmentation des dents

7.2.1 Difficultés principales

Bien que les dents soient des structures contrastées et bien visibles sur les images scanner, leur segmentation automatique présente pourtant plusieurs difficultés :

- La présence d'artefacts dus aux plombages dentaires introduit des biais d'intensité importants sur les images.
- La variabilité anatomique des dents est importante, notamment au niveau des racines : certaines dents peuvent avoir un nombre variable de racines (un

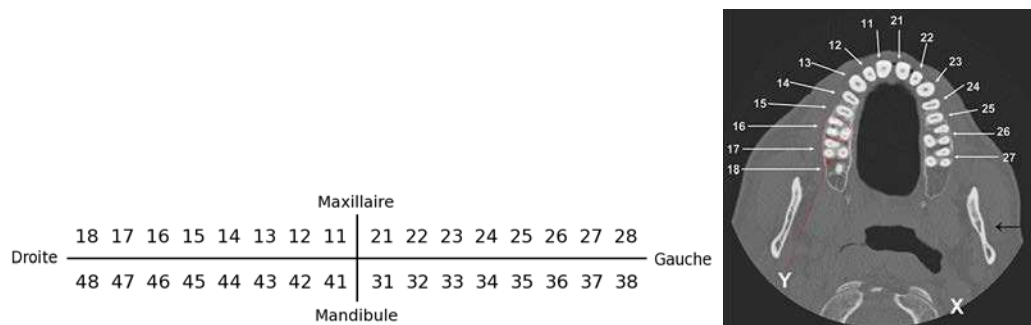


FIG. 7.2 – Numérotation des dents : schéma et denta-scanographie. A droite, la flèche Y montre les dents polyradiculées, c'est-à-dire avec plusieurs racines (source : [Thariat 2010b]).

exemple de dents polyradiculées est montré par la flèche Y sur le schéma 7.2), et la forme des racines peut être variable.

- Au niveau des racines, les dents sont peu contrastées par rapport aux structures osseuses environnantes, c'est-à-dire les mâchoires (mandibule et maxillaire).
- Au niveau des couronnes, les dents se touchent et la délimitation entre deux dents adjacentes peut poser problème. Ainsi il est relativement simple de segmenter correctement l'ensemble des dents d'un seul bloc, mais plus difficile d'obtenir les segmentations séparées de chaque dent de façon précise.
- Si les deux mâchoires sont jointives, la délimitation entre dents supérieures et dents inférieures peut être compliquée.

7.2.2 État de l'art

Plusieurs articles dans la littérature ont traité le problème de la segmentation automatique 3D des dents sur des images CT (ou CBCT). La plupart des auteurs [Hosntalab 2010, Keyhaninejad 2006, Gao 2008] abordent le problème de façon séquentielle, c'est-à-dire en suivant une trame du type suivant :

1. Extraction de l'ensemble des structures osseuses (mâchoires et dents confondues), basée sur un seuillage [Gao 2008, Hosntalab 2010] ou sur des levels sets [Keyhaninejad 2006].
2. Séparation des mâchoires supérieure et inférieure, souvent en recherchant la coupe axiale pour laquelle la somme des intensités est minimale.
3. Extraction du ou des arc(s) dentaire(s), souvent par des opérations de morphologie mathématique à partir des masques des mâchoires.
4. Le long de l'arc dentaire, détection de la frontière séparant deux dents pour chaque couple de dents adjacentes. [Gao 2008] recherche en chaque point de l'arc dentaire le plan qui minimise la somme des intensités, en partant

du plan orthogonal à l'arc dentaire, puis en autorisant des inclinaisons limitées en 3D autour de ce plan orthogonal. Un profil est ensuite tracé à partir de ces intégrales d'intensité le long de l'arc dentaire, et les minima locaux donnent les points de frontière (et donc les plans correspondants). [Keyhaninejad 2006, Hosntalab 2010] recherchent également le plan minimisant la somme des intensités, mais après projection panoramique de l'image le long de l'arc dentaire.

5. Segmentation de chaque dent à partir des plans qui la délimitent. [Keyhaninejad 2006, Hosntalab 2010] utilisent les levels sets.

[Gao 2010] utilisent des levels sets en 2D, avec des paramètres et des schémas différents suivant les coupes : pour les coupes correspondant aux racines, les dents sont segmentées successivement de façon indépendante les unes des autres, tandis que pour les coupes correspondant aux couronnes, une méthode de level sets couplés est appliquée pour gérer les jonctions entre les dents. Enfin, [Le 2009] ont présenté un algorithme pour déterminer la limite entre mâchoires inférieure et supérieure lorsque celle-ci n'est pas un simple plan (par exemple quand deux dents s'emboîtent l'une dans l'autre ou bien quand les mâchoires se chevauchent, une coupe peut alors contenir les deux dents à la fois).

Du point de vue applicatif, la segmentation automatique des dents pour l'estimation de la dose reçue lors d'un traitement radiothérapeutique n'a pas été abordée à notre connaissance.

7.3 Description de l'algorithme

Dans cette partie, nous décrivons la base de données utilisée en section 7.3.1, puis les détails de l'algorithme en section 7.3.2.

7.3.1 Base de données

Nous avons utilisé rétrospectivement des images scanner de patients ayant subi une radiothérapie de la sphère ORL. Ces images proviennent de différents centres : le centre Antoine Lacassagne à Nice, le centre Georges François Leclerc à Dijon, et l'Université Catholique de Louvain à Bruxelles. Parmi la totalité des images, nous avons exclu celles qui ne se prêtaient pas à l'étude, c'est-à-dire :

- celles pour lesquelles la résolution axiale était insuffisante (la plus mauvaise résolution axiale gardée est de 3.8 mm, mais la plupart sont de 2.5 mm),
- celles pour lesquelles plus de sept dents étaient manquantes car arrachées,
- celles comportant de trop importants artefacts dentaires et pour lesquelles on ne distinguait plus du tout les dents.

A l'issue de ce tri, nous avons ainsi sélectionné 21 images. Parmi ces 21 images, seules 7 n'ont aucun artefact dentaire. La figure 7.3 montre des exemples parmi les 21 images de la base.

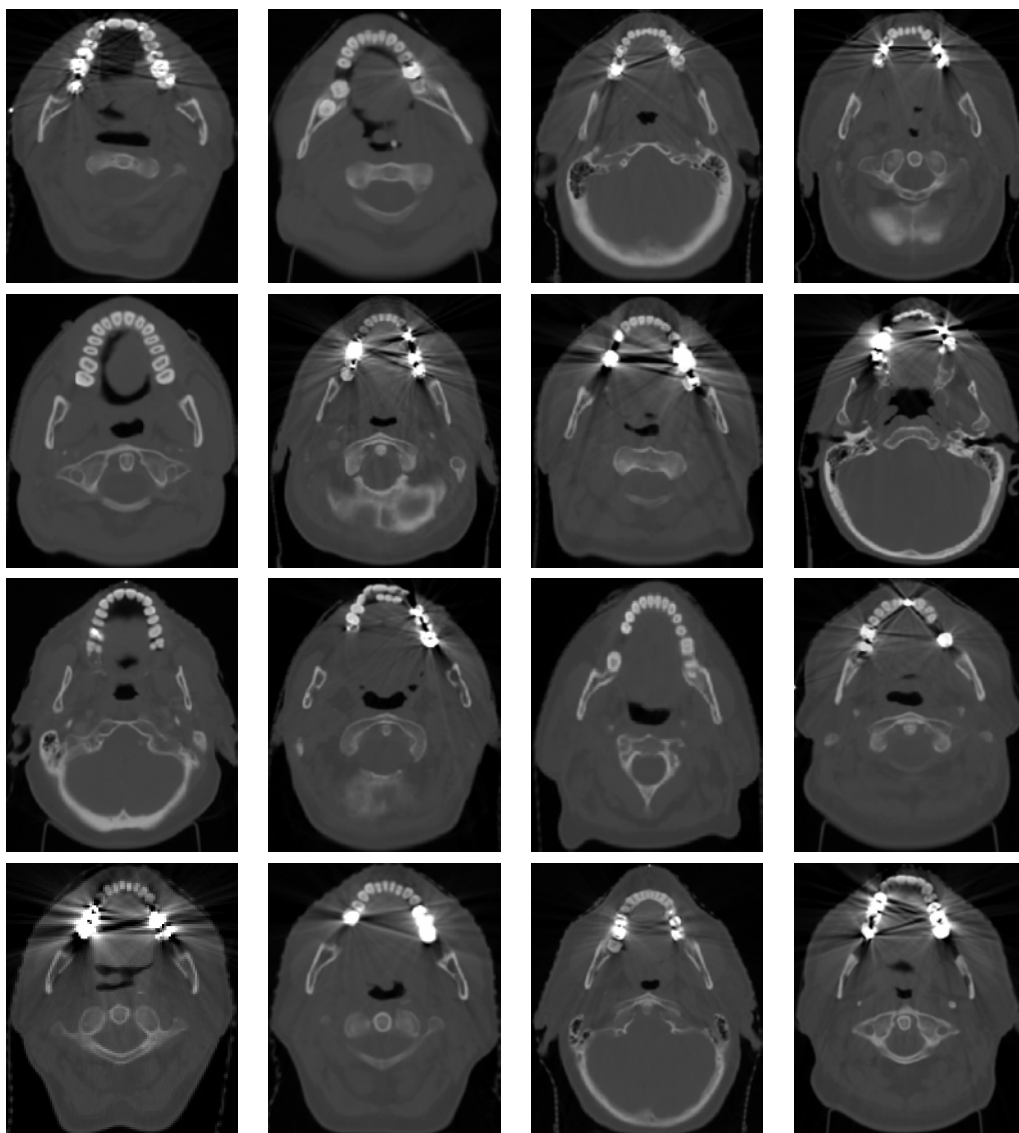


FIG. 7.3 – Exemples d'images utilisées.

Pour chaque image de cette base, les dents présentes ont été contourées manuellement par un radiothérapeute (Juliette Thariat du Centre Antoine Lacassagne) en utilisant un fenêtrage adapté à leur visualisation (800-2500 unités Hounsfield). Pour certaines des 21 images, la dosimétrie associée n'était pas disponible mais nous les avons gardé pour pouvoir exploiter leur segmentation manuelle lors de l'étape de segmentation automatique des dents. Afin de doubler la base de données, nous avons symétrisé chaque image et sa segmentation manuelle par rapport à son axe médian sagittal, conduisant ainsi à une base de $N = 42$ images manuellement contourées.

7.3.2 Algorithme proposé

Pour segmenter automatiquement les dents sur un patient test donné, nous proposons d'utiliser l'approche de segmentation par atlas. Comme détaillé dans les chapitres 3 et 5, différentes stratégies sont possibles pour tirer profit de notre base de segmentations manuelles. Tout d'abord, la construction et l'utilisation d'un atlas moyen (image moyenne et segmentation moyenne) a l'avantage d'être simple à mettre en œuvre mais elle paraît peu adaptée pour le cas particulier des dents. En effet, la plupart des images de notre base comportent des artefacts dentaires (plus ou moins prononcés selon les cas), et l'image d'intensité moyenne construite par exemple avec [Guimond 2000] est en conséquence relativement floue au niveau des dents (cf. figure 7.4). Sur cette image moyenne, les frontières entre les différentes dents ne sont absolument pas visibles. Cela risque de réduire la précision des recalages avec l'image moyenne introduisant ainsi des imprécisions à deux niveaux : d'une part lors de la mise en correspondance des segmentations manuelles de la base dans le référentiel de l'image moyenne pour le calcul de segmentation moyenne, et d'autre part lors du recalage entre le patient test et l'image moyenne. Nous proposons donc d'utiliser une approche de type multi-atlas basée sur des recalages non-linéaires directs entre chaque image de la base et le patient test. Cela permet en effet de mieux guider le recalage en exploitant au maximum les séparations visibles entre chaque dent ainsi que les contrastes des dents par rapport aux tissus environnants (qui sont plus prononcés sur les images de la base que sur l'image moyenne). Par ailleurs, les biais d'intensité causés par les artefacts risquent de perturber fortement une quelconque sélection ou pondération des atlas les plus similaires au patient. Pour cette raison, nous avons choisi d'exploiter la totalité des atlas de la base sans appliquer de sélection ni de pondération. Nous décrivons plus en détail notre approche dans la suite.

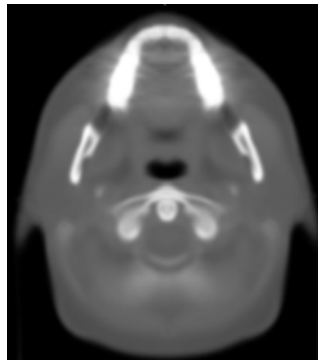


FIG. 7.4 – Image d'intensité moyenne floue sur la zone des dents.

Les principales étapes de l'approche proposée sont détaillées sur le schéma 7.5. Étant donné un patient test externe à la base, la première étape consiste à recalculer de façon affine puis non-linéaire chacune des 42 images de la base sur ce patient test. Les algorithmes de recalage utilisés sont ceux décrits dans l'annexe A. Cela permet

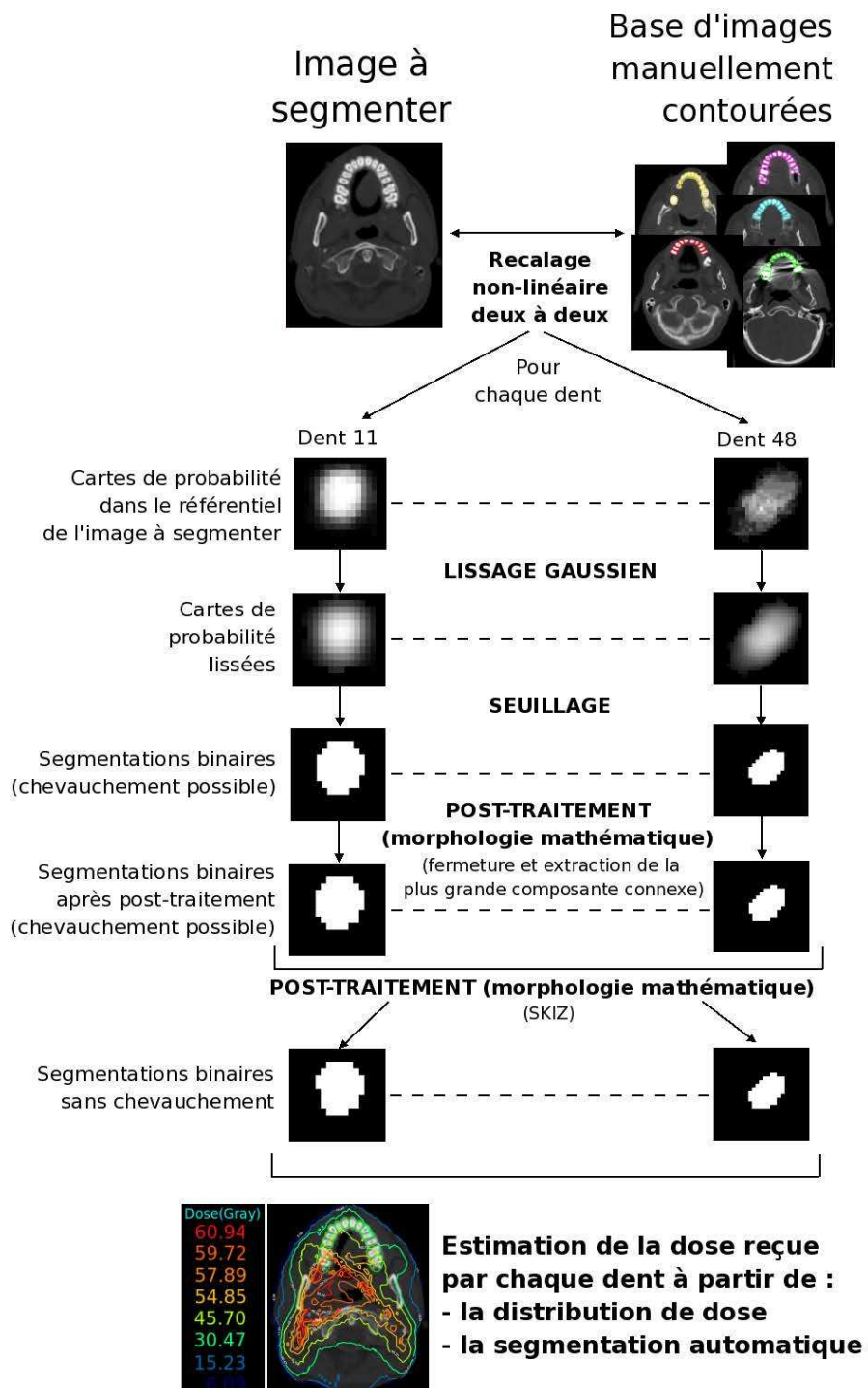


FIG. 7.5 – Résumé de la méthode de segmentation automatique des dents.

de déformer l'ensemble des segmentations manuelles de la base sur notre patient test. Chacune des segmentations manuelles ainsi déformées représente alors une segmentation potentielle des dents sur l'image test, et leur fusion permet une estimation robuste de la segmentation de l'image test. Pour cette fusion de segmentations, les différentes méthodes présentées dans le chapitre 4 peuvent *a priori* être utilisées. Nous choisissons d'utiliser l'algorithme de vote par majorité lissé présenté en section 4.5.2.2 car il est rapide, précis et robuste à un faible nombre de segmentations (ce qui est notre cas pour les dents X8). Dans un premier temps, une segmentation moyenne binaire de chaque structure est estimée indépendamment par les étapes suivantes : calcul d'une carte de probabilité à partir des segmentations manuelles de cette structure, puis lissage gaussien et seuillage, et enfin fermeture et extraction de la plus grande composante connexe pour éliminer les voxels potentiellement isolés. Les segmentations binaires des différentes dents ainsi calculées peuvent cependant se chevaucher. Pour corriger ce chevauchement, nous appliquons pour chaque couple de structures voisines un SKIZ (squelette par zones d'influence) des deux segmentations non-chevauchantes à l'intérieur de l'union des deux segmentations chevauchantes (cf. section 4.5.2.2 pour plus de détails).

Enfin, pour chaque dent, l'histogramme dose-volume (HDV) ainsi que la dose maximum peuvent alors être estimés en utilisant conjointement la segmentation automatique de la dent et la dosimétrie dans un TPS standard. Nous avons utilisé pour cela le logiciel ISOgray de DOSIsoft qui permet une estimation précise de ces caractéristiques de dose grâce à un tirage aléatoire d'un très grand nombre de points à l'intérieur du volume considéré.

7.4 Résultats

Cette partie décrit dans un premier temps le protocole d'évaluation utilisé en section 7.4.1, puis présente ensuite les évaluations en terme de qualité de segmentation en section 7.4.2 et en terme d'estimation de dose en section 7.4.3.

7.4.1 Protocole d'évaluation

Pour évaluer notre algorithme, nous avons appliqué un protocole d'évaluation de type Leave-One-Out sur notre base d'images. Cependant, certaines des images de la base n'ont pas la dosimétrie associée et l'algorithme ne peut pas être évalué en terme de dose sur ces images. Parmi les images restantes, certaines correspondent à une radiothérapie conformationnelle 3D standard, et d'autres à une radiothérapie conformationnelle avec modulation d'intensité (RCMI). Afin d'évaluer notre algorithme dans le cas le plus difficile, c'est-à-dire avec les distributions de dose les plus complexes, nous avons choisi de limiter l'évaluation aux seuls 8 patients de la base pour lesquels la dosimétrie associée est de type RCMI. Pour chacun de ces 8 patients, nous avons exclu de la base des 42 images le patient considéré ainsi que son symétrique, et nous avons appliqué l'algorithme à partir des 40 images restantes et de leurs segmentations manuelles. Les paramètres de l'algorithme ont été opti-

misés pour maximiser les mesures de similarité entre les contours automatiques et les contours manuels pour ces 8 patients donnés (selon les mesures DICE, VOL et $RHD_{\tau=95\%}$ présentées dans l'annexe B). Les valeurs retenues pour le lissage puis le seuillage de la carte de probabilité lors de la fusion de segmentation sont les suivantes : un écart-type de 1 pour le filtrage gaussien, et un seuil de 0.25.

Afin d'analyser l'apport de l'algorithme proposé par rapport à l'utilisation d'un atlas moyen classique, nous avons également mis en œuvre et évalué cette seconde méthode. L'image moyenne a été construite avec l'algorithme de Guimond et al. [Guimond 2000], et la segmentation moyenne associée avec l'algorithme de vote par majorité lissé présenté en section 4.5.2.2. Un protocole de Leave-One-Out a également été utilisé : pour chacun des 8 patients, l'atlas moyen est construit à partir des 40 images restantes (après suppression du patient en question et de son symétrique). Comme pour l'algorithme multi-atlas, les paramètres de lissage et de seuillage ont été optimisés pour les 8 patients, et les valeurs retenues sont les suivantes : un écart-type de 1 pour le filtrage gaussien, et un seuil de 0.3.

Dans la suite, nous notons MULTIAT la méthode multi-atlas proposée et AVAT l'utilisation de l'atlas moyen.

7.4.2 Évaluation en terme de qualité de segmentation

L'objectif premier de ce travail n'est pas le contourage des dents à proprement parler, mais l'estimation de la dose reçue par chaque dent. Cependant, plus les contours automatiques sont proches des contours manuels, meilleure sera l'estimation de la dose. Nous présentons donc dans cette section une analyse qualitative puis quantitative de la qualité des contours automatiques obtenus avec les méthodes MULTIAT et AVAT en les comparant aux contours manuels.

7.4.2.1 Évaluation qualitative

Les figures 7.6 et 7.7 présentent des résultats visuels de segmentation automatique obtenue selon les deux méthodes MULTIAT (colonne de gauche, contours verts) et AVAT (colonne de droite, contours jaunes). Les contours manuels sont représentés en bleu sur les deux colonnes. Les coupes montrées se situent au niveau du maxillaire (os de la mâchoire supérieure) pour la figure 7.6, et au niveau de la mandibule (os de la mâchoire inférieure) pour la figure 7.7. Nous avons choisi de présenter des coupes se situant au niveau des racines des dents et non de leurs extrémités car c'est à ce niveau que la dose reçue est la plus préjudiciable et donc qu'il est important de segmenter correctement les dents avec les méthodes automatiques. Visuellement, les résultats sont un peu moins bons au niveau de la jonction entre les deux mâchoires lorsque celles-ci sont jointives (non présenté ici). Comme mentionné dans l'état de l'art, la délimitation précise entre les deux mâchoires est un problème réputé difficile dans la segmentation des dents.

Tout d'abord, on remarque que les deux méthodes fournissent des segmentations automatiques de tailles similaires entre elles, et similaires aux tailles des segmenta-

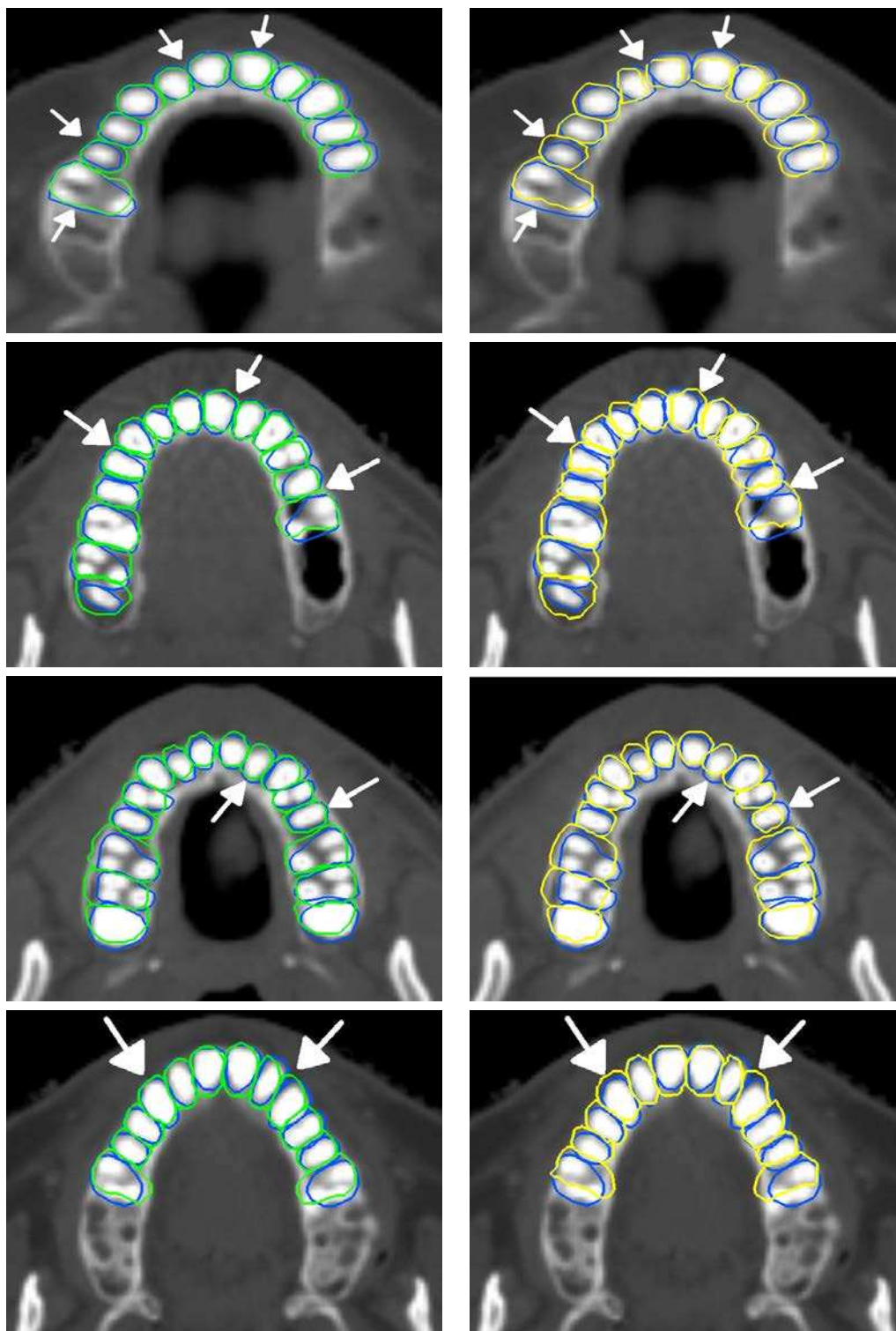


FIG. 7.6 – Résultats visuels au niveau du maxillaire. Chaque ligne correspond à une coupe donnée parmi les coupes des 8 patients tests. La colonne de gauche compare les contours automatiques obtenus avec la méthode MULTIAT (en vert) aux contours manuels (en bleu). La colonne de droite compare les contours automatiques obtenus avec la méthode AVAT (en jaune) aux contours manuels (en bleu).

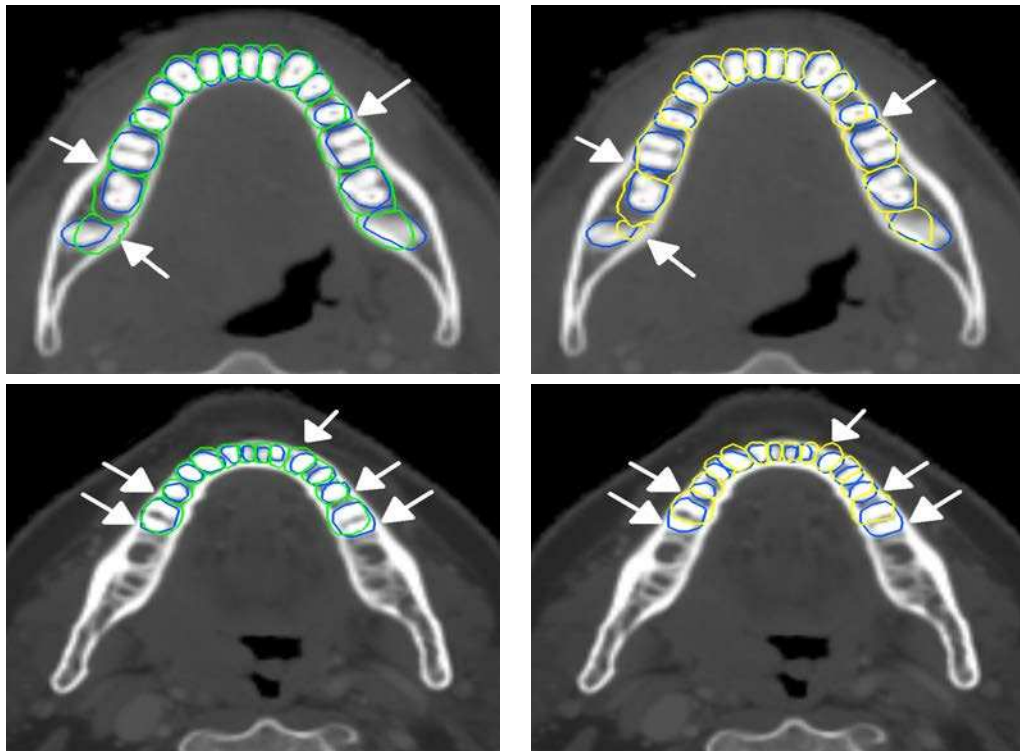


FIG. 7.7 – Résultats visuels au niveau de la mandibule. Chaque ligne correspond à une coupe donnée parmi les coupes des 8 patients tests. La colonne de gauche compare les contours automatiques obtenus avec la méthode MULTIAT (en vert) aux contours manuels (en bleu). La colonne de droite compare les contours automatiques obtenus avec la méthode AVAT (en jaune) aux contours manuels (en bleu).

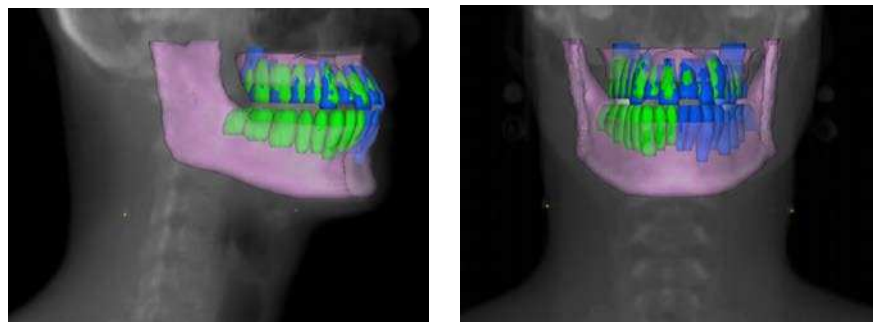


FIG. 7.8 – Résultats visuels après reconstruction 3D des segmentations automatiques obtenues avec la méthode MULTIAT (en vert) et des segmentations manuelles (en bleu) (les deux étant superposées sur la mâchoire supérieure). Les segmentations manuelles de la mandibule et du maxillaire sont représentées en rose.

tions manuelles. Les principales différences entre les contours automatiques obtenus en utilisant les deux méthodes se situent au niveau des frontières entre dents adjacentes. Les flèches blanches sur les figures mettent en évidence que ces frontières sont en général mieux localisées avec la méthode MULTIAT qu'avec la méthode AVAT (en se référant aux contours manuels bleus). Ce constat est général aux 8 patients test, et confirme que au niveau des frontières entre les dents, le recalage avec l'image moyenne est moins précis que les recalages directs avec les images de la base.

Les imprécisions observées sont probablement imputables aux nombreux artefacts dentaires dans les images de notre base (cf. figure 7.3). En effet, la qualité des segmentations automatiques est globalement moins bonne pour les dents X5, X6, X7 et X8 (dents sur lesquelles les artefacts sont concentrés) que pour les dents de devant X1, X2, X3 et X4 (qui ne présentent pas d'artefacts).

Enfin, la figure 7.8 présente des résultats visuels après reconstruction 3D des contours manuels et automatiques en utilisant la méthode MULTIAT.

7.4.2.2 Évaluation quantitative

Nous présentons maintenant une analyse quantitative de la qualité des contours automatiques obtenus avec les méthodes MULTIAT et AVAT en les comparant aux contours manuels selon les mesures présentées dans l'annexe B. La figure 7.9 présente les résultats obtenus sous forme de boîtes à moustaches pour les mesures DICE, VOL et $RHD_{\tau=95\%}$ sur l'ensemble des dents des 8 patients test (soit 214 dents au total).

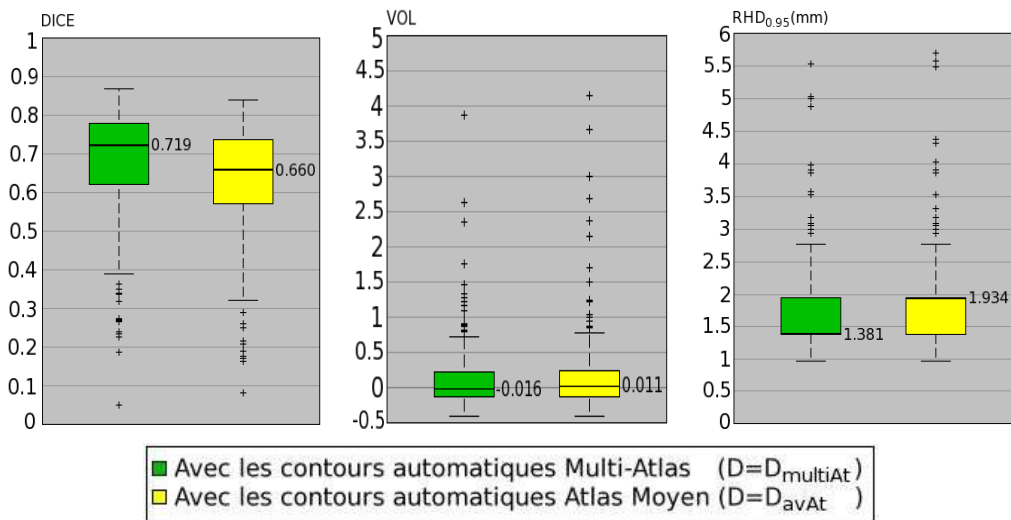


FIG. 7.9 – Boîtes à moustaches des mesures DICE, VOL et $RHD_{\tau=95\%}$ obtenues avec les contours automatiques des méthodes MULTIAT et AVAT sur l'ensemble des dents des 8 patients test (soit 214 dents au total). Les boîtes colorées représentent l'écart inter-quartile et les lignes horizontales à l'intérieur de ces boîtes représentent les valeurs médianes. Les croix noires correspondent aux données aberrantes qui sont en dehors des moustaches.

Nous avons également calculé les p-values pour chacune des trois mesures avec un test de Student pour des observations paires, avec un seuil de signification de 0.05.

En terme d'indice de Dice (graphique de gauche), la méthode MULTIAT fournit des résultats significativement meilleurs que la méthode AVAT (moyenne de 0.67 contre 0.62, avec une p-value de 3.10^{-17}).

En terme de volume (graphique du milieu), les deux méthodes MULTIAT et AVAT conduisent à des résultats très proches, ce qui confirme les résultats qualitatifs présentés sur les figures 7.6 et 7.7. On note cependant une tendance à la sur-segmentation pour les deux méthodes (les deux quartiles supérieurs sont clairement plus larges que les deux quartiles inférieurs). Cette sur-segmentation pourra potentiellement conduire à une surestimation de la dose reçue par les dents. Afin de déterminer si l'une des deux méthodes fournit des volumes plus proches des volumes manuels que l'autre (c'est-à-dire des mesures de VOL plus proches de 0, qu'importe leurs signes), nous avons également analysé les résultats de la mesure $|\text{VOL}|$ (non présentés sur la figure). Nous obtenons pour MULTIAT une moyenne de 0.287, et pour AVAT une moyenne de 0.313, avec une p-value de 0.034. Ceci indique que la méthode MULTIAT fournit des segmentations dont le volume est globalement plus proche de la réalité que ceux fournis par la méthode AVAT.

Enfin, pour la distance de Hausdorff robuste (graphique de droite), les résultats obtenus avec les deux méthodes sont assez proches, mais légèrement meilleurs pour la méthode MULTIAT (moyenne de 1.827 mm pour MULTIAT, et de 1.903 mm pour AVAT) avec une p-value de 0.014 indiquant que la différence est significative.

En conclusion, la méthode MULTIAT fournit selon les mesures DICE, VOL et $\text{RHD}_{\tau=95\%}$ des volumes automatiques significativement plus proches des volumes manuels que ceux fournis par la méthode AVAT. Bien que significatives, les différences entre les deux méthodes sont malgré tout de petite amplitude.

7.4.3 Évaluation en terme d'estimation de dose

Dans la section précédente, nous avons montré que la méthode MULTIAT fournit des contours automatiques plus proches des contours manuels que la méthode AVAT. Le premier objectif de cette section est d'évaluer dans quelle mesure ces différences se répercutent sur l'estimation de la dose à partir des contours automatiques. Le second objectif est de comparer les deux méthodes automatiques à la méthode visuelle d'estimation de la dose telle qu'elle est faite en routine clinique (détaillée en section 7.1.2.2). Dans cette partie, nous utilisons comme vérité terrain les doses (maximum ou HDV) estimées à partir des contours manuels (c'est-à-dire la méthode exacte détaillée en section 7.1.2.1).

7.4.3.1 Évaluation qualitative des histogrammes dose-volume (HDV)

Le principe des HDVs est expliqué sur la figure 2.13 : les abscisses correspondent aux doses en Gray, et pour une abscisse D donnée, l'ordonnée sur la courbe repré-

sente le pourcentage du volume qui reçoit une dose supérieure à D . La figure 7.10 montre des exemples de HDVs pour les 8 dents du quart supérieur droit de l'un des 8 patients test, obtenus avec les contours manuels (courbes bleues), avec les contours automatiques issus de la méthode MULTIAT (courbes vertes) et avec les contours automatiques issus de la méthode AVAT (courbes jaunes). Sur ces exemples, les courbes vertes (MULTIAT) sont plus proches des courbes bleues (manuel) que les courbes jaunes (AVAT) pour les dents 12, 13, 16, 17 et 18. La tendance inverse est observée pour la dent 14, tandis que les courbes vertes et jaunes sont équidistantes des courbes bleues pour les dents 11 et 15. La dose maximum reçue par une dent donnée correspond à l'intersection de la courbe correspondante avec l'axe des abscisses. La dose maximum semble approximativement bien estimée par les deux méthodes automatiques pour les dents 11 à 17, mais surestimée pour la dent 18. Dans la section suivante, nous allons évaluer de façon plus complète et plus fine l'estimation de la dose maximum reçue par chaque dent avec les deux méthodes automatiques.

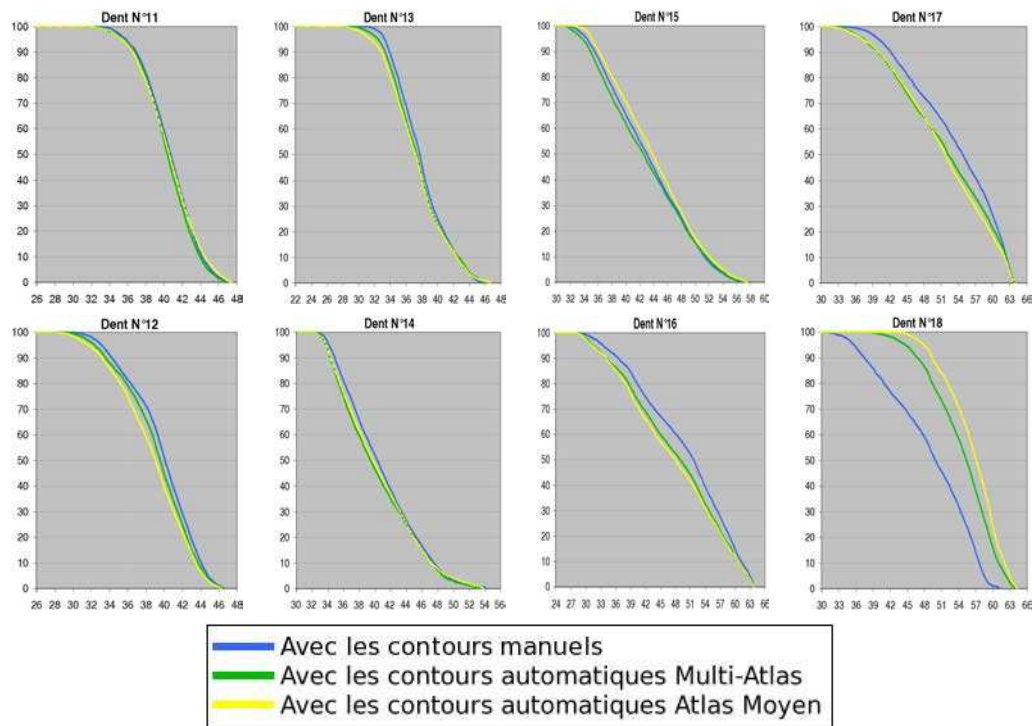


FIG. 7.10 – Exemple d’histogrammes dose-volume (HDV) obtenus sur les 8 dents du quart supérieur droit d’un patient test. Les courbes bleues, vertes et jaunes correspondent respectivement aux HDVs obtenus avec les contours manuels, avec les contours automatiques issus de la méthode MULTIAT, et avec les contours automatiques issus de la méthode AVAT.

7.4.3.2 Évaluation quantitative de la dose maximum

Dans le paragraphe précédent, nous avons évalué qualitativement des exemples de HDVs obtenus avec les contours automatiques et avec les contours manuels. Nous nous intéressons maintenant plus en détail à l'estimation de la dose maximum. Comme expliqué dans l'introduction, c'est en général la dose maximum reçue par la dent (et en particulier par sa racine) qui est estimée pour définir le niveau de risque associé à une pose d'implant ou à une extraction.

Nous présentons dans cette section une analyse quantitative de l'estimation de la dose maximum reçue par chaque dent selon différentes méthodes :

- la méthode dite *exacte* qui utilise les contours manuels (détaillée section 7.1.2.1),
- la méthode dite *visuelle* telle qu'elle est faite en routine clinique (détaillée en section 7.1.2.2) ; nous nous plaçons ici dans le cas le plus favorable où le dossier électronique du patient est disponible sur le TPS : le radiothérapeute peut alors estimer la dose maximum de façon interactive et avec un fenêtrage adapté (mais sans contours),
- la méthode de contourage automatique MULTIAT,
- la méthode de contourage automatique AVAT.

Pour un patient donné et une dent donnée, nous notons respectivement D_{gt} , D_{visu} , $D_{multiAt}$ et D_{avAt} les doses maximum estimées avec la méthode exacte, la méthode visuelle, la méthode MULTIAT et la méthode AVAT. Nous considérons D_{gt} comme la vérité terrain, et souhaitons comparer les précisions obtenues par rapport à cette vérité terrain en utilisant D_{visu} , $D_{multiAt}$ et D_{avAt} . Pour cela, nous utilisons l'erreur signée $D - D_{gt}$ ainsi que l'erreur absolue $|D - D_{gt}|$ pour $D = D_{visu}$, $D = D_{multiAt}$ et $D = D_{avAt}$. L'erreur signée nous indique dans quelle mesure la dose est surestimée ou sous-estimée, et l'erreur absolue permet de quantifier la précision de l'estimation. La figure 7.11 présente sous forme de boîtes à moustaches les statistiques obtenues pour les erreurs signées ($D - D_{gt}$, graphique de gauche) et absolues ($|D - D_{gt}|$, graphique de droite) sur les 214 dents des 8 patients test avec la méthode visuelle ($D = D_{visu}$, barres rouges), la méthode MULTIAT ($D = D_{multiAt}$, barres vertes) et la méthode AVAT ($D = D_{avAt}$, barres jaunes). Les chiffres écrits à côté des barres correspondent aux valeurs médianes. Le tableau 7.12 présente les valeurs moyennes correspondantes.

Notre objectif est double : tout d'abord nous souhaitons comparer la précision de nos méthodes automatiques MULTIAT et AVAT à la précision de la méthode visuelle qui est pratiquée en routine clinique. D'autre part, nous souhaitons aussi déterminer si l'une des deux méthodes automatiques est plus précise que l'autre.

Précision des méthodes automatiques par rapport à la méthode visuelle

Tout d'abord, on remarque que la méthode visuelle est clairement moins précise que les méthodes automatiques. En effet, elle présente en moyenne une erreur absolue de 5.48 Gy alors que les méthodes automatiques conduisent à une erreur absolue inférieure à 2 Gy en moyenne (1.87 Gy et 1.94 Gy, cf. tableau 7.12). Les différences

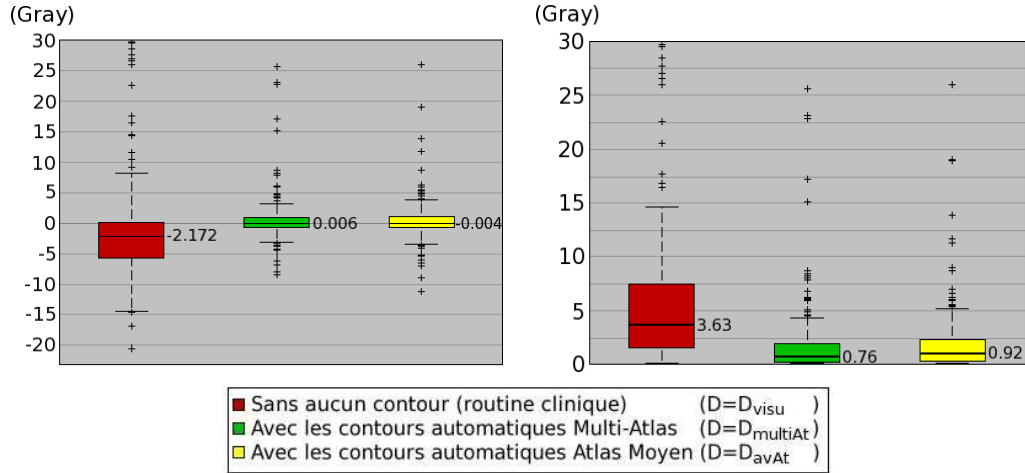


FIG. 7.11 – Boîtes à moustaches obtenues pour les erreurs signées ($D - D_{gt}$, graphique de gauche) et absolues ($|D - D_{gt}|$, graphique de droite) d'estimation de la dose maximum sur les 214 dents des 8 patients test avec différentes méthodes d'estimation de D : la méthode visuelle telle qu'elle est pratiquée en routine clinique ($D = D_{visu}$, barres rouges), la méthode automatique MULTIAT ($D = D_{multiAt}$, barres vertes) et la méthode automatique AVAT ($D = D_{avAt}$, barres jaunes). Les boîtes colorées représentent l'écart inter-quartile et les lignes horizontales à l'intérieur de ces boîtes représentent les valeurs médianes. Les croix noires correspondent aux données aberrantes qui sont en dehors des moustaches.

	$D = D_{visu}$	$D = D_{multiAt}$	$D = D_{avAt}$
$\overline{D - D_{gt}}$	-1.40 Gy	0.56 Gy	0.48 Gy
$\overline{ D - D_{gt} }$	5.48 Gy	1.87 Gy	1.94 Gy

FIG. 7.12 – Valeurs moyennes obtenues sur les 214 dents des 8 patients test pour les erreurs signées ($D - D_{gt}$) et absolues ($|D - D_{gt}|$) d'estimation de la dose maximum avec la méthode visuelle ($D = D_{visu}$), la méthode automatique MULTIAT ($D = D_{multiAt}$) et la méthode automatique AVAT ($D = D_{avAt}$).

entre les erreurs absolues commises avec la méthode visuelle d'une part et avec les deux méthodes automatiques d'autre part sont statistiquement significatives (p-values $\approx 10^{-18}$). Cela est confirmé par le graphique de droite de la figure 7.11. De plus, la méthode visuelle est celle qui fournit la plus grande sous-estimation (-20.6 Gy) et la plus grande surestimation (+29.7 Gy) (cf. données extrêmes sur la figure 7.11).

D'autre part, les erreurs d'estimation de dose commises avec la méthode visuelle correspondent dans environ 75% des cas à une sous-estimation de la dose : cela est mis en évidence par le graphique de gauche de la figure 7.11. Cette sous-estimation de dose est de 1.4 Gy en moyenne (cf. tableau 7.12). Les méthodes automatiques quant à elles ont tendance à surestimer la dose, ce qui est cohérent avec les résultats de la mesure VOL présentés en figure 7.9 (tendance à la sur-segmentation). Cependant, cette surestimation est de 0.56 Gy et 0.48 Gy en moyenne (cf. tableau 7.12), et donc de plus faible amplitude que la sous-estimation de la méthode visuelle (-1.4 Gy). De plus, une surestimation de la dose est préférable d'un point de vue clinique, car une sous-estimation correspond à un risque de complications dentaires sous-évalué.

Enfin, les écart-types et écarts inter-quartile sont significativement réduits avec les deux méthodes automatiques par rapport à la méthode visuelle (cf. figure 7.11).

Précision relative des deux méthodes automatiques Nous souhaitons maintenant comparer les deux méthodes automatiques l'une par rapport à l'autre. A première vue, les différences entre les deux méthodes sont relativement faibles. Les chiffres obtenus pour l'erreur absolue sont plutôt favorables à la méthode MULTIAT. En effet, l'erreur absolue moyenne en utilisant MULTIAT est de 1.87 Gy contre 1.94 Gy pour la méthode AVAT (cf. tableau 7.12). Cette tendance est confirmée par les boîtes à moustaches du graphique de droite de la figure 7.11 : avec la méthode MULTIAT, la médiane est inférieure (0.76 Gy contre 0.92 Gy) et l'écart inter-quartile est réduit. Concernant les chiffres obtenus pour l'erreur signée, les boîtes à moustaches du graphique de gauche de la figure 7.11 ne révèlent quasiment aucune différence. Les valeurs moyennes présentées dans le tableau 7.12 (0.56 Gy et 0.48 Gy) semblent indiquer que la surestimation de la dose est plus importante avec la méthode MULTIAT qu'avec la méthode AVAT. Cependant, les différences entre les deux méthodes ne sont pas statistiquement significatives, que ce soit pour les erreurs absolues ou signées (p-values respectives de 0.47 et 0.44).

Enfin, nous présentons sur la figure 7.13 le pourcentage de cas (parmi les 214 dents) pour lesquels l'erreur absolue d'estimation de la dose maximum est inférieure à X Gy, pour $X = 1$ à 30 Gy, et ce pour les différentes méthodes (visuelle et automatiques). On constate ici encore que la méthode visuelle est clairement moins précise que les deux méthodes automatiques. En effet, l'erreur absolue est inférieure à 2 Gy dans environ 75 % des cas avec les méthodes automatiques, contre 30 % des cas seulement pour la méthode visuelle. Pour un seuil à 5 Gy, les chiffres sont 92 % pour les méthodes automatiques contre 59 % pour la méthode visuelle. Les différences entre les deux courbes de méthodes automatiques sont faibles. La courbe

de MULTIAT (verte) est légèrement au-dessus de la courbe de AVAT (jaune). Ceci confirme nos observations précédentes : l'erreur absolue d'estimation de la dose maximum est globalement plus faible avec la méthode MULTIAT qu'avec la méthode AVAT, mais ces différences restent marginales et non significatives d'après nos tests de p-values. Cela signifie que les différences en terme de qualité de contours que nous avons mises en évidence en section 7.4.2 (qui étaient en faveur de MULTIAT et de façon significative) sont petites à l'échelle des gradients de dose au voisinage des dents pour ces 8 patients test.

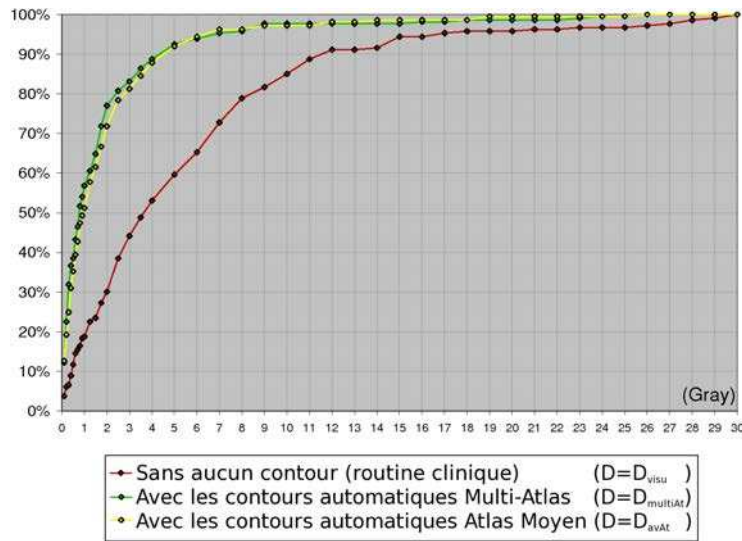


FIG. 7.13 – Pour une abscisse X donnée, les valeurs en ordonnée représentent le nombre de cas (en pourcentage de l'ensemble des 214 dents des 8 patients test) pour lesquels l'erreur absolue d'estimation de dose $|D - D_{gt}|$ est inférieure à X Gy (pour $D = D_{visu}$ en rouge, $D = D_{multiAt}$ en vert et $D = D_{avAt}$ en jaune). Par exemple, l'erreur d'estimation est inférieure à 5 Gy dans 59 % des cas avec l'estimation visuelle, et dans 92 % des cas avec les contours automatiques issus des méthodes MULTIAT et AVAT.

7.4.4 Évaluation dent par dent

Dans les résultats quantitatifs présentés jusque-là, nous avons considéré l'ensemble des dents des 8 patients test sans distinction. Dans l'idéal, il aurait été préférable de calculer les mesures d'évaluation dent par dent, mais nous disposions d'un trop petit nombre de patients test (les 8 ayant une dosimétrie IMRT) pour que ces mesures aient un sens. Cependant, nous avons malgré tout constaté des résultats globalement moins précis pour les molaires (dents X6, X7 et X8) que pour les autres dents. Cette tendance est commune aux mesures de qualité de contours (DICE, $RHD_{\tau=95\%}$) et aux mesures d'erreurs d'estimation de la dose maximum. Sur certaines des images présentées sur les figures 7.6 et 7.7, on peut en effet visualiser

des contours automatiques moins précis pour les dents X8 que pour les autres. Sur la figure 7.10, les courbes obtenues avec les contours automatiques sont plus éloignées de la courbe de référence pour la dent 18 que pour les autres dents.

Plusieurs raisons peuvent expliquer cette différence de précision suivant les dents. Tout d'abord, comme nous l'avons déjà évoqué, les molaires sont les dents présentant le plus d'artefacts dentaires dans notre base à cause des plombages. Ces artefacts limitent la qualité du recalage et donc la qualité de la segmentation automatique.

De plus, les molaires sont les dents pour lesquelles la variabilité anatomique est la plus marquée. Notamment, elles peuvent avoir un nombre variable de racines, et la forme de ces racines est également variable. Il est donc possible que le recalage non-linéaire ne permette pas de compenser l'ensemble des différences anatomiques entre deux patients donnés.

Par ailleurs, les molaires (X6, X7, X8) sont les dents qui sont le plus souvent manquantes dans notre base d'images. En particulier, les dents de sagesse (X8) sont manquantes pour environ 75 % de nos patients, ce qui correspond au taux observé en général dans la population (dents non sorties ou souvent arrachées). De plus, les molaires sont parfois arrachées avant le traitement de radiothérapie par prévention. Pour ces dents, la segmentation automatique est donc calculée à partir d'un plus faible nombre de segmentations manuelles que pour les autres dents (et de plus avec des segmentations d'une plus grande variabilité comme nous l'avons mentionné plus haut), ce qui conduit à une estimation moins robuste, et potentiellement moins précise.

D'autre part, lorsque certaines dents sont présentes sur l'image test mais absentes sur une image donnée de la base, cela perturbe le recalage non-linéaire. Le cas typique est le recalage entre un patient test avec dents de sagesse et un patient de la base sans dents de sagesse (ou l'inverse). Ceci est illustré sur la figure 7.14.

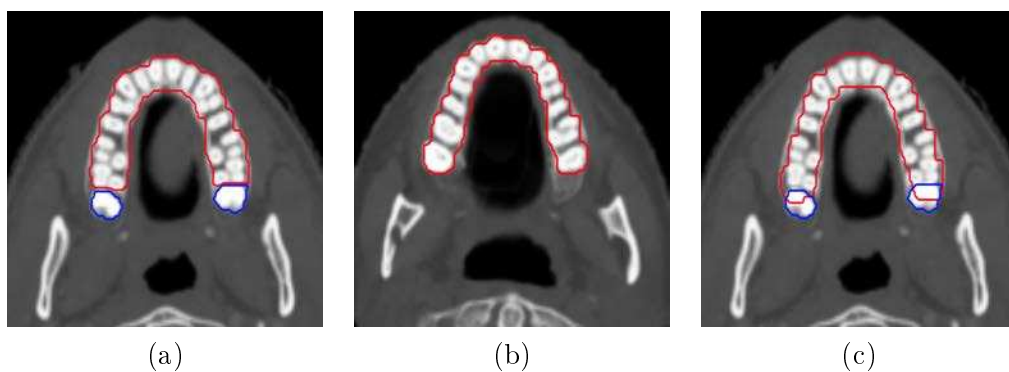


FIG. 7.14 – Illustration du décalage spatial introduit lors du recalage non-linéaire entre un patient P avec des dents de sagesse (a) et une image I de la base sans dents de sagesse (b). Les contours de l'image I déformés non-linéairement sur le patient P sont montrés en rouge sur l'image (c). Les segmentations manuelles des dents de sagesse du patient P sont ajoutées en bleu sur l'image (c) afin de bien visualiser le décalage des dents voisines (dents X6 et X7) sur les dents de sagesse.

En effet, le recalage non-linéaire produit alors un décalage spatial artificiel des dents X6 et X7 vers les dents X8. Une fois déformés sur l'image du patient test, les contours de l'image de la base sont mal localisés sur l'image test, ce qui biaise l'estimation de la segmentation pour ces dents-là. Cet effet vient se combiner à l'effet mentionné précédemment : on dispose pour les dents X6, X7 et X8 de moins de contours que pour les autres dents, et de plus ces contours sont déformés sur le patient test de façon moins précise si des dents sont manquantes dans l'image test et pas dans l'image de la base (ou l'inverse).

Une solution pour s'affranchir de ce problème de décalage spatial serait de séparer notre base d'images en plusieurs sous-bases : par exemple en séparant les patients avec et sans dents de sagesse. Cependant, cela nécessiterait d'avoir une plus grande base de données si l'on veut s'assurer un nombre suffisant de patients dans chaque sous-base. De plus, certains patients ont certaines dents de sagesse mais pas d'autres, ce qui impliquerait un trop grand nombre de sous-bases pour couvrir toutes les configurations possibles. Une alternative pourrait alors consister à sélectionner pour chaque région de chaque mâchoire les images de la base les plus proches du patient test en terme d'intensité, et à calculer pour chaque région les segmentations des dents sur l'image test à partir des seules images sélectionnées sur cette région. Des régions spécifiques pourraient par exemple être définies au niveau des dents X6, X7 et X8 pour gérer efficacement notre problème. La phase de sélection serait néanmoins très sensible à la présence d'artefacts dans les images.

7.5 Discussion

7.5.1 Synthèse des résultats

Les résultats présentés dans la section précédente montrent que notre méthode de segmentation automatique des dents permet de fournir une estimation plus précise de la dose reçue par chaque dent que la méthode d'estimation visuelle de la dose telle qu'elle est pratiquée en routine clinique (c'est-à-dire sans aucun contour). D'autre part, la méthode visuelle sous-estime la dose dans 75 % des cas, ce qui peut potentiellement conduire à une sous-estimation des complications dentaires au delà d'un certain seuil.

D'autre part, notre méthode permet d'estimer la dose maximum avec une précision inférieure à 2 Gy dans 77 % des cas, et une précision moyenne de 1.87 Gy. Du point de vue clinique, ce degré de précision est suffisant pour évaluer les risques de complications dentaires. De plus, notre méthode a plutôt tendance à sur-segmenter les dents, et donc à surestimer la dose maximum. Le graphique de gauche de la figure 7.11 montre en effet que la sous-estimation de dose la plus importante avec MULTIAT est de -8.4 Gy. Cette valeur reste tout à fait raisonnable, notamment si l'on compare à la méthode visuelle pour laquelle la sous-estimation est d'amplitude plus grande que -5.7 Gy dans 25% des cas (premier quartile). La sur-segmentation la plus importante avec MULTIAT est +25.6 Gy mais cette grande amplitude est moins gênante dans ce sens-là elle n'entraîne pas une sous-estimation du risque.

Enfin, nous avons mis en évidence que la méthode `MULTIAT` proposée fournit des contours automatiques de meilleure qualité que ceux obtenus en utilisant un atlas moyen classique (méthode `AVAT`) : visuellement, les frontières entre dents adjacentes sont globalement mieux localisées, et quantitativement les mesures `DICE`, `VOL` et `RHD $\tau=95\%$` sont significativement meilleures. Cependant, nous avons également montré que cette meilleure segmentation des dents ne se répercutait pas de façon significative sur l'estimation de la dose maximum reçue par chaque dent. En effet, les erreurs d'estimation de dose sont globalement inférieures avec `MULTIAT` qu'avec `AVAT`, mais les différences ne sont pas statistiquement significatives. Cela signifie que les améliorations en terme de contours sont trop petites à l'échelle des gradients de dose pour ces patients. Une analyse sur davantage de patients permettrait de savoir si ce constat est spécifique aux 8 patients testés ou pas. Si cette tendance est confirmée (notamment sur des cas de distributions de dose très complexes obtenues par `CyberKnife`), l'utilisation de l'atlas moyen permettrait de réduire le coût en temps de calcul par rapport à la méthode proposée tout en fournissant une estimation de dose de précision similaire.

7.5.2 Comparaison avec la dose sur la mandibule ou le maxillaire

Dans la littérature, les risques d'ostéoradionécrose ou d'échec d'implant sont souvent estimés à partir de la dose reçue par la mâchoire sous-jacente (maxillaire ou mandibule) dans son intégralité, c'est-à-dire sans se restreindre localement à la dent considérée. Dans notre étude, nous avons considéré la dose reçue par chaque dent comme représentative de la dose reçue par sa racine (et donc localement par l'os sous-jacent), ce que nous considérons comme *a priori* plus précis. En effet, l'estimation de la dose sur l'intégralité de la mâchoire sous-jacente peut conduire à une surestimation du risque de complications dentaires, si par exemple une dose élevée a été reçue localement à un endroit éloigné de la dent considérée sur la mâchoire. D'autre part, les risques d'ostéoradionécrose ne sont pas les mêmes sur toutes les dents, par exemple pour la mandibule ils sont supérieurs pour les segments postérieurs. Estimer le risque localement au niveau de chaque dent semble donc plus pertinent et plus précis. Toutefois, afin de réellement prendre en compte la dose reçue par l'os sous-jacent (et pas seulement par la racine de la dent), on pourrait envisager de compléter les segmentations automatiques des dents de quelques coupes du côté des racines, tout en s'assurant que les contours rajoutés sur ces coupes correspondent bien à une structure osseuse et pas à des tissus mous.

Nous avons malgré tout appliqué notre méthode pour la segmentation de la mandibule et du maxillaire, et obtenu une estimation de la dose maximum avec une précision comprise dans l'intervalle $[-1\text{Gy}; +1\text{Gy}]$ pour les 8 patient test, ce qui tout à fait acceptable.

D'autre part, lorsque la dose est estimée sur la mâchoire, il est souvent plus courant d'estimer la dose maximale reçue sur un certain volume (en absolu ou en pourcentage du volume de la mâchoire) au lieu d'estimer la dose maximale ponctuelle comme nous l'avons fait. Par exemple, [Studer 2006] reportent un risque d'ostéora-

dionécrose lorsque un volume supérieur à 1 cm^3 de la mandibule a reçu plus de 70 Gy. Cependant, comme les dents sont des très petites structures (moins de 2 cm^3 pour 97 % des cas dans notre base), les recommandations valables pour la mandibule et le maxillaire ne s'appliquent pas, et nous avons considéré que l'estimation ponctuelle de la dose maximum avait davantage de sens.

7.5.3 Applications cliniques

Tout d'abord, la méthode proposée permet de fournir une estimation automatique de la dose reçue par chaque dent. L'avantage est que l'estimation de dose peut se faire directement lors de la planification de traitement car elle ne nécessite pas de temps médecin. Les doses maximales reçues par chaque dent peuvent alors être stockées directement dans le dossier du patient et sont donc directement accessibles par le dentiste.

Le contournage automatique des dents peut avoir d'autres applications cliniques intéressantes dans le cadre de la planification en radiothérapie des cancers ORL. Par exemple, la segmentation automatique des dents peut permettre d'analyser les problèmes dosimétriques liés aux artefacts. En effet, les artefacts dentaires introduisent un biais dans les calculs de dose dans les TPS et limitent l'optimisation de la dosimétrie. Si l'on dispose des contours automatiques des dents, on peut alors remplacer artificiellement les densités dans les dents plombées par les densités réelles des amalgames. Cela ne permettra pas de s'affranchir de l'artefact sur les tissus mous environnants mais permettra au moins de corriger la densité de l'amalgame, qui est souvent mal estimée par les systèmes.

D'autre part, si un arrachage de dents est décidé après l'acquisition du scanner de planification, son impact sur la dosimétrie peut être facilement estimé en remplaçant sur l'image scanner les densités à l'intérieur des contours automatiques des dents concernées par la densité de l'air.

7.5.4 Perspectives

Les perspectives de ce travail sont principalement liées à la validation. En effet, une validation de la méthode sur davantage de patients est nécessaire pour confirmer les résultats quantitatifs présentés à partir de nos 8 patients test. Une telle validation serait également nécessaire pour déterminer si les différences observées entre la méthode proposée et l'utilisation de l'atlas moyen sont significatives ou pas en terme d'estimation de dose.

Une validation multi-centrique de la méthode via le logiciel ISOgray est en cours de lancement au sein du GORTEC (Groupe d'Oncologie Radiothérapie des tumeurs de la Tête Et du Cou). Cette validation permettra notamment de tester la robustesse de la méthode pour des patients ayant subi une radiothérapie par CyberKnife (et donc avec des gradients de dose plus abrupts), ou pour des patients avec un bouchon buccal dans la bouche sur l'image scanner. Il serait également intéressant d'étudier la robustesse de la méthode en présence de très importants artefacts dentaires puisque

ce type de cas avait été écarté dans notre base.

D'autre part, l'augmentation de la base d'images utilisées pour segmenter un patient test donné (et notamment avec des images avec coupes très fines et sans artefacts) permettra potentiellement d'améliorer la qualité des résultats.

7.6 Conclusion et synthèse des contributions

Dans ce chapitre, nous avons proposé une méthode permettant de segmenter automatiquement les dents sur une image scanner pour en déduire les doses reçues par chaque dent lors du traitement. Ces informations de dose peuvent alors permettre au dentiste d'estimer les risques de complications associés à une extraction dentaire ou à la pose d'un implant après la radiothérapie.

Nous avons proposé une analyse quantitative en terme de qualité de contours des contours automatiques obtenus (par rapport aux contours manuels), et en terme de précision de la dose estimée à partir des contours automatiques (en comparaison de l'estimation de dose à partir des contours manuels). Dans cette analyse, nous avons considéré les doses reçues par chaque dent comme étant représentatives des doses reçues par leurs racines (et donc par les os sous-jacent). Nous avons présenté les résultats obtenus sur nos 8 patients test. Sur ces patients, la méthode proposée a permis de déterminer les doses reçues par chaque dent avec une précision inférieure à 2 Gy en moyenne (et également inférieure à 2 Gy dans 77 % des cas parmi les 214 dents considérées). Ce niveau de précision est suffisant d'un point de vue clinique pour l'estimation des risques associés. D'autre part, nous avons montré que ce degré de précision est bien supérieur à celui utilisé en routine clinique avec la méthode d'estimation visuelle de la dose sans aucun contours (qui est de l'ordre de 5 Gy en moyenne).

De plus, nous avons comparé notre méthode de segmentation automatique des dents à l'utilisation d'un atlas moyen classique. Nous avons montré que les résultats obtenus avec la méthode proposée sont significativement meilleurs que ceux obtenus avec l'atlas moyen en terme de qualité de contours par rapport aux contours manuels. En revanche, bien que légèrement favorables à la méthode proposée, les différences entre les deux méthodes sont non significatives en terme d'estimation de la dose maximum.

Enfin, la segmentation automatique des dents peut également être utilisée à d'autres fins dans le contexte de la planification en radiothérapie, comme par exemple pour l'analyse dosimétrique des problèmes causés par les artefacts.

Conclusion

Sommaire

8.1 Synthèse des contributions	207
8.1.1 Construction d'une segmentation moyenne	208
8.1.2 Construction d'atlas anatomiques adaptés au patient	209
8.1.3 Segmentation automatique des dents	212
8.1.4 Intégration dans ISOgray et validation clinique	213
8.2 Perspectives	214
8.2.1 Perspectives liées aux données	214
8.2.2 Perspectives liées à l'évaluation	215
8.2.3 Perspectives liées à la méthodologie	215

8.1 Synthèse des contributions

Dans cette thèse, nous avons proposé des algorithmes de segmentation automatique basés sur l'utilisation d'atlas anatomiques. Le fil conducteur de nos travaux peut se résumer par la question suivante : comment exploiter au mieux une base d'images manuellement contourées pour concevoir des atlas anatomiques adaptés à l'anatomie du patient à segmenter ? Cette problématique est d'un intérêt majeur pour la segmentation par atlas, car l'utilisation d'atlas similaires au patient permet d'atteindre une meilleure précision lors du recalage et par conséquent une segmentation automatique de meilleure qualité. Nous avons appliqué et évalué les algorithmes proposés à la segmentation des structures de la région ORL pour la radiothérapie.

Dans le but d'exploiter au mieux la base d'atlas, nous avons successivement étudié les deux axes de recherche suivants. Tout d'abord, nous avons exploré la stratégie de l'utilisation d'un atlas unique pour tous les patients. Dans ce cas, l'atlas similaire au plus grand nombre de patients possible est un atlas moyen construit à partir de la totalité de la base. Nous avons donc cherché à optimiser la construction d'un tel atlas, et en particulier la construction de la segmentation moyenne. Afin de dépasser les limites de l'atlas moyen unique (et en particulier ses difficultés à gérer la grande variabilité anatomique de la région ORL), nous avons ensuite étudié et défini des algorithmes de construction d'atlas basés sur la stratification de la base, ou encore sur la sélection et la fusion des atlas de la base les plus pertinents pour chaque patient.

Les principales difficultés auxquelles nous avons été confronté lors de ce travail de recherche sont la variabilité anatomique inter-individuelle et la variabilité de contourage présentes dans la région ORL. Du point de vue de cette application, notre objectif était double : à la fois améliorer globalement la qualité des segmentations automatiques et réduire les cas d'échecs en définissant des algorithmes les plus robustes possibles aux anatomies atypiques (patient corpulent ou très maigre, flexion de colonne importante).

Nous détaillons dans la suite les principales contributions de cette thèse.

8.1.1 Construction d'une segmentation moyenne

Dans un premier temps, nous avons proposé dans le chapitre 4 plusieurs solutions pour optimiser la construction d'un atlas moyen à partir de la base d'images manuellement segmentées, en particulier la phase d'estimation d'une segmentation moyenne. La principale difficulté liée à ce travail est la gestion adéquate des contours manquants dans les segmentations manuelles de notre base.

Tout d'abord, nous avons présenté en section 4.2.4 une analyse cohérente de la sur-segmentation obtenue en utilisant l'algorithme STAPLE [Warfield 2004]. Nous avons en effet montré que cette sur-segmentation provient pour une faible part d'une approximation dans la prise en compte des contours manquants dans l'implémentation initiale (que nous avons corrigée), et en majeure partie de la très grande variabilité anatomique et de contourage qui conduit à un masque de calcul de grande taille.

Nous avons mis en place une correction *a posteriori* de la sur-segmentation basée sur des érosions optimales pour chaque structure. L'atlas moyen résultant a été évalué en clinique, et cela a donné lieu à une publication de type résumé dans une conférence française de radiothérapie :

- A. Isambert, V. Grégoire, F. Bidault, M. Benkebil, **L. Ramus**, G. Malandain, J. Bourhis et D. Lefkopoulos. *Évaluation d'un algorithme de segmentation automatique des structures d'intérêt dans la sphère ORL*. Actes de la SFRO'08, 2008.

La correction *a posteriori* de la sur-segmentation n'est pas satisfaisante du point de vue méthodologique et elle ne permet pas de prendre en compte des amplitudes de sur-segmentation différentes selon les endroits des structures. Nous avons donc proposé en section 4.3.1 une approche de gestion des contours manquants pouvant s'appliquer à n'importe quel algorithme de calcul de segmentation moyenne. Cette approche permet de gérer le statut particulier du fond de l'image (qui ne correspond pas à la même entité pour toutes les segmentations de la base) par rapport aux structures anatomiques. Le principe consiste à ne pas mettre en compétition simultanément l'ensemble des labels $s \in [0 \dots L]$ les uns contre les autres, mais à estimer pour chaque structure $s \in [1 \dots L]$ indépendamment la probabilité de chaque voxel d'appartenir ou de ne pas appartenir à cette structure. Pour chaque voxel, la probabilité d'appartenance au fond est alors définie comme la probabilité de ne pas appartenir à la structure anatomique la plus probable en ce voxel. Nous avons

montré que l'application de ce schéma de calcul à l'algorithme STAPLE permet de réduire significativement la sur-segmentation et d'atteindre des structures moyennes de taille acceptable.

Par ailleurs, nous avons proposé en section 4.4 un algorithme de calcul de segmentation moyenne basé sur l'estimation du degré d'accord local entre les segmentations manuelles de la base. Cet algorithme a fourni des résultats comparables en terme de qualité de segmentation à ceux obtenus avec la précédente approche de gestion des structures manquantes appliquée à l'algorithme STAPLE. Une version intermédiaire de l'algorithme proposé a donné lieu à une publication dans une conférence internationale sélective avec comité de revue et à une présentation orale à cette conférence :

- **L. Ramus** et G. Malandain. *Using Consensus Measures for Atlas Construction*. Proceedings of ISBI'09, Boston, Massachusetts, USA, pages 1190-1193, 2009. IEEE.

8.1.2 Construction d'atlas anatomiques adaptés au patient

Tout d'abord, un état de l'art des différentes stratégies de segmentation par atlas pour la radiothérapie de la région ORL a été publiée dans un journal en français :

- **L. Ramus**, J. Thariat, P.Y. Marcy, Y. Pointreau, G. Bera, O. Commowick et G. Malandain. *Outils de contourage, utilisation et construction d'atlas anatomiques : exemples des cancers de la tête et du cou*. Cancer/Radiothérapie, 14(3), pages 206-212, 2010.

Nous avons présenté et évalué dans les chapitres 5 et 6 respectivement des algorithmes de construction d'atlas adaptés à l'anatomie du patient. Ces algorithmes étaient soit basés sur la stratification de la base d'atlas, soit sur un schéma multi-atlas comportant des étapes de sélection et de fusion des atlas de la base les plus pertinents pour chaque patient. Nous détaillons dans la suite les contributions méthodologiques relatives à la sélection d'atlas en section 8.1.2.1 et à la fusion d'atlas en section 8.1.2.2, puis les contributions liées à l'évaluation en section 8.1.2.3.

8.1.2.1 Contributions méthodologiques liées à la sélection d'atlas

Nous avons proposé deux nouveaux critères de sélection d'atlas en section 5.2.5, et un protocole non biaisé d'évaluation des critères de sélection d'atlas en section 5.3.

Le premier critère de sélection proposé est basé sur des méta-informations adaptées à la radiothérapie des tumeurs ORL, à savoir la localisation et le stade de la tumeur primitive. Pour analyser ce critère de sélection, nous l'avons utilisé pour stratifier la base d'atlas et construire ensuite un sous-atlas moyen pour chaque catégorie. Évalué pour les tumeurs de l'oropharynx seulement, le critère proposé s'est révélé pertinent pour les tumeurs de stade N2 seulement pour lesquelles il a fourni des meilleurs résultats qu'une sélection aléatoire. Cependant, le sous-atlas correspondant n'a pas permis de dépasser les performances de l'atlas moyen global. Ce

critère de sélection nécessite donc probablement une base d'images beaucoup plus grande que la nôtre pour que le sous-atlas soit estimé de façon suffisamment robuste. Ce travail a donné lieu à une publication sur résumé dans une conférence internationale :

- **L. Ramus**, S. Aouadi, Y. Pointreau, G. Bera, Y. Tao, A. Beaudré, V. Grégoire et G. Malandain. *Design and Use of Patient Specific Atlases for Radiotherapy Planning of the Head and Neck Region*. Acts of ESTRO'10, Barcelona, Spain, 2010.

Le second critère de sélection proposé est basé sur l'estimation du degré de contraction ou de dilatation de structures anatomiques. Plus précisément, il consiste à sélectionner pour une région donnée les atlas de la base qui subissent en moyenne sur cette région le même degré de contraction/dilatation que le patient pour se déformer sur l'atlas moyen. Nous avons évalué ce critère dans le cadre de la fusion des segmentations dans le référentiel du patient. Nous l'avons en particulier comparé au critère de sélection basé sur l'amplitude de la déformation non-linéaire [Commowick 2007a]. Les deux critères ont donné des résultats similaires pour la sélection globale. Lorsque la sélection d'atlas est appliquée de façon régionale, notre critère s'est révélé en moyenne sur toutes les structures moins performant que le critère basé sur l'amplitude de la déformation non-linéaire. Le critère de sélection proposé a été publié dans une conférence internationale sélective avec comité de revue :

- **L. Ramus**, O. Commowick et G. Malandain. *Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces*. Proceedings of MIC-CAI'10, Part III, volume 6363 of LNCS, Beijing, China, pages 155-162, 2010. Springer.

Enfin, le protocole proposé pour l'évaluation non biaisée des méthodes de sélection d'atlas est basé sur leur partitionnement à partir des corrélations entre les classements d'atlas qu'elles fournissent, du plus similaire au moins similaire au patient. Ce protocole permet d'évaluer la phase de sélection indépendamment du reste de l'algorithme, c'est-à-dire du nombre d'atlas sélectionnés et de la phase de fusion. Dans l'évaluation des méthodes de sélection d'atlas présentée en section 6.2, ce protocole a conduit à des conclusions proches de celles apportées par l'étude des performances des atlas en fonction de leur rang dans le classement. Ce travail a donné lieu à une publication à une conférence internationale sélective avec comité de revue :

- **L. Ramus** et G. Malandain. *Assessing Selection Methods in the Context of Multi-atlas Based Segmentation*. Proceedings of ISBI'10, Rotterdam, The Netherlands, pages 1321-1324, 2010. IEEE.

8.1.2.2 Contributions méthodologiques liées à la fusion d'atlas

Nous avons proposé deux schémas pour la fusion d'atlas. La première contribution, décrite en section 5.4.3.3, consiste à construire un atlas moyen composite adapté à l'anatomie du patient à partir d'un ensemble d'atlas sélectionnés sur des ré-

gions. L'image de cet atlas moyen composite est construite selon un schéma itératif du type de [Guimond 2000], mais dans lequel les images moyennées sont pondérées par des poids variant spatialement et dépendant également de leur similarité à l'image du patient. Ce travail a donné lieu à une publication dans une conférence internationale sélective avec comité de revue, que nous avons déjà mentionnée dans le paragraphe précédent :

- **L. Ramus**, O. Commowick et G. Malandain. *Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces*. Proceedings of MICCAI'10, Part III, volume 6363 of LNCS, Beijing, China, pages 155-162, 2010. Springer.

La seconde contribution, décrite en section 5.4.4.2, est un schéma de fusion locale pondérée des segmentations des atlas dans le référentiel du patient. Dans ce type d'algorithme, seules les segmentations des atlas sont fusionnées, il n'y a pas de calcul d'image moyenne adaptée au patient comme dans la première contribution. Cette contribution a donné lieu à une publication dans un workshop international sélectif avec comité de revue :

- **L. Ramus** and G. Malandain. *Multi-atlas Based Segmentation : Application to the head and neck region for radiotherapy planning*. Proceedings of MICCAI'10 Workshop Medical Image Analysis for the Clinic - A Grand Challenge, Beijing, China, pages 281-288, 2010.

Ce workshop proposait une base d'images commune à tous les participants et avait pour but de comparer les résultats obtenus sur cette base avec les différents algorithmes présentés pour la segmentation des parotides. Notre algorithme de fusion locale des segmentations a obtenu la troisième place sur six participants.

Ces deux contributions se sont toutes deux avérées significativement plus performantes que l'atlas moyen global. La pondération locale des segmentations a permis d'obtenir les meilleurs résultats. L'atlas moyen composite s'est cependant révélé plus pertinent pour les patients ayant une anatomie atypique.

8.1.2.3 Évaluation approfondie des approches de segmentation multi-atlas

Nous avons également proposé en section 6.3 une évaluation approfondie des méthodes de segmentation appelées *multi-atlas* dans la littérature, c'est-à-dire les méthodes basées sur la fusion des segmentations des atlas après propagation dans le référentiel du patient. Nous avons en particulier étudié (1) l'influence de l'échelle à laquelle est faite la sélection d'atlas (globale, régionale ou locale), (2) l'influence de la phase de propagation (directe ou indirecte via l'atlas moyen global), (3) l'influence du nombre d'atlas fusionnés, et (4) l'influence du critère de sélection utilisé.

Concernant l'échelle de sélection des atlas, nous avons montré que la sélection est d'autant plus pertinente qu'elle est ciblée : la pondération locale est plus pertinente que la sélection régionale, qui est elle-même plus pertinente que la sélection globale, qui fournit quant à elle des résultats similaires à l'atlas moyen global. Pour les sélections régionale et locale, nous avons également mis en évidence que la propa-

gation indirecte donne des résultats significativement moins bons que la propagation directe dès lors que le critère de sélection d’atlas est pertinent. Dans le cas de la pondération locale cependant, les différences entre les segmentations automatiques obtenues par propagation directe ou indirecte sont petites, et elles restent marginales au niveau visuel. Le meilleur compromis entre précision des contours et rapidité est donc la pondération locale des segmentations des atlas après propagation indirecte dans le référentiel du patient. Cette approche a par conséquent été choisie pour être intégrée dans ISOgray.

Enfin, nous avons mis en évidence que la propagation directe n’est préférable à la propagation indirecte que dans le cas où la méthode de sélection d’atlas est suffisamment pertinente. Dans le cas où les atlas sélectionnés ne sont pas suffisamment similaires au patient, la propagation indirecte donne de meilleurs résultats. On se retrouve typiquement dans ce cas de figure lorsque la méthode de sélection utilisée est peu robuste, ou lorsqu’un trop grand nombre d’atlas est sélectionné.

8.1.3 Segmentation automatique des dents

Dans le chapitre 7, nous avons mis en œuvre le principe de segmentation multi-atlas pour la segmentation des dents, c’est-à-dire la fusion des segmentations des atlas après propagation directe dans le référentiel du patient. L’objectif de ce travail est de fournir au dentiste une estimation *a posteriori* de la dose reçue par chaque dent pour qu’il puisse évaluer le niveau de risque associé à un soin dentaire post-irradiation donné (typiquement l’extraction d’une dent ou la pose d’un implant). La principale difficulté liée à la segmentation automatique des dents est la présence de nombreux artefacts dus aux plombages ainsi qu’une grande variabilité anatomique (surtout au niveau des racines).

Concernant la méthodologie, nous avons fait ici des choix différents de ceux faits pour la segmentation des structures ORL. En effet, la présence de nombreux artefacts dentaires sur les images CT introduit des biais d’intensité qui rendent la sélection d’atlas particulièrement sensible. Cela est particulièrement vrai pour la sélection locale qui s’était révélée la plus pertinente pour les structures ORL. De plus, le nombre d’images disponibles est plus faible en comparaison de la base ORL. Pour ces deux raisons, nous avons choisi d’utiliser l’ensemble des atlas disponibles sans effectuer de sélection. Par ailleurs, nous avons choisi une propagation directe de chaque atlas sur le patient car les nombreux artefacts dentaires contribuent à une image moyenne très floue au niveau des dents ce qui réduirait de façon non négligeable les performances obtenues avec une propagation indirecte via l’atlas moyen.

L’algorithme multi-atlas proposé s’est révélé significativement plus performant que l’utilisation d’un atlas moyen en terme de qualité de contours par rapport aux contours manuels. Cependant, cette amélioration ne s’est pas répercutée de façon significative sur l’estimation de la dose : les différences étaient favorables à l’algorithme multi-atlas mais trop petites à l’échelle des gradients de dose et non significatives. Nous avons également montré que l’algorithme multi-atlas proposé fournit une pré-

cision en terme de dose de l'ordre de 2 Gy en moyenne. Ce degré de précision est bien meilleur à celui atteint en routine clinique (c'est-à-dire avec la méthode d'estimation visuelle de la dose sans aucun contours).

Les travaux liés à la segmentation des dents ont donné lieu à un article de journal en co-première auteur actuellement en presse et à plusieurs conférences sur résumé en deuxième auteur :

- J. Thariat and **L. Ramus (co-first authors)**, P. Maingon, G. Odin, V. Grégoire, V. Darcourt, P.Y. Marcy, N. Guevara, M.H. Orlanducci, S. Marcie, G. Poissonnet, A. Bozec, O. Dassonville, F. Demard, L. Castillo, J. Santini and G. Malandain. *DENTALMAPS : Automatic Dental Delineation for Radiotherapy Planning in Head and Neck Cancer*. Int J Radiat Oncol Biol Phys, In Press, 2011.
- J. Thariat, **L. Ramus**, P. Maingon, V. Grégoire, V. Darcourt, S. Marcie et G. Malandain. *Atlas de contourage automatique des dents pour l'optimisation des soins dentaires*. Actes de la SFRO'10, 2010.
- J. Thariat, **L. Ramus**, P. Maingon, V. Grégoire, S. Marcie et G. Malandain. *Construction of a Dental Atlas for Automatic Teeth Segmentation to Improve Post-irradiation Dental Care Management*. Acts of ASTRO'10, 2010.

8.1.4 Intégration dans ISOgray et validation clinique

Les principales contributions de cette thèse ont été intégrées ou sont en cours d'intégration dans le logiciel ISOgray de DOSIsoft.

Tout d'abord, l'atlas moyen construit à partir de l'algorithme proposé en section 4.3.2 (approche de gestion des structures manquantes appliquée à l'algorithme STAPLE) a remplacé l'atlas construit avec STAPLE dans le module de segmentation de ISOgray. Bien que non détaillé dans la thèse, des structures anatomiques ont été ajoutées dans cet atlas moyen, notamment les aires ganglionnaires IB et V ainsi que deux jeux différents de parotides : un avec lobe accessoire et l'autre sans lobe accessoire (le médecin devant alors choisir la segmentation adéquate suivant si le patient a ou n'a pas de lobe accessoire). Une validation clinique a eu lieu et a confirmé l'amélioration par rapport à la version précédente de l'atlas moyen.

Concernant les algorithmes multi-atlas adaptés au patient, la méthode de pondération locale des segmentations des atlas dans le référentiel du patient après propagation indirecte (meilleur compromis qualité/temps de calcul) a été intégrée au module de segmentation des structures ORL et est en cours de validation.

Enfin, la segmentation des dents a été intégrée dans un module destiné à la validation clinique. Cette validation clinique qui est actuellement en cours va permettre de confirmer ou d'infirmer les conclusions déduites sur notre base d'images concernant les performances relatives de l'algorithme multi-atlas et de l'atlas moyen. De plus, la robustesse des deux approches va être testée sur des patients traités par CyberKnife et présentant des forts gradients de dose.

8.2 Perspectives

Le travail de recherche effectué lors de cette thèse a ouvert la voie à de nombreuses perspectives que nous détaillons dans cette section.

8.2.1 Perspectives liées aux données

Dès lors qu'une sélection ou une pondération des atlas est appliquée, la base d'atlas a tout intérêt à être la plus grande possible pour optimiser les performances. Le seul inconvénient lié à l'agrandissement de la base d'atlas est le temps de calcul qui est augmenté en fonction si les atlas sont mis en correspondance avec le patient de façon directe, c'est-à-dire sans utiliser l'atlas moyen comme référentiel intermédiaire. Pour la méthode que nous avons choisie (pondération locale des segmentations après propagation indirecte), le surcoût en temps de calcul lié à l'augmentation de la base d'atlas serait très faible, ce qui constitue donc un avantage supplémentaire en faveur de cette méthode. Une perspective importante est donc l'augmentation de la base d'atlas et l'évaluation des performances ainsi atteintes.

Pour la segmentation des structures ORL, il serait intéressant d'évaluer le critère de sélection basé sur la localisation et le stade de tumeur avec une base d'atlas plus grande pour déterminer si les performances obtenues dépassent l'atlas moyen global. Pour la stratification de la base selon ce critère et la construction de sous-atlas, une base d'atlas plus grande et dans laquelle toutes les catégories sont suffisamment représentées est indispensable pour une utilisation clinique. D'autre part, la symétrisation de la base d'atlas n'a été utilisée que pour les sous-atlas par localisation et stade de tumeur. La symétrisation de la base d'atlas pourrait permettre d'améliorer les résultats pour tous les autres algorithmes proposés. L'évaluation de cette amélioration est une piste importante pour l'exploitation optimale des données.

Pour la segmentation des dents, la base d'atlas utilisée est de plus petite taille que pour les structures ORL, et les atlas de cette base comportent de très nombreux artefacts dentaires. L'augmentation de la base d'atlas permettrait peut être d'améliorer les résultats, mais certainement de façon moins marquée que pour l'ORL puisqu'aucune sélection d'atlas n'est appliquée pour les dents. Par contre, l'amélioration de la qualité de la base d'atlas (atlas sans artefacts dentaires et avec des coupes fines) serait probablement un facteur important pour optimiser la qualité des segmentations résultantes. Dans le cadre de l'algorithme multi-atlas, cela permettrait d'améliorer la qualité des recalages non-linéaires entre le patient et les atlas. Cela permettrait également de construire un atlas moyen de bien meilleure qualité, qui pourra peut être alors fournir à moindre coût (un seul recalage) des résultats de segmentation aussi bons que ceux du schéma multi-atlas.

Enfin, les algorithmes proposés pourraient également être testés pour la segmentation automatique d'une autre localisation anatomique, par exemple pour les aires ganglionnaires du thorax.

8.2.2 Perspectives liées à l'évaluation

Au cours de cette thèse, notre objectif était double : améliorer la qualité des segmentations en moyenne sur tous les patients, et réduire les cas d'échecs. Pour l'évaluation des algorithmes, nous avons utilisé dans toute la thèse des mesures quantitatives moyennes sur l'ensemble des patients (en utilisant un schéma de type Leave-One-Out). Cela permet uniquement d'évaluer les performances relatives à notre premier objectif. Il serait intéressant d'évaluer les algorithmes proposés sur des sous-groupes de patients ayant des anatomies atypiques puisque ces patients correspondent en général à la majeure partie des cas d'échecs. En particulier, nous avons remarqué que la construction et l'utilisation d'un atlas composite moyen permettait de surpasser les autres algorithmes pour ce genre de patients. Une étude plus approfondie permettrait de confirmer cette tendance.

Par ailleurs, les images de notre base sont uniquement des cas N0/N1 (70 sur 105) et N2 (35 sur 105). Un point important pour la clinique serait l'étude de la robustesse des algorithmes proposés sur des patients ayant une tumeur de stade N3, et donc entraînant de grandes déformations par rapport à une anatomie normale.

Pour la segmentation des dents, une évaluation clinique est en cours au Centre Antoine Lacassagne, puis une autre sera lancée au sein du GORTEC (Groupe d'Oncologie Radiothérapie des tumeurs de la Tête Et du Cou). Ces études devraient permettre d'étudier la robustesse de l'algorithme face à des gradients de dose importants sur des patients traités par CyberKnife, ainsi que sa robustesse à la présence de blocs d'immobilisation dans la bouche.

Enfin, il serait intéressant d'étudier la corrélation entre nos mesures quantitatives de qualité de segmentation, et l'appréciation visuelle du radiothérapeute. Cette appréciation visuelle pourrait par exemple être quantifiée par des scores allant de 1 à 5 comme cela avait été fait pour les sous-atlas par localisation et stade de tumeur. Menée sur une dizaine de patients, une telle étude permettrait de confirmer (ou d'infirmer) la validité clinique de l'évaluation des algorithmes que nous avons conduite sur notre base d'images.

8.2.3 Perspectives liées à la méthodologie

8.2.3.1 Stratification de la base

Dans cette thèse, nous n'avons étudié la stratification de la base que d'après le critère de localisation et de stade de tumeur. Avec ce critère de stratification, nous n'avons pas obtenu de meilleurs résultats que l'atlas moyen global. Une perspective importante du point de vue méthodologique est la stratification de la base d'après des critères basés sur les images ou les segmentations.

Tout d'abord, une stratification d'après les intensités permettrait de regrouper ensemble les images de la base visuellement similaires. Une telle stratification pourrait être effectuée avec l'algorithme proposé par Blezek et al. [Blezek 2007].

D'autre part, une piste intéressante serait l'étude de la stratification des images de la base selon la similarité entre leurs segmentations manuelles. En effet, dans

toute la thèse, nous n'avons utilisé que des critères de sélection basés intensité ou déformation, mais les contours manuels des images constituent une représentation simplifiée de l'anatomie des atlas qui peut être exploitée au même titre que l'image. Bien entendu, un tel critère basé sur les segmentations des structures d'intérêt ne peut pas être utilisé pour la sélection d'atlas les plus similaires au patient puisque la segmentation du patient est justement l'inconnue du problème. En revanche, ce critère peut être utilisé pour comparer entre elles les images de la base, et les regrouper en sous-groupes homogènes. Concrètement, la stratification pourrait être faite en appliquant l'algorithme d'affinité-propagation [Frey 2007] à partir des indices de Dice moyens sur toutes les structures pour chaque paire d'images. Ensuite, un atlas moyen pourrait être construit pour chaque sous-groupe, et l'atlas moyen le plus approprié pour un patient donné pourrait être sélectionné par des méthodes basées intensité.

Par ailleurs, on s'attend intuitivement à ce que la stratification d'après les segmentations soit plus pertinente appliquée structure par structure (plutôt qu'appliquée à partir des indices de Dice moyens sur toutes les structures). Dans le cas d'une stratification de la base indépendante pour chaque structure, le schéma suivant pourrait être utilisé : pour chaque structure, construction de morceaux d'atlas moyens correspondant aux différents sous-groupes obtenus, puis sélection du morceau d'atlas moyen le plus similaire au patient par un critère basé intensité. Cela permettrait de construire un atlas moyen composite adapté au patient, mais avec une mise en œuvre différente de celle que nous avons proposée dans la thèse, la fusion se faisant ici hors-ligne et non à la volée. Un lissage adapté pourrait permettre d'assurer une certaine continuité aux frontières.

8.2.3.2 Couplage des algorithmes basés atlas avec une classification en intensité

L'ensemble des algorithmes proposés dans cette thèse sont basés sur la mise en correspondance de l'image à segmenter avec un ou plusieurs atlas sélectionnés ou construits de façon adéquate. Les informations contenues dans l'image à segmenter ne sont donc exploitées que pour le recalage et la sélection ou pondération d'atlas, c'est-à-dire uniquement dans le cadre du calcul d'une mesure de similarité entre les atlas et l'image à segmenter : cette mesure de similarité sert soit à définir les correspondances entre l'image et le(s) atlas, soit à définir les poids donnés à chaque atlas pour le calcul de la segmentation du patient. Les algorithmes basés atlas proposés dans cette thèse permettent d'atteindre une bonne précision mais des petites erreurs locales aux frontières des structures persistent souvent.

D'autre part, lorsque les structures sont un minimum contrastées et leurs contours visibles sur l'image, l'intensité d'un voxel ainsi que celles de ses voisins représentent une information pertinente en tant que telle pour son étiquetage. En pratique, comme la plupart de nos structures d'intérêt sont peu contrastées, cette information d'intensité ne peut pas être exploitée d'emblée de façon satisfaisante. Cependant, elle peut avantageusement être combinée aux *a priori* spatiaux fournis

par les algorithmes basés sur les atlas. Nous décrivons dans la suite différentes pistes de travail pour combiner ces deux types complémentaires d'information.

Tout d'abord, les *a priori* spatiaux fournis par les atlas peuvent être utilisés pour définir les zones d'incertitude à la frontière des structures. Pour cela, une solution intuitive consiste à considérer l'ensemble des voxels pour lesquels la probabilité d'appartenance au label de probabilité maximale est inférieure à un certain seuil [Lötjönen 2009, Qazi 2010]. Dans le cas où l'on ne dispose pas de la segmentation probabiliste, la zone d'incertitude peut être définie en soustrayant la segmentation binaire érodée à la segmentation binaire dilatée. Tout voxel situé dans les zones d'incertitude peut ensuite être classifié dans l'une ou l'autre des classes en prenant en compte à la fois l'*a priori* spatial en ce voxel ainsi que son intensité et éventuellement celles de ses voisins. Pour les structures bien contrastées, cette classification peut se faire de façon très simple par des opérations de seuillage : avant de les combiner avec l'algorithme STAPLE, Leung et al. excluent des segmentations des atlas normalisées sur le patient les voxels dont l'intensité est en dehors d'une certaine fenêtre [Leung 2010]. Pour des structures moins contrastées, la classification peut se faire avec un algorithme de type EM [Lötjönen 2009, Murgasova 2009], ou encore avec un algorithme de type graph-cut [van der Lijn 2008, Wolz 2009]. Ces algorithmes nécessitent de définir les distributions d'intensité de chaque structure d'intérêt ainsi que du fond de l'image. Ces distributions d'intensité peuvent être définies à partir d'une base d'apprentissage [van der Lijn 2008] ou à partir de l'image à segmenter directement [Wolz 2009, Lötjönen 2009]. Dans ce dernier cas, elles peuvent être facilement estimées à partir des voxels situés à l'intérieur de la structure mais n'appartenant pas à la zone d'incertitude. Pour les structures suffisamment homogènes (par exemple les parotides, les sous-maxillaires, la mandibule, la moelle, le tronc cérébral), une distribution d'intensité gaussienne nous paraît raisonnable. Enfin, des termes de régularisation pourraient également être pris en compte afin d'assurer des bordures lisses, par exemple à l'aide de champs de Markov comme dans [van der Lijn 2008].

Par ailleurs, Sdika a proposé un algorithme permettant de déformer les segmentations d'un atlas sur l'image à segmenter tout en assurant une certaine cohérence avec la classification des voxels de cette image selon leurs intensités [Sdika 2010]. Concrètement, cette approche nécessite la classification en intensité des deux images (atlas et patient) ainsi que leur mise en correspondance non-linéaire. A partir de ces éléments, chaque voxel x de l'image du patient est alors étiqueté avec le label du voxel le plus proche de x et classifié comme appartenant à la même classe dans l'atlas déformé. Cette méthode permet de repérer et de corriger les imprécisions obtenues avec l'atlas seul. De plus, l'auteur l'a mise en œuvre avec un schéma multi-atlas et a montré de très bons résultats.

Enfin, Wang et al. ont proposé une méthode permettant de corriger les biais de segmentation produits par n'importe quel algorithme de segmentation [Wang 2010a]. Pour toute nouvelle image segmentée par l'algorithme en question, cette méthode utilise une base d'apprentissage d'une part pour détecter les voxels mal classifiés, et d'autre part pour les reclassifier correctement.

Pour conclure, la prise en compte des informations d'intensité en tant que telles

pour la segmentation est la piste la plus importante d'évolution et d'amélioration des algorithmes basés atlas proposés dans cette thèse. Par information d'intensité, nous comprenons toutes les caractéristiques pouvant être calculées pour un voxel donné à partir de son intensité et de celles de ses voisins (intensité, gradient, caractéristiques de texture, etc). Les différentes approches de la littérature évoquées dans ce paragraphe représentent autant de pistes de travail et de réflexion.

8.2.3.3 Couplage des algorithmes basés atlas avec un modèle déformable

Dans le paragraphe précédent, nous avons évoqué des pistes d'amélioration basées sur la combinaison des approches basées atlas avec une classification des voxels selon des informations d'intensité. Une autre façon d'exploiter les informations d'intensité de l'image du patient à segmenter est d'utiliser un modèle déformable. Cette approche est particulièrement adaptée pour récupérer les bordures des structures qui sont un minimum visibles sur l'image. La combinaison des algorithmes basés atlas et des modèles déformables peut se faire de façon séquentielle ou simultanée.

Dans la version séquentielle [Han 2010, Hollensen 2010], les algorithmes basés atlas permettent de fournir une initialisation précise de la segmentation qui sera ensuite affinée avec le modèle déformable. Cette bonne initialisation constitue un élément clé pour la bonne convergence du modèle déformable vers les bordures réelles des structures. Han et al. ont montré que le raffinement de la segmentation par modèle déformable après un algorithme multi-atlas permet d'améliorer de façon flagrante les résultats pour la segmentation des parotides [Han 2010]. Le modèle déformable permet de plus d'assurer une certaine régularité des contours.

Par ailleurs, les modèles déformables et la segmentation par atlas peuvent aussi être mis en œuvre de façon conjointe. Par exemple, Gorthi et al. ont proposé un algorithme dans lequel les avantages des deux types d'approches sont avantageusement combinés pour la segmentation des aires ganglionnaires [Gorthi 2009b]. Des structures voisines des aires ganglionnaires et dont les bords sont bien visibles sur les images (mandibule, vertèbres, trachée) sont définies sur l'atlas et servent à guider le recalage entre l'atlas et le patient. Concrètement, le recalage et la segmentation de ces structures bien contrastées sont ici menés de façon conjointe : la segmentation utilise un modèle déformable par contours actifs, et le champ de déformation est estimé à partir de l'évolution du modèle déformable. Le champ de déformation résultant est ensuite appliqué aux segmentations des aires ganglionnaires de l'atlas pour les déformer sur le patient.

8.2.3.4 Utilisation d'informations de positions relatives de structures

Une autre piste d'amélioration des algorithmes basés atlas concerne l'utilisation d'informations de positions relatives des différentes structures. Ce type d'information est déjà implicitement exploité dans la segmentation par atlas, mais plusieurs pistes peuvent être envisagées pour les exploiter explicitement.

Une première façon de tirer davantage profit de ce genre d'informations consiste à segmenter des structures supplémentaires. Cette approche a été utilisée dans

[Gorthi 2009b] pour guider le recalage comme nous l'avons mentionné précédemment. Par ailleurs, la limite latérale des aires ganglionnaires est définie en grande partie par la frontière avec le muscle sternocléidomastoïdien. En raison de son homogénéité, ce muscle est plus facile à segmenter automatiquement que les aires ganglionnaires. Sa segmentation par des techniques basées atlas couplées à la prise en compte des intensités (par modèle déformable ou classification par exemple) pourrait alors être utilisée pour corriger les segmentations des aires ganglionnaires sur la frontière commune avec le muscle.

D'autre part, les informations de relations de voisinage entre les structures voisines peuvent être exploitées de façon beaucoup plus poussée. Des connaissances *a priori* peuvent en effet être utilisées pour construire un modèle structurel de l'anatomie à segmenter qui servira à guider la segmentation [Bloch 2005]. En pratique, des contraintes sont fixées en terme d'adjacence, de connexité, de distances et de position relative directionnelle entre les différentes structures à segmenter, et la segmentation s'effectue en recherchant dans l'image les régions qui vérifient toutes les contraintes du modèle. Ces approches pourraient très bien s'appliquer aux structures ORL, par exemple pour modéliser les repères anatomiques délimitant les aires ganglionnaires. Elles pourraient également être utilisées pour prendre en compte les tumeurs dans la segmentation, comme cela a été fait dans [Khotanlou 2007, Atif 2007]. Enfin, Fischl et al. ont également introduit des *a priori* sur les positions relatives des structures par le biais de champs de Markov anisotropiques qui permettent de considérer les probabilités conditionnelles d'appartenance de deux voxels voisins à deux structures données [Fischl 2002].

8.2.3.5 Perspectives spécifiques à la segmentation des structures ORL

Pour finir, nous avons déjà mentionné qu'une raison fréquente d'échec du recalage sur la région du cou est liée à la flexion de la colonne sur l'image du patient à segmenter. Pour s'affranchir de cette cause d'échec, Han et al. ont proposé un algorithme permettant de corriger la position de la tête sur l'image du patient pour la rendre similaire à l'atlas [Han 2009]. Après un premier recalage linéaire, cet algorithme estime en deux dimensions dans le plan sagittal médian deux transformations rigides (translation + rotation) appliquées avec deux centres de rotation distincts : l'un au niveau du tronc cérébral et l'autre approximativement à mi-hauteur de la moelle épinière. La même transformation est ensuite appliquée à chaque coupe sagittale, puis le recalage non-linéaire est effectué. Il serait intéressant d'évaluer l'amélioration apportée par une telle correction de position dans la phase de recalage.

Algorithme de recalage utilisé

Cette annexe décrit les algorithmes de recalage affine et non-linéaire utilisés pour tous les travaux présentés dans cette thèse.

A.1 Algorithme global

Soit une image de référence R et une image flottante F . Nous recherchons la transformation non-linéaire $\hat{T}_{F \leftarrow R}$ qui permet de ré-échantillonner l'image flottante sur l'image de référence, c'est-à-dire $F \circ \hat{T}_{F \leftarrow R}$ superposée à R . Pour calculer $\hat{T}_{F \leftarrow R}$, nous adoptons l'approche hiérarchique décrite dans l'algorithme 1 : nous recherchons d'abord une transformation très contrainte (transformation rigide), puis nous relâchons progressivement les contraintes en augmentant les degrés de liberté (transformation affine puis non-linéaire). A chaque étape, le recalage précédent permet de fournir une bonne initialisation de la transformation recherchée, ce qui limite le risque de tomber dans un optimum local par rapport au cas où l'on rechercherait d'emblée une transformation non-linéaire.

Algorithme 1 Algorithme de recalage permettant de ré-échantillonner une image flottante F sur une image de référence R

- 1: Recalage rigide par bloc : $\hat{T}_{F \leftarrow R}^{rig} = \text{RECRIG}(F, R, T^{init} = Id)$ (cf. A.2).
 - 2: Recalage affine par bloc : $\hat{T}_{F \leftarrow R}^{aff} = \text{RECAFF}(F, R, T^{init} = \hat{T}_{F \leftarrow R}^{rig})$ (cf. A.2).
 - 3: Recalage non-linéaire par bloc : $\hat{T}_{F \leftarrow R}^{nl} = \text{REC�L}(F, R, T^{init} = \hat{T}_{F \leftarrow R}^{aff})$ (cf. A.3).
-

Les trois recalages successifs (rigide, affine, non-linéaire) utilisent tous une approche de recalage par appariement de blocs et sont décrits plus en détail dans les sections A.2 (rigide, affine) et A.3 (non-linéaire).

A.2 Algorithme utilisé pour le recalage rigide et affine

Pour les recalages rigides et affines, nous utilisons l'algorithme par appariement de blocs proposé par Ourselin et al. [Ourselin 2000], et résumé dans l'algorithme 2. Cet algorithme est encapsulé dans un schéma itératif proche de celui de l'algorithme ICP (Iterative Closest Point) : il alterne entre appariement des primitives et estimation de la correction de transformation selon une minimisation aux moindres carrés tamisés pondérés (Least Trimmed Squares Weighted). Il utilise également une approche multi-résolution.

Algorithme 2 Algorithme de recalage rigide/affine par appariement de blocs

-
- 1: $T^0 \leftarrow T^{init}$.
 - 2: **pour** $p = 1 \dots M$, itérations sur les niveaux de pyramide, **faire**
 - 3: **pour** $l = 1 \dots L$, itérations, **faire**
 - 4: Calcul d'appariements (x_v, y_v) entre les images R et $F \circ T^{l-1}$ par appariement de blocs (cf. A.2.1).
 - 5: Calcul de correction de transformation δT^l par moindres carrés tamisés pondérés (cf. A.2.2).
 - 6: Composition de la correction : $T^l = T^{l-1} \circ \delta T^l$.
-

A.2.1 Estimation des appariements

Les primitives, appelées blocs, sont ici des points x_v régulièrement répartis sur l'image de référence R et enrichis de leur voisinage proche. On note $B(x_v) \subset R$ le bloc centré en x_v . A une itération l donnée, chacun de ces blocs $B(x_v)$ est apparié au bloc $B(y_v)$ de l'image flottante qui lui est le plus similaire selon un critère iconique local. On recherche y_v en se restreignant à un voisinage de x_v dans l'image flottante courante : $y_v \in V(x_v) \subset F \circ T^{l-1}$. Pour nos images, les blocs que nous considérons sont relativement petits ($5 \times 5 \times 5$), et ils ne contiennent donc en général que deux tissus différents au maximum. Par conséquent, l'hypothèse de la relation affine entre les intensités des deux blocs paraît raisonnable, et nous utilisons donc le coefficient de corrélation au carré comme mesure de similarité entre les blocs :

$$B(y_v) = \arg \max_{B(y), y \in V(x_v)} CC^2(B(x_v), B(y)) \quad (\text{A.1})$$

$$= \arg \max_{B(y), y \in V(x_v)} \frac{Cov^2(B(x_v), B(y))}{Var(B(x_v))Var(B(y))} \quad (\text{A.2})$$

A.2.2 Estimation de la correction de transformation

La correction de transformation δT^l expliquant au mieux ces nouveaux appariements (x_v, y_v) est estimée par moindres carrés tamisés pondérés (Least Squares Trimmed Weighted). Cela permet de rejeter automatiquement les appariements aberrants (c'est-à-dire mal expliqués par la déformation calculée), tout en pondérant les appariements restants par les valeurs de similarités atteintes (c'est-à-dire les $CC^2(B(x_v), B(y_v))$). D'autre part, il est également possible d'exclure d'emblée les appariements pour lesquels les blocs de l'image de référence ont une trop faible variance (blocs uniformes).

A.3 Algorithme utilisé pour le recalage non-linéaire

Pour le recalage non-linéaire, nous utilisons une généralisation de l'algorithme présenté dans la section précédente pour le cas des transformations non-linéaires

[Commowick 2007b, Garcia 2010]. Les différentes étapes de cette extension sont les suivantes :

Algorithme 3 Algorithme de recalage dense par appariement de blocs

- 1: Initialisation de la transformation $T^0 \leftarrow T^{init}$.
 - 2: **pour** $p = 1 \dots M$, itérations sur les niveaux de pyramide, **faire**
 - 3: **pour** $l = 1 \dots L$, itérations, **faire**
 - 4: Estimation des appariements : $C \leftarrow \text{apparie}(R, F \circ T^{l-1})$ (cf. A.3.1).
 - 5: Estimation du champ de correction : $\delta T^l \leftarrow \text{interp}(C)$ (cf. A.3.2).
 - 6: Rejet des appariements aberrants : $\tilde{C} \leftarrow \text{prune}(C, \delta T^l)$ (cf. A.3.3).
 - 7: Interpolation du champ de correction : $\delta \tilde{T}^l \leftarrow \text{interp}(\tilde{C})$ (cf. A.3.2).
 - 8: Composition de la correction $T^l = T^{l-1} \circ \delta \tilde{T}^l$.
-

A.3.1 Estimation des appariements

Tout d’abord, l’étape d’estimation des appariements est similaire à celle présentée dans A.2.1. Elle permet de récupérer un ensemble de couples (x_v, y_v) auxquels sont associées les valeurs de similarité atteintes pour ces appariements, c’est-à-dire les valeurs $CC_v = CC^2(B(x_v), B(y_v))$. Ces valeurs permettent de quantifier la confiance que l’on peut avoir dans les appariements.

A.3.2 Estimation de la correction de transformation

Les appariements fournis par l’étape précédente peuvent être vus comme un champ de déplacement épars C , pour lequel $C(x_v) = y_v - x_v$ pour les points x_v étant des centres de blocs dans R , et $C(x) = 0$ pour tous les autres points x . De même, un champ scalaire de confiance dans les appariements peut être construit à partir des $CC_v : k(x_v) = CC_v$ pour les points x_v étant des centres de blocs dans R , et $k(x) = 0$ pour tous les autres points x .

Afin d’estimer un champ dense à partir du champ épars C , Commowick et al. convoluent le champ épars C avec une gaussienne 3D G_σ tout en prenant en compte le champ scalaire de confiance des appariements :

$$\delta T^l = \frac{G_\sigma * kC}{G_\sigma * k} \quad (\text{A.3})$$

Dans cette convolution, l’écart-type σ permet de régler la largeur de la gaussienne selon l’espace entre les blocs dans l’image de référence. Ce lissage du champ épars C peut être vu comme une régularisation fluide.

A.3.3 Rejet des appariements aberrants

L’étape précédente fournit un champ de déplacement dense δT^l . Cependant, les appariements initiaux (x_v, y_v) peuvent être erronés, par exemple sur les régions homogènes, ou encore en présence de tumeurs sur une des deux images. L’étape

précédente s'apparente à une régularisation de type fluide : elle ne permet que de réduire l'impact de ces appariements aberrants, mais pas de les éliminer.

Commowick et al. proposent de détecter les appariements aberrants en comparant pour chaque centre de bloc x_v les déplacements avant ($C(x_v) = y_v - x_v$) et après ($\delta T^l(x_v)$) interpolation du champ épars. Les couples (x_v, y_v) pour lesquels ces deux déplacements sont trop différents seront alors considérés comme appariements aberrants.

Plus précisément, la différence moyenne entre ces deux déplacements avant et après interpolation est notée e , et son écart-type σ_e :

$$e = \frac{1}{N} \sum_v \|C(x_v) - \delta T^l(x_v)\|, \quad (\text{A.4})$$

$$\sigma_e^2 = \frac{1}{N} \sum_v (e - \|C(x_v) - \delta T^l(x_v)\|)^2 \quad (\text{A.5})$$

où N est le nombre de blocs dans l'image de référence. A partir de ces deux valeurs, un appariement (x_v, y_v) est alors considéré comme aberrant s'il vérifie la condition suivante :

$$\|C(x_v) - \delta T^l(x_v)\| > e + \alpha \sigma_e. \quad (\text{A.6})$$

Une fois que les appariements aberrants ont été détectés, ils sont supprimés du champ épars C , et l'interpolation de la transformation est recalculée comme à l'étape A.3.2. Cela fournit le champ de correction $\delta \tilde{T}^l$ qui sera composé avec T^{l-1} .

Mesures d'évaluation de segmentation automatique

Cette annexe décrit les mesures quantitatives permettant de comparer une segmentation automatique à une segmentation de référence (section B.1) puis leur mise en œuvre pour le cas particulier des structures de la sphère ORL (section B.2).

B.1 Mesures d'évaluation

L'évaluation d'une méthode de segmentation automatique se fait en général en comparant pour un ensemble d'images les segmentations automatiques avec des segmentations de référence. Pour les images médicales, les segmentations de référence correspondent typiquement à des segmentations manuelles faites par un médecin. On considère ici l'évaluation de segmentations binaires. Pour des segmentations multilabels, les méthodes d'évaluation présentées ici sont appliquées label par label indépendamment.

On peut globalement classer les mesures de comparaison de segmentations 3D selon deux types : celles qui comparent les volumes en tant que tels (détaillées en section B.1.1), et celles qui comparent les surfaces (détaillées en section B.1.2). Dans la suite, on adopte les notations suivantes : la segmentation automatique sera représentée par le volume V_{auto} et la surface S_{auto} , et la segmentation de référence par le volume V_{ref} et la surface S_{ref} .

B.1.1 Mesures volumiques

Les mesures volumiques consistent à comparer les volumes V_{auto} et V_{ref} .

B.1.1.1 Comparaison absolue des deux volumes

Dans certains cas, il peut être intéressant de comparer les tailles relatives des deux volumes V_{auto} et V_{ref} indépendamment de leur position spatiale. Cela peut notamment permettre de mettre en évidence une sur-segmentation (segmentation automatique trop large par rapport à la segmentation de référence) ou à l'inverse une sous-segmentation. Pour cela, nous proposons d'utiliser la mesure suivante :

$$\text{VOL}(V_{\text{auto}}, V_{\text{ref}}) = \log \left(\frac{\text{card}(V_{\text{auto}})}{\text{card}(V_{\text{ref}})} \right) \quad (\text{B.1})$$

Cette mesure a l'avantage de traduire les sous- et sur-segmentations de façon symétrique par rapport au cas $V_{\text{auto}} = V_{\text{ref}}$. En effet, pour tout facteur $\lambda > 0$, on a : $\text{VOL}(\lambda \cdot V_{\text{ref}}, V_{\text{ref}}) = -\text{VOL}(\frac{V_{\text{ref}}}{\lambda}, V_{\text{ref}})$.

B.1.1.2 Mesures de chevauchement entre les deux volumes

Dans la plupart des cas, on ne souhaite pas simplement comparer les volumes en absolu, mais plutôt comparer leurs positions relatives et leur chevauchement. Pour cela, différentes mesures existent dans la littérature.

Très fréquemment utilisé, l'indice de Dice [Dice 1945] est défini par :

$$\text{DICE}(V_{\text{auto}}, V_{\text{ref}}) = \frac{2 \cdot \text{card}(V_{\text{auto}} \cap V_{\text{ref}})}{\text{card}(V_{\text{auto}}) + \text{card}(V_{\text{ref}})} \tag{B.2}$$

D'autres mesures, comme la sensibilité et la spécificité, sont basées sur la quantification du nombre de voxels correctement ou incorrectement classifiés, c'est-à-dire sur les quantités suivantes illustrées sur la figure B.1 :

- **les Vrais Positifs (VP)** : voxels étiquetés comme appartenant à l'objet à la fois dans la segmentation de référence et dans la segmentation automatique : $VP = V_{\text{auto}} \cap V_{\text{ref}}$;
- **les Vrais Négatifs (VN)** : voxels étiquetés comme n'appartenant pas à l'objet à la fois dans la segmentation de référence et dans la segmentation automatique : $VN = (M \setminus V_{\text{auto}}) \cap (M \setminus V_{\text{ref}})$ où M représente le masque de calcul ;
- **les Faux Négatifs (FN)** : voxels étiquetés comme appartenant à l'objet dans la segmentation de référence, mais pas dans la segmentation automatique : $FN = V_{\text{ref}} \setminus (V_{\text{auto}} \cap V_{\text{ref}})$;
- **les Faux Positifs (FP)** : voxels étiquetés comme appartenant à l'objet dans la segmentation automatique, mais pas dans la segmentation de référence : $FP = V_{\text{auto}} \setminus (V_{\text{auto}} \cap V_{\text{ref}})$.

Le nombre de vrais négatifs dépend directement du masque de calcul M . Un choix courant consiste à utiliser comme masque l'union dilatée de V_{auto} et de V_{ref} . En effet, considérer l'image entière en guise de masque introduit un biais puisqu'alors le nombre de vrais négatifs est disproportionné par rapport aux autres valeurs VP , FP et FN .

A partir de ces mesures, la sensibilité et la spécificité sont définies comme suit :

$$\text{SENS}(V_{\text{auto}}, V_{\text{ref}}) = \frac{\text{card}(VP)}{\text{card}(VP) + \text{card}(FN)} = \frac{\text{card}(V_{\text{auto}} \cap V_{\text{ref}})}{\text{card}(V_{\text{ref}})} \tag{B.3}$$

$$\text{SPEC}(V_{\text{auto}}, V_{\text{ref}}) = \frac{\text{card}(VN)}{\text{card}(VN) + \text{card}(FP)} = \frac{\text{card}((M \setminus V_{\text{auto}}) \cap (M \setminus V_{\text{ref}}))}{\text{card}(M \setminus V_{\text{ref}})} \tag{B.4}$$

Comme le mettent en évidence ces équations, la sensibilité correspond au taux de détection de l'objet, tandis que la spécificité correspond au taux de détection

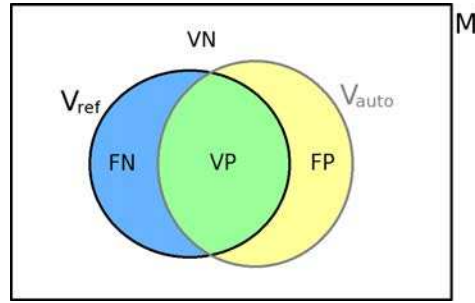


FIG. B.1 – Illustration du statut des voxels à l'intérieur du masque M selon leur localisation par rapport à la segmentation automatique V_{auto} (contour gris) et à la segmentation de référence V_{ref} (contour noir) : Vrais Positifs (VP) en vert, Faux Négatifs (FN) en bleu, Faux Positifs (FP) en jaune, et Vrais Négatifs (VN) en blanc.

du fond. Elle constituent donc des mesures complémentaires. Elles sont toutes deux comprises entre 0 et 1, et leur valeur optimale commune est 1. Une mesure combinée de ces deux critères consiste à mesurer la distance au point ($\text{SENS} = 1, \text{SPEC} = 1$) dans l'espace sensibilité-spécificité selon la formule suivante :

$$\text{DIST}(V_{\text{auto}}, V_{\text{ref}}) = \sqrt{(1 - \text{SENS})^2 + (1 - \text{SPEC})^2} \quad (\text{B.5})$$

Le cas idéal ($\text{SENS} = 1, \text{SPEC} = 1$) correspond au cas $V_{\text{auto}} = V_{\text{ref}}$ et est donc difficilement atteignable. Si l'on est proche de ce point idéal, il est souvent difficile d'améliorer la sensibilité sans dégrader la spécificité et réciproquement. La comparaison de plusieurs méthodes de segmentation automatique est alors difficile. La mesure DIST permet de s'affranchir de ce problème en fournissant une mesure de qualité combinée de la sensibilité et spécificité.

Les mesures de sensibilité et spécificité sont également reliées à la notion de sur- et de sous-segmentation. En effet, une spécificité faible et une sensibilité élevée correspondent en général à une sur-segmentation (peu de FN et beaucoup de FP). Au contraire, une spécificité élevée et une sensibilité faible correspondent à une sous-segmentation (beaucoup de FN et peu de FP). Ceci est illustré sur la figure B.2

B.1.2 Mesures surfaciques

Les mesures surfaciques consistent à quantifier la distance entre les surfaces S_{auto} et S_{ref} . Chacune de ces deux surfaces correspond à un ensemble fini de points, dits points de contours.

B.1.2.1 Distance de Hausdorff

La distance de Hausdorff, introduite dans [Huttenlocher 1993], permet de quantifier la plus élevée des distances point à point entre les surfaces S_{auto} et S_{ref} . Plus

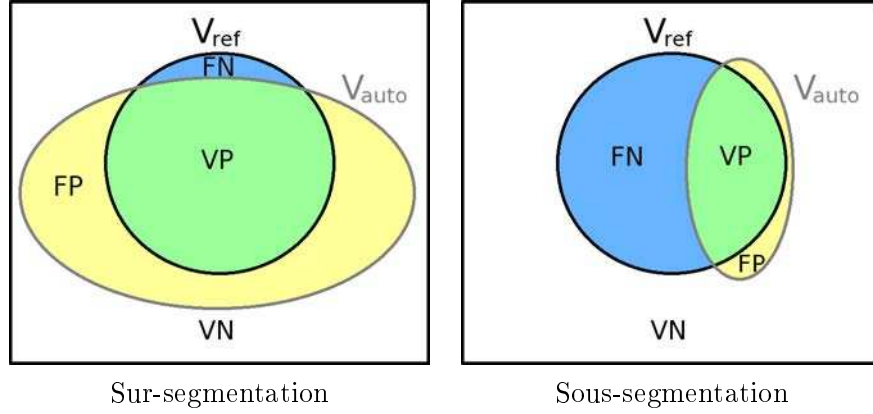


FIG. B.2 – Mise en évidence du lien entre spécificité, sensibilité et sur-, sous-segmentations. A gauche, cas de sur-segmentation (peu de FN et beaucoup de FP , sensibilité élevée, spécificité faible). A droite, cas de sous-segmentation (beaucoup de FN et peu de FP , sensibilité faible, spécificité élevée).

précisément, on associe à chaque point de contour de S_{ref} la distance au point de S_{auto} qui lui est le plus proche, et on note $h(S_{\text{ref}}, S_{\text{auto}})$ la distance la plus élevée que l'on a obtenu en parcourant ainsi tous les points de S_{ref} . Afin d'obtenir une mesure symétrique, on recommence le calcul en inter-changeant les rôles de S_{auto} et de S_{ref} , ce qui nous donne une deuxième valeur $h(S_{\text{auto}}, S_{\text{ref}})$. Ceci est illustré sur la figure B.3. La distance de Hausdorff est alors définie comme la distance la plus élevée parmi les deux valeurs ainsi obtenues :

$$\text{HD}(S_{\text{auto}}, S_{\text{ref}}) = \max(h(S_{\text{auto}}, S_{\text{ref}}), h(S_{\text{ref}}, S_{\text{auto}})) \quad (\text{B.6})$$

où :

$$h(S, S') = \max_{x \in S} \left(\min_{y \in S'} |x - y| \right) \quad (\text{B.7})$$

B.1.2.2 Distance de Hausdorff robuste

L'inconvénient majeur de la distance de Hausdorff est qu'elle est très sensible au bruit, puisqu'elle cherche à estimer le pire des cas (c'est-à-dire la distance point à point la plus élevée entre les deux surfaces). Elle n'est donc pas forcément représentative de la qualité de la segmentation dans sa globalité. Par exemple, si les deux segmentations sont identiques sauf que l'une d'elle possède très localement une extension par rapport à l'autre, la distance de Hausdorff va directement dépendre de la longueur de cette extension, aussi locale soit-elle.

Une façon de rendre plus robuste le calcul de la distance de Hausdorff est d'exclure les distances les plus élevées lors de la recherche des distances maximales

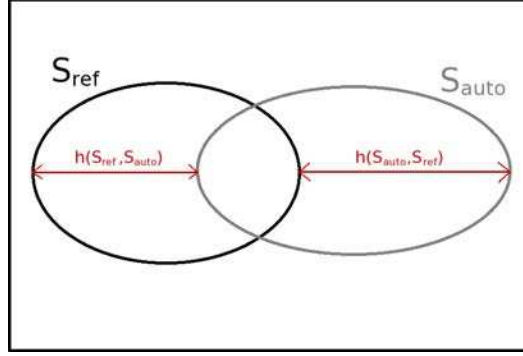


FIG. B.3 – Illustration du calcul de la distance de Hausdorff.

$h(S_{\text{auto}}, S_{\text{ref}})$ et $h(S_{\text{ref}}, S_{\text{auto}})$. En notant τ le pourcentage de valeurs gardées ($\tau = 1$ équivaut à la distance de Hausdorff classique), cela se formalise de la façon suivante :

$$\text{RHD}_{\tau}(S_{\text{auto}}, S_{\text{ref}}) = \max(h_{\tau}(S_{\text{auto}}, S_{\text{ref}}), h_{\tau}(S_{\text{ref}}, S_{\text{auto}})) \quad (\text{B.8})$$

où, en notant N_S le nombre de points du contour S , et $E(\cdot)$ la partie entière :

$$h_{\tau}(S, S') = K^{\text{ièmePlusPetiteValeur}} \left(\min_{x \in S, K=E(\tau \cdot N_S)} |x - y| \right) \quad (\text{B.9})$$

Pour $\tau = 1$, on obtient bien la formule de la distance de Hausdorff classique :

$$h_{\tau=1}(S, S') = K^{\text{ièmePlusPetiteValeur}} \left(\min_{x \in S, K=N_S} |x - y| \right) \quad (\text{B.10})$$

$$= \max_{x \in S} \left(\min_{y \in S'} |x - y| \right) \quad (\text{B.11})$$

$$= h(S, S') \quad (\text{B.12})$$

B.2 Mise en œuvre pour les structures de la région ORL

B.2.1 Particularités des structures considérées

Pour évaluer les différents algorithmes de segmentation automatique proposés dans cette thèse, nous avons adapté les mesures présentées dans la section précédente aux particularités des structures de la région ORL. En effet, parmi ces structures, certaines ont la particularité d'être séparées par un plan axial défini d'après des repères anatomiques précis. Comme illustré sur la figure 2.11, c'est typiquement le cas des aires ganglionnaires II, III et IV, pour lesquelles les plans axiaux faisant frontière sont par exemple définis selon la localisation de certains cartilages spécifiques.

Lors de l'évaluation des segmentations automatiques obtenues selon divers algorithmes, notre objectif est de nous affranchir le plus possible de la position de ces frontières axiales. En effet, elles correspondent typiquement à des détails très fins

de l'anatomie du cou que l'on peut difficilement espérer prendre en compte de façon automatique, même avec le recalage non-linéaire. Cela est d'autant plus vrai que ces repères anatomiques peuvent être très variables d'un individu à l'autre (pour les aires IV par exemple). Nous ne voulons donc pas pénaliser une segmentation automatique parce que ses limites axiales supérieure et/ou inférieure ne correspondent pas à celles de la segmentation de référence. Ce qui nous intéresse est de comparer les segmentations automatique et manuelle sur leurs plans communs.

Si toutes les aires ganglionnaires II, III et IV étaient systématiquement contournées dans nos segmentations manuelles, alors une solution simple serait de considérer l'union des aires II, III et IV lors de la comparaison des segmentations automatique et manuelle. Cependant, comme notre base de segmentations manuelles comporte des structures manquantes, nous proposons une mise en œuvre alternative.

B.2.2 Mise en œuvre proposée

Pour chacune des aires ganglionnaires II, III, et IV, on va calculer les mesures de qualité entre :

- la segmentation manuelle de l'aire considérée,
- l'union des segmentations automatiques des aires II, III et IV restreinte aux plans « pertinents ».

Les plans « pertinents » sont définis de façon différente suivant que la structure considérée est l'aire II, III ou IV (cf figure B.4) :

- **pour l'aire II** : l'aire II est l'aire la plus craniale parmi II, III et IV. Nous souhaitons donc juste nous affranchir de sa frontière caudale. Nous définissons donc les plans « pertinents » comme l'ensemble des plans de la segmentation manuelle de II ainsi que l'ensemble des plans des segmentations automatiques II, III et IV auxquels on retire ceux situés en dessous de la segmentation manuelle de II.
- **pour l'aire IV** : l'aire IV est l'aire la plus caudale parmi II, III et IV. Nous souhaitons donc juste nous affranchir de sa frontière craniale. Nous définissons donc les plans « pertinents » comme l'ensemble des plans de la segmentation manuelle de IV ainsi que l'ensemble des plans des segmentations automatiques II, III et IV auxquels on retire ceux situés en dessus de la segmentation manuelle de IV.
- **pour l'aire III** : l'aire III est située entre II et IV. Nous souhaitons donc nous affranchir de ses frontières craniale et caudale. Nous définissons donc les plans « pertinents » comme l'ensemble des plans de la segmentation manuelle de III.

(ou l'expression « l'ensemble des plans d'une segmentation » désigne ici l'ensemble des plans pour lesquels au moins un voxel de cette segmentation est étiqueté comme appartenant à la structure considérée). Une fois les plans pertinents sélectionnés, nous calculons les mesures de qualité sur un masque dilaté de l'union des segmentations manuelle et automatique sur ces plans.

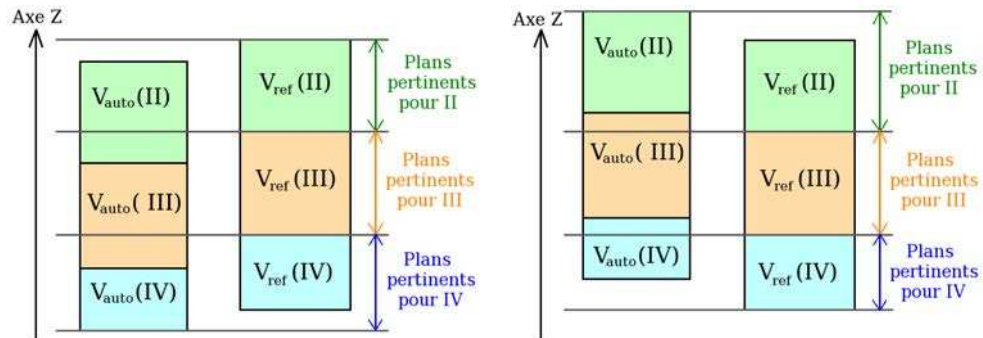


FIG. B.4 – Mise en évidence de la définition des plans « pertinents » sur lesquels nous nous restreignons pour le calcul des mesures de qualité entre la segmentation manuelle d'une aire donnée et l'union des segmentations automatiques des aires II, III et IV. Dans chacun des deux schémas, la segmentation automatique est à gauche, et la segmentation manuelle est à droite.

Nous appliquons également cette stratégie pour les segmentations de la moelle épinière et du tronc cérébral puisqu'ils sont aussi séparés par un plan axial. Pour toutes les autres structures, nous calculons les mesures de qualité de façon standard, c'est-à-dire sur le masque dilaté de l'union des segmentations automatique et manuelle.

Variabilité de contourage des structures ORL

Cette annexe décrit la variabilité de contourage manuel inter-expert pour les structures ORL. Pour trois images manuellement contourées de notre base, nous avons demandé à un second radiothérapeute de contourer manuellement les structures ORL. Afin d'éviter tout biais, ce radiothérapeute n'avait pas accès au premier contourage manuel. Pour ces trois patients, nous avons ensuite pu estimer le degré de variabilité inter-expert dans le contourage manuel, de façon quantitative et qualitative.

De façon quantitative, nous avons quantifié la variabilité de contourage par l'indice de DICE entre les deux contourages manuels disponibles pour chaque structure et pour chaque patient. Les résultats sont exposés sur la figure C.1. De façon qualitative, les figures C.2, C.3 et C.4 présentent les deux jeux de contours manuels pour différentes structures pour les trois patients.

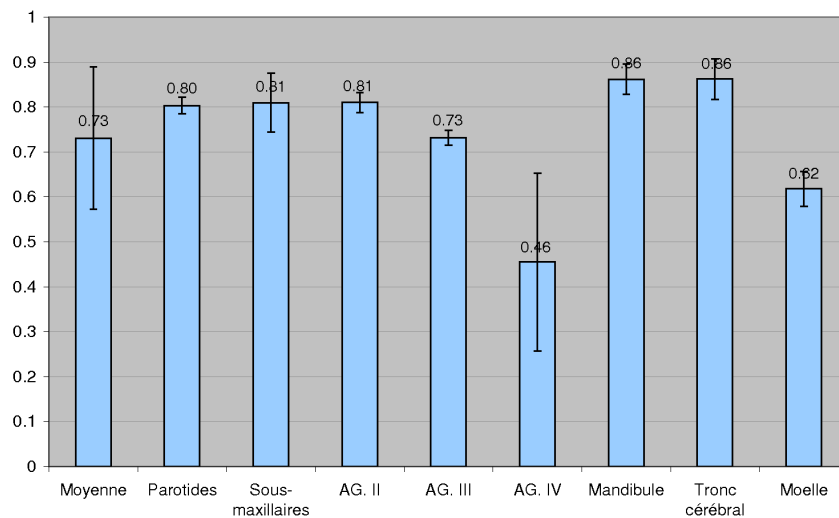


FIG. C.1 – Indice de DICE moyen calculé sur trois patients entre deux contourages manuels faits par deux experts différents. Les valeurs sont indiquées sur toutes les structures confondues (première barre), puis structure par structure.

Sur le graphique C.1, nous remarquons que les différences entre les deux contours manuels sont particulièrement importantes pour les aires ganglionnaires III et IV et pour la moelle épinière.

Pour la moelle épinière, ces différences ne représentent pas une variabilité de contourage réelle, elle provient simplement du fait que le second radiothérapeute a contouré le canal médullaire (c'est-à-dire l'enceinte osseuse) alors que le premier a contouré le cordon médullaire. Ceci est visible sur les figures C.2 et C.3.

Pour les aires ganglionnaires, les différences entre contours proviennent principalement de trois facteurs :

- Des paquets de graisse sont inclus dans l'aire ganglionnaire par l'un des expert et pas l'autre (cf. figures C.2 et C.3) ;
- Pour les aires ganglionnaires II et IV, un facteur de variabilité entre les deux jeux de contours est la limite supérieure (pour les aires II) ou inférieure (pour les aires IV), comme on peut le voir sur les coupes coronales de la figure C.4. Cet effet est plus important pour les aires IV ;
- La limite postérieure des aires ganglionnaires II, III et IV est déterminée d'après la limite postérieure du muscle sternocléidomastoïdien, qui est parfois difficile à visualiser.

Ces trois sources de différences ne caractérisent pas réellement une variabilité de contourage, mais davantage un protocole de segmentation. Cependant, malgré les consensus de segmentation proposés pour les aires ganglionnaires (notamment [Grégoire 2000]) les radiothérapeutes ont souvent des règles et des habitudes de contourage différents (en particulier d'un centre à l'autre, comme c'est le cas ici). Ainsi, bien que liées davantage à des définitions variables suivant les radiothérapeutes, les différences résultantes pour la segmentation des aires ganglionnaires représentent malgré tout un facteur important de variabilité de contourage.

Pour l'ensemble des autres structures (parotides, sous-maxillaires et tronc cérébral), l'indice de DICE moyen sur les trois patients est plus élevé que pour les structures précédentes et compris entre 0.8 et 0.9. Cela représente malgré tout une variabilité de contourage importante. La variabilité la plus faible est observée pour la mandibule et le tronc cérébral (DICE moyen de 0.86). Pour la mandibule, cette faible variabilité vient du fait qu'il s'agit d'une structure osseuse donc très contrastée. Pour les parotides, la variabilité de contourage se situe principalement au niveau des lobes accessoires et profonds, comme illustré sur la figure C.2. Un facteur particulièrement important pouvant induire des différences de segmentation pour les parotides est le fenêtrage d'intensités utilisé lors du contourage. Enfin, pour l'ensemble des structures, les différences de contourage peuvent aussi venir de la résolution de l'image (512×512 ou 256×256) ainsi que du degré d'interpolation utilisé. En effet, dans certains cas, la totalité des coupes ne sont pas contourées (une sur deux seulement par exemple), et une interpolation est alors utilisée pour obtenir une segmentation complète. Dans notre étude, la résolution de l'image était la même pour les deux radiothérapeutes, et la totalité des coupes a été contourée manuellement (aucune interpolation utilisée).

Dans notre base, les 105 images ont été contourées par des radiothérapeutes potentiellement différents, mais elles ont toutes été validées (et corrigées si besoin) par un même radiothérapeute. Cette phase de validation et de correction permet de limiter la variabilité inter-expert, mais elle ne permet pas de s'en affranchir totalement.

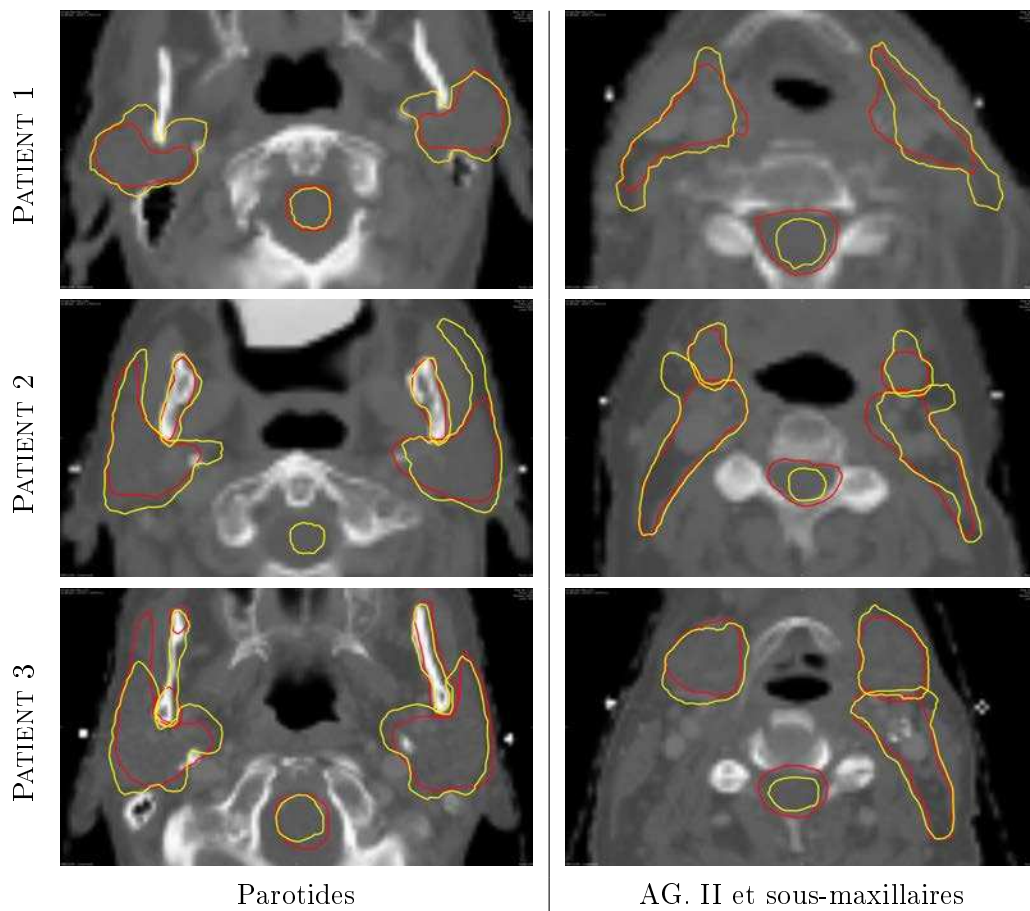


FIG. C.2 – Illustration de la variabilité de contourage inter-expert sur les parotides, les sous-maxillaires et les aires ganglionnaires II. Les contours jaunes représentent les contours manuels du premier expert, et les contours rouges ceux du deuxième expert.

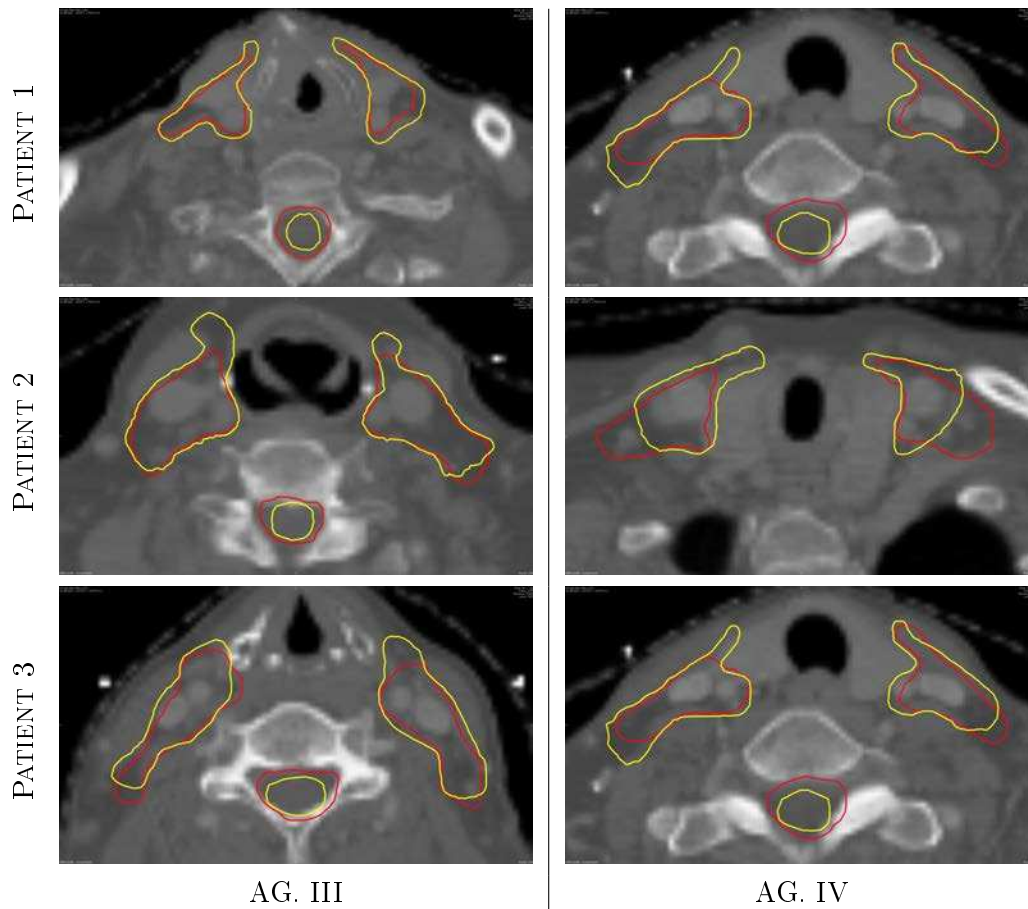


FIG. C.3 – Illustration de la variabilité de contourage inter-expert sur les aires ganglionnaires III et IV. Les contours jaunes représentent les contours manuels du premier expert, et les contours rouges ceux du deuxième expert.

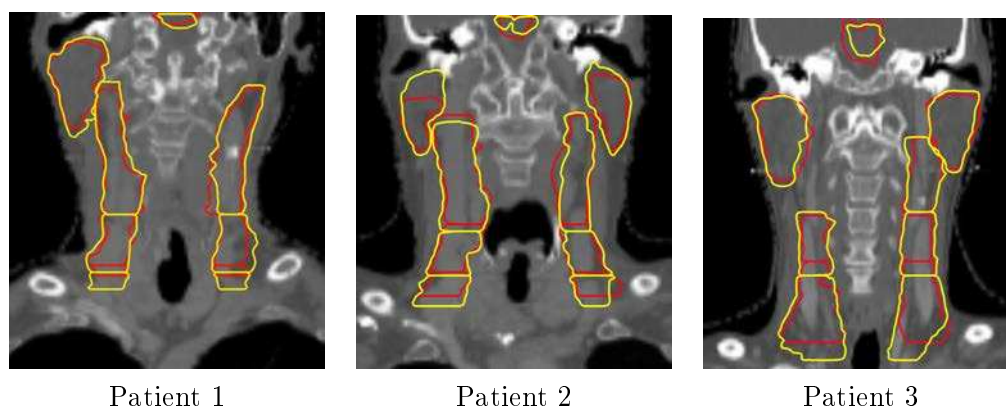


FIG. C.4 – Illustration de la variabilité de contourage inter-expert sur les aires ganglionnaires II, III et IV en coupes coronales. Les contours jaunes représentent les contours manuels du premier expert, et les contours rouges ceux du deuxième expert.

De plus, à la variabilité inter-expert résiduelle viennent s'ajouter une variabilité intra-expert ainsi qu'une certaine influence du diagnostic sur le contourage.

La variabilité de contourage représente une difficulté majeure pour nos outils de segmentation automatique, et ce à deux niveaux. D'une part elle pose problème dans le processus de segmentation automatique en lui-même. Par exemple, dans les algorithmes de segmentation multi-atlas adaptés au patient présentés dans les chapitres 5 et 6, un atlas sélectionné comme étant très similaire en intensité au patient peut biaiser la segmentation finale du patient si son contourage est sujet à une variabilité importante par rapport à la majorité des autres atlas sélectionnés. D'autre part, la variabilité inter-expert de contourage nous pose également problème lors de l'évaluation des algorithmes. En effet, notre seul moyen d'évaluation est basé sur la comparaison des contours automatiques aux contours manuels, puisqu'une évaluation clinique par des radiothérapeutes n'est pas possible à grande échelle lors de la phase de définition et de test des algorithmes. Concernant ces problèmes d'évaluation, les chiffres présentés sur le graphique C.1 représentent donc en quelque sorte la borne supérieure que l'on peut espérer atteindre avec nos méthodes de segmentation automatique (DICE moyen de l'ordre de 0.8, mais plus faible pour les aires ganglionnaires III et IV (environ 0.7 et 0.5 respectivement)).

Bibliographie

- [Akhondi-Asl 2010] A. Akhondi-Asl, K. Jafari-Khouzani et H. Soltanian-Zadeh. *Multiple-atlas-based automatic segmentation of hippocampus for lateralization in temporal lobe epilepsy*. In Proceedings of the 7th IEEE International Symposium on Biomedical Imaging (ISBI'10), pages 836–839. IEEE Press, 2010. (Cité page 179.)
- [Aljabar 2007] P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal et D. Rueckert. *Classifier selection strategies for label fusion using large atlas databases*. In Proceedings of the 10th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'07), Part I, volume 4791 of *LNCS*, pages 523–531. Springer Verlag, 2007. (Cité pages 55, 105, 110, 125 et 126.)
- [Aljabar 2009] P. Aljabar, R.A. Heckemann, A. Hammers, J.V. Hajnal et D. Rueckert. *Multi-atlas based segmentation of brain images : atlas selection and its effect on accuracy*. *Neuroimage*, vol. 46, no. 3, pages 726–738, 2009. (Cité pages 55, 105, 106, 110, 112, 124 et 126.)
- [Arsigny 2006a] V. Arsigny, O. Commowick, X. Pennec et N. Ayache. *A Log-Euclidean Framework for Statistics on Diffeomorphisms*. In Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06), Part I, volume 4190 of *LNCS*, pages 924–931. Springer Verlag, October 2006. (Cité page 120.)
- [Arsigny 2006b] Vincent Arsigny, Olivier Commowick, Xavier Pennec et Nicholas Ayache. *A Log-Euclidean Framework for Statistics on Diffeomorphisms*. In Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06), Part I, volume 4190 of *LNCS*, pages 924–931. Springer Verlag, Octobre 2006. (Cité page 121.)
- [Artaechevarria 2009] X. Artaechevarria, A. Munoz-Barrutia et C. Ortiz de Solorzano. *Combination strategies in multi-atlas image segmentation : application to brain MR data*. *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pages 1266–1277, August 2009. (Cité pages 55, 106, 111, 124, 126 et 127.)
- [Atif 2007] J. Atif, C. Hudelot, O. Nempont, N. Richard, B. Batrancourt, E. Angelini et I. Block. *GRAFIP : A framework for the representation of healthy and pathological cerebral information*. In Proceedings of the 4th IEEE International Symposium on Biomedical Imaging (ISBI'07), pages 205–208, april 2007. (Cité page 219.)
- [Bach Cuadra 2004] M. Bach Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J. Villemure et J.-P. Thiran. *Atlas-based Segmentation of Pathological MR Brain Images using a Model of Lesion Growth*. *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pages 1301–1314, Octobre 2004. (Cité page 42.)

- [Barnes 2008] J. Barnes, J. Foster, R.G. Boyes, T. Pepple, E.K. Moore, J.M. Schott, C. Frost, R.I. Scahill et N.C. Fox. *A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus*. NeuroImage, vol. 40, no. 4, pages 1655–1671, 2008. (Cité page 56.)
- [Benoudiba 2007] F. Benoudiba, Y. Menu, J-B. Charrier et B. Lallemant. *Guide de lecture d'imagerie médicale : Classification des niveaux ganglionnaires cervicaux*, 2007. Laboratoires Sanofi Aventis. (Cité page 24.)
- [Besl 1992] P.J. Besl et N.D. McKay. *A Method for Registration of 3D Shapes*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pages 239–256, 1992. (Cité page 44.)
- [Betrouni 2009] N. Betrouni. *Le recalage en imagerie médicale : de la conception à la validation*. IRBM, vol. 30, no. 2, pages 60–71, 2009. (Cité page 41.)
- [Bezdek 1981] James C. Bezdek. Pattern recognition with fuzzy objectives function algorithms. Plenum Press Ed., 1981. (Cité page 35.)
- [Bhatia 2004] Kanwal K. Bhatia, Joseph V. Hajnal, Basant K. Puri, A. D. Edwards et Daniel Rueckert. *Consistent Groupwise Non-Rigid Registration for Atlas Construction*. In Proceedings of the 1st IEEE International Symposium on Biomedical Imaging (ISBI'04), volume 1, pages 908–911. IEEE Press, 2004. (Cité page 52.)
- [Bhatia 2007] K.K. Bhatia, P. Aljabar, J.P. Boardman, L. Srinivasan, M. Murgasova, S.J. Counsell, M.A. Rutherford, J. Hajnal, A.D. Edwards et D. Rueckert. *Groupwise combined segmentation and registration for atlas construction*. In Proceedings of the 10th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'07), Part I, volume 4791 of LNCS, pages 532–540. Springer Verlag, 2007. (Cité page 51.)
- [Blezek 2007] D.-J. Blezek et J.-V. Miller. *Atlas stratification*. Medical Image Analysis, vol. 11, no. 5, pages 443–457, 2007. (Cité pages 54, 117, 118 et 215.)
- [Bloch 2005] I. Bloch, O. Colliot, O. Camara et T. Géraud. *Fusion of spatial relationships for guiding recognition, example of brain structure recognition in 3D MRI*. Pattern Recognition Letters, vol. 26, no. 4, pages 449–457, 2005. ICAPR 2003. (Cité pages 37 et 219.)
- [Boisserie 2001] G. Boisserie et D. Hasboun. *Utilisation de l'imagerie multimodalité en radiothérapie*. Cancer/Radiothérapie, vol. 5, no. Supplément 1, pages 15s–35s, 2001. (Cité page 41.)
- [Bondiau 2005] Pierre-Yves Bondiau, Grégoire Malandain, Stéphane Chanalet, Pierre-Yves Marcy, Jean-Louis Habrand, Francois Fauchon, Philippe Paquis, Adel Courdi, Olivier Commowick, Isabelle Rutten et Nicholas Ayache. *Atlas-based automatic segmentation of MR images : validation study on the brainstem in radiotherapy context*. Int J Radiat Oncol Biol Phys, vol. 61, no. 1, pages 289–298, Janvier 2005. (Cité pages 39 et 50.)

- [Bonniaud 2006] G. Bonniaud, A. Isambert, F. Dhermain, A. Beaudré, I. Ferreira, M. Ricard et D. Lefkopoulos. *Recalage d'images en radiothérapie : considérations pratiques et contrôle de qualité*. Cancer/Radiothérapie, vol. 10, no. 5, pages 222–230, 2006. (Cité page 41.)
- [Borgefors 1988] G. Borgefors. *Hierarchical Chamfer Matching : A Parametric Edge Matching Algorithm*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 6, pages 849–865, Novembre 1988. (Cité page 43.)
- [Brown 1992] L.G. Brown. *A Survey of Image Registration Techniques*. ACM Computing Surveys, vol. 24, no. 4, pages 325–376, Décembre 1992. (Cité page 41.)
- [Cachier 2001] P. Cachier, J.-F. Mangin, X. Pennec, D. Rivière, D. Papadopoulos-Orfanos, J. Régis et N. Ayache. *Multisubject Non-Rigid Registration of Brain MRI using Intensity and Geometric Features*. In Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'01), volume 2208 of LNCS, pages 734–742. Springer Verlag, 2001. (Cité page 46.)
- [Castadot 2008] Pierre Castadot, John Aldo Lee, Adriane Parraga, Xavier Geets, Benoît Macq et Vincent Grégoire. *Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors*. Radiotherapy Oncology, vol. 89, no. 1, pages 1–12, 2008. (Cité page 41.)
- [Castadot 2010] P. Castadot, J.-A. Lee, X. Geets et V. Grégoire. *Adaptive Radiotherapy of Head and Neck Cancer*. Seminars in Radiation Oncology, vol. 20, no. 2, pages 84–93, 2010. Adaptive Radiotherapy. (Cité page 41.)
- [Chao 2007] K.S.C. Chao, S. Bhide, H. Chen, J. Asper, S. Bush, G. Franklin, V. Kavadi, V. Liengswangwong, W. Gordon, A. Raben, J. Strasser, C. Koprowski, S. Frank, G. Chronowski, A. Ahamad, R. Malyapa, L. Zhang et L. Dong. *Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach*. Int J Radiat Oncol Biol Phys, vol. 68, no. 5, pages 1512–1521, August 2007. (Cité pages 30, 32 et 33.)
- [Chupin 2009] M. Chupin, A. Hammers, R.S.N. Liu, O. Colliot, J. Burdett, E. Bardinnet, J.S. Duncan, L. Garnero et L. Lemieux. *Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints : method and validation*. Neuroimage, vol. 46, no. 3, pages 749–761, July 2009. (Cité page 34.)
- [Ciofalo 2005] C. Ciofalo et C. Barillot. *Brain segmentation with competitive level sets and fuzzy control*. In Proceedings of the 19th International Conference on Information Processing in Medical Imaging (IPMI'05), volume 3565 of LNCS, pages 333–344. Springer Verlag, 2005. (Cité page 34.)
- [Commowick 2007a] O. Commowick et G. Malandain. *Efficient Selection of the Most Similar Image in a Database for Critical Structures Segmentation*. In

- Proceedings of 10th the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'07), Part II, volume 4792 of *LNCS*, pages 203–210. Springer Verlag, 2007. (Cité pages 55, 56, 107, 108, 109, 110, 112, 119, 138 et 210.)
- [Commowick 2007b] Olivier Commowick. *Création et Utilisation d'Atlas Anatomiques Numériques pour la Radiothérapie (Design and Use of Anatomical Atlases for Radiotherapy)*. Thèse de sciences (phd thesis), Université de Nice – Sophia-Antipolis, February 2007. (Cité pages 39, 49, 65, 66, 68 et 223.)
- [Commowick 2008a] O. Commowick, V. Arsigny, A. Isambert, J. Costa, F. Dhermain, F. Bidault, P.-Y. Bondiau, N. Ayache et G. Malandain. *An Efficient Locally Affine Framework for the Smooth Registration of Anatomical Structures*. *Medical Image Analysis*, vol. 12, no. 4, pages 427–441, 2008. (Cité page 48.)
- [Commowick 2008b] O. Commowick, V. Grégoire et G. Malandain. *Atlas-Based Delineation of Lymph Node Levels in Head and Neck Computed Tomography Images*. *Radiotherapy Oncology*, vol. 87, no. 2, pages 281–289, 2008. (Cité page 66.)
- [Commowick 2009] O. Commowick, S.-K. Warfield et G. Malandain. *Using Frankenstein's Creature Paradigm to Build a Patient Specific Atlas*. In Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'09), Part II, volume 5762 of *LNCS*, pages 993–1000. Springer Verlag, 2009. (Cité pages 55, 57, 110, 119, 120, 121 et 122.)
- [Commowick 2010] O. Commowick et S.K. Warfield. *Incorporating Priors on Expert Performance Parameters for Segmentation Validation and Label Fusion : A Maximum a Posteriori STAPLE*. In Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'10), volume 6363 of *LNCS*, pages 25–32. Springer Verlag, 2010. (Cité pages 66 et 94.)
- [Cootes 1994] T.F. Cootes, A. Hill, C.J. Taylor et J. Haslam. *Use of Active Shape Models for Locating Structures in Medical Images*. *Image Vision Computing*, vol. 12, no. 6, pages 355–366, 1994. (Cité page 64.)
- [Cootes 1995] T.F. Cootes, C.J. Taylor, D.H. Cooper et J. Graham. *Active shape models - their training and application*. *Computer Vision and Image Understanding*, vol. 61, no. 1, pages 38–59, 1995. (Cité page 36.)
- [Cootes 2001] T.F. Cootes, G.J. Edwards et C.J. Taylor. *Active Appearance Models*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pages 681–685, 2001. (Cité page 36.)
- [Costa 2007] J. Costa, H. Delingette, S. Novellas et N. Ayache. *Automatic Segmentation of Bladder and Prostate Using Coupled 3D Deformable Models*. In Proceedings of the 10th International Conference on Medical Image Com-

- puting and Computer-Assisted Intervention (MICCAI'07), volume 4791 of *LNCS*, pages 252–260. Springer Verlag, October 2007. (Cité page 38.)
- [Craene 2005] M. De Craene, A. du Bois d'Aische, B. Macq et S.K. Warfield. *Multi-subject variational registration for probabilistic unbiased atlas generation*. volume 3, pages 601–604, sep. 2005. (Cité page 51.)
- [D'Agostino 2004] Emiliano D'Agostino, Frederik Maes, Dirk Vandermeulen et Paul Suetens. *Non-rigid Atlas-to-Image Registration by Minimization of Class-Conditional Image Entropy*. In Proceedings of the 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'04), volume 3216 of *LNCS*, pages 745–753. Springer Verlag, 2004. (Cité page 42.)
- [Dawant 1999a] B. M. Dawant, S. L. Hartmann et S. Gadamsetty. *Brain atlas deformation in the presence of large space-occupying tumors*. In Proceedings of the 2nd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'99), volume 1679 of *LNCS*, pages 589–596. Springer Verlag, 1999. (Cité page 42.)
- [Dawant 1999b] B. M. Dawant, S. L. Hartmann, J-P. Thirion, F. Maes, D. Vandermeulen et P. Demaerel. *Automatic 3-D Segmentation of Internal Structures of the Head in MR Images Using a Combination of Similarity and Free-Form Transformations : Part I, Methodology and Validation on Normal Subjects*. IEEE Transactions on Medical Imaging, vol. 18, no. 10, pages 909–916, 1999. (Cité pages 39 et 50.)
- [De Craene 2004] M. De Craene, A. du Bois d'Aische, B. Macq et S.K. Warfield. *Multi-Subject Registration for Unbiased Statistical Atlas Construction*. In Proceedings of the 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'04), volume 3216 of *LNCS*, pages 655–662. Springer Verlag, Septembre 2004. (Cité page 51.)
- [de Graeff 2001] A. de Graeff, J. R. J. de Leeuw, W. J. G. Ros, G. J. Hordijk, G. H. Blijham et J. A. M. Winnubst. *Sociodemographic factors and quality of life as prognostic indicators in head and neck cancer*. European Journal of Cancer, vol. 37, no. 3, pages 332–339, 2001. (Cité page 8.)
- [Dempster 1977] A. Dempster, N. Laird et D. Rubin. *Maximum Likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society, vol. 39 (Series B), 1977. (Cité pages 35 et 66.)
- [Dice 1945] L.R. Dice. *Measures of the amount of ecologic association between species*. Ecology, vol. 26, no. 3, pages 297–302, 1945. (Cité page 226.)
- [Emami 2003] B. Emami, A. Sethi et G.-J Petruzelli. *Influence of MRI on target volume delineation and IMRT planning in nasopharyngeal carcinoma*. Int J Radiat Oncol Biol Phys, vol. 57, no. 2, pages 481–488, 2003. (Cité pages 20 et 41.)

- [Epstein 1987] J.-B. Epstein, F.-L. Wong et P. Stevenson-Moore. *Osteoradionecrosis : clinical experience and a proposal for classification*. J Oral Maxillofac Surg, vol. 45, no. 2, pages 104–110, February 1987. (Cit  page 182.)
- [Epstein 1997] J. Epstein, E. van der Meij, M. McKenzie, F. Wong, M. Lepawsky et P. Stevenson-Moore. *Postradiation osteonecrosis of the mandible : a long-term follow-up study*. Oral Surg Oral Med Oral Pathol Oral Radiol Endod, vol. 83, no. 6, pages 657–662, June 1997. (Cit  page 182.)
- [Ericsson 2008] A. Ericsson, P. Aljabar et D. Rueckert. *Construction of a patient-specific atlas of the brain : Application to normal aging*. In Proceedings of the 5th IEEE International Symposium on Biomedical Imaging (ISBI'08), pages 480–483. IEEE Press, 2008. (Cit  page 104.)
- [Ferlay 2010] J. Ferlay, H.R. Shin, F. Bray, D. Forman, C. Mathers et D.M. Parkin. *GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide : IARC CancerBase No. 10 [Internet]*, 2010. (Cit  pages 5 et 7.)
- [Fischl 2002] B. Fischl, D.-H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen et A.-M. Dale. *Whole Brain Segmentation : Automated Labeling of Neuroanatomical Structures in the Human Brain*. Neuron, vol. 33, no. 3, pages 341–355, 2002. (Cit  page 219.)
- [Fonov 2011] V. Fonov, A.C. Evans, K. Botteron, C.R. Almli, R.C. McKinstry et D.L. Collins. *Unbiased average age-appropriate atlases for pediatric studies*. Neuroimage, vol. 54, no. 1, pages 313–327, January 2011. (Cit  page 118.)
- [Frangi 2001] A.F. Frangi, D. Rueckert, J.A. Schnabel et W.J. Niessen. *Automatic 3D ASM Construction via Atlas-Based Landmarking and Volumetric Elastic Registration*. In Proceedings of the 17th International Conference on Information Processing in Medical Imaging (IPMI'01), volume 2082 of LNCS, pages 78–91. Springer Verlag, 2001. (Cit  page 36.)
- [Frey 2007] B.J. Frey et D. Dueck. *Clustering by passing messages between data points*. Science, vol. 315, no. 5814, pages 972–976, February 2007. (Cit  pages 54, 115, 118, 139 et 216.)
- [Gao 2008] H. Gao et O. Chae. *Automatic Tooth Region Separation for Dental CT Images*. In Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology (ICCIT'08), volume 1, pages 897–901, 2008. (Cit  page 185.)
- [Gao 2010] Hui Gao et Oksam Chae. *Individual tooth segmentation from CT images using level set method with shape and intensity prior*. Pattern Recognition, vol. 43, no. 7, pages 2406–2417, 2010. (Cit  page 186.)
- [Garcia 2010] V. Garcia, O. Commowick et G. Malandain. *A Robust and Efficient Block Matching Framework for Non Linear Registration of Thoracic CT Images*. In Proceedings of the Workshop Evaluation of Methods for Pulmonary Image REgistration (EMPIRE'10), September 2010. Satellite Workshop of MICCAI'10. (Cit  pages 49 et 223.)

- [Gee 1993] J.-C. Gee, M. Reivich et R. Bajcsy. *Elastically deforming 3D atlas to match anatomical brain images*. J Comput Assist Tomogr, vol. 17, no. 2, pages 225–236, Mar-Apr 1993. (Cit e page 50.)
- [Gorthi 2009a] S. Gorthi, V. Duay, M. Bach Cuadra, P.-A. Tercier, A.S. Allal et J.-P. Thiran. *Active contour-based segmentation of head and neck with adaptive atlas selection*. In Proceedings of Head And Neck Auto-Segmentation Challenge Workshop, September 2009. Satellite Workshop of MICCAI'09. (Cit e pages 56, 105 et 106.)
- [Gorthi 2009b] S. Gorthi, V. Duay, N. Houhou, M. Bach Cuadra, U. Schick, M. Becker, A.S. Allal et J.-P. Thiran. *Segmentation of Head and Neck Lymph Node Regions for Radiotherapy Planning Using Active Contour-Based Atlas Registration*. IEEE Journal of Selected Topics in Signal Processing, vol. 3, no. 1, pages 135–147, feb. 2009. (Cit e pages 218 et 219.)
- [Gorthi 2010] S. Gorthi, M. Bach Cuadra, U. Schick, P.-A. Tercier, A.S. Allal et J.-P. Thiran. *Multi-atlas based segmentation of head and neck CT images using active contour framework*. In Proceedings of Head And Neck Auto-Segmentation Challenge Workshop, pages 313–321, September 2010. Satellite Workshop of MICCAI'10. (Cit e pages 105, 106, 124 et 126.)
- [Grabner 2006] G unther Grabner, Andrew L. Janke, Marc M. Budge, David Smith, Jens Pruessner et D. Louis Collins. *Symmetric Atlasing and Model Based Segmentation : An Application to the Hippocampus in Older Adults*. In Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06), Part II, volume 4191 of LNCS, pages 58–66. Springer Verlag, Octobre 2006. (Cit e page 52.)
- [Gr egoire 2000] V. Gr egoire, E. Coche, G. Cosnard, M. Hamoir et H. Reyhler. *Selection and delineation of lymph node target volumes in head and neck conformal radiotherapy. Proposal for standardizing terminology and procedure based on the surgical experience*. Radiotherapy Oncology, vol. 56, no. 2, pages 135–150, 2000. (Cit e pages 23 et 234.)
- [Gr egoire 2003] V. Gr egoire, P. Levendag, K. K Ang, J. Bernier, M. Braaksma, V. Budach, C. Chao, E. Coche, J. S Cooper, G. Cosnard, A. Eisbruch, S. El-Sayed, B. Emami, C. Grau, M. Hamoir, N. Lee, P. Maingon, K. Muller et H. Reyhler. *CT-based delineation of lymph node levels and related CTVs in the node-negative neck : DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines*. Radiotherapy Oncology, vol. 69, no. 3, pages 227–236, D ecembre 2003. (Cit e page 25.)
- [Guimond 2000] A. Guimond, J. Meunier et J.-P. Thirion. *Average Brain Models : A Convergence Study*. Computer Vision and Image Understanding, vol. 77, no. 2, pages 192–210, 2000. (Cit e pages 52, 62, 68, 99, 119, 120, 121, 123, 133, 188, 191 et 211.)
- [Han 2008] X. Han, M.-S. Hoogeman, P.-C. Levendag, L.-S. Hibbard, D.-N. Teguh, P. Voet, A.-C. Cowen et T.-K. Wolf. *Atlas-based auto-segmentation of head*

- and neck CT images*. In Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'08), Part II, volume 5242 of *LNCS*, pages 434–441. Springer Verlag, 2008. (Cité pages 39, 55, 105, 110 et 126.)
- [Han 2009] X. Han, L.S. Hibbard, N. O'Connell et V. Willcut. *Automatic Segmentation of Head and Neck CT Images by GPU-Accelerated Multi-atlas Fusion*. In Proceedings of Head And Neck Auto-Segmentation Challenge Workshop, 2009. Satellite Workshop of MICCAI'09. (Cité page 219.)
- [Han 2010] X. Han, L.S. Hibbard, N.P. O'Connell et V. Willcut. *Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion*. In Proceedings of Head And Neck Auto-Segmentation Challenge Workshop, pages 297–304, September 2010. Satellite Workshop of MICCAI'10. (Cité pages 34, 39, 124, 126, 179 et 218.)
- [Heckemann 2006] R.-A. Heckemann, J.-V. Hajnal, P. Aljabar, D. Rueckert et A. Hammers. *Automatic anatomical brain MRI segmentation combining label propagation and decision fusion*. *NeuroImage*, vol. 33, no. 1, pages 115–126, 2006. (Cité pages 55, 58, 123 et 126.)
- [Heckemann 2009] R.A. Heckemann, A. Hammers, P. Aljabar, D. Rueckert et J.V. Hajnal. *The mirror method of assessing segmentation quality in atlas label propagation*. In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI'09), pages 1194–1197. IEEE Press, 2009. (Cité pages 107, 110, 112, 124 et 126.)
- [Hollensen 2010] C. Hollensen, M.F. Hansen, L. Hojgaard, L. Specht et R. Larsen. *Segmenting the parotid gland using registration and level set methods*. In Proceedings of Head And Neck Auto-Segmentation Challenge Workshop, pages 305–312, September 2010. Satellite Workshop of MICCAI'10. (Cité pages 34, 39 et 218.)
- [Hosntalab 2010] M. Hosntalab, R. Aghaeizadeh Zoroofi, A. Abbaspour Tehrani-Fard et S. Gholamreza. *Classification and numbering of teeth in multi-slice CT images using wavelet-Fourier descriptor*. *Int J Comput Assist Radiol Surg*, vol. 5, no. 3, pages 237–249, May 2010. (Cité pages 185 et 186.)
- [Houhou 2005] N. Houhou, V. Duay, A. Allal et J.-P. Thiran. *Medical images registration with a hierarchical atlas*. In Proceedings of the 13th European Signal Processing Conference (EUSIPCO'05), Parallel Computing in Electrical Engineering. IEEE, 2005. (Cité pages 42 et 45.)
- [Hsiung 2006] C.-Y Hsiung, H.-M Ting, H.-Y Huang, C.-H Lee, E.-Y Huang et H.-C Hsu. *Parotid-sparing intensity-modulated radiotherapy (IMRT) for nasopharyngeal carcinoma : Preserved parotid function after IMRT on quantitative salivary scintigraphy, and comparison with historical data after conventional radiotherapy*. *Int J Radiat Oncol Biol Phys*, vol. 66, pages 454–461, 2006. (Cité page 20.)

- [Hu 2008] K. Hu, A. Lin, A. Young, G. Kubicek, J.W. Piper, A.S. Nelson, J. Dolan, R. Masino et M. Machtay. *Timesavings for contour generation in head and neck IMRT : multi-institutional experience with an atlas-based segmentation method*. In Acts of the 50th Annual Meeting of the American Society for Therapeutic Radiology and Oncology (ASTRO'08), volume 72, page S391, 2008. (Cité pages 30, 32 et 33.)
- [Huttenlocher 1993] D. Huttenlocher, D. Klanderman et A. Rucklidge. *Comparing images using the Hausdorff distance*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 9, pages 850–863, Septembre 1993. (Cité page 227.)
- [Isambert 2008] A. Isambert, V. Grégoire, F. Bidault, M. Benkebil, L. Ramus, G. Malandain, J. Bourhis et D. Lefkopoulos. *Evaluation d'un algorithme de segmentation automatique des structures d'intérêt dans la sphère ORL*. In Acts of the 19ème Congrès Annuel de la Société Française de Radiothérapie Oncologique (SFRO'08), 2008. Cancer/Radiothérapie 12(6-7) :743, 2008. (Cité page 62.)
- [Isgum 2009] I. Isgum, M. Staring, A. Rutten, M. Prokop, M.-A. Viergever et B. van Ginneken. *Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans*. IEEE Transactions on Medical Imaging, vol. 28, no. 7, pages 1000–1010, July 2009. (Cité pages 55, 111, 124 et 126.)
- [Johnson 2008] S. Johnson, J.-T. McDonald et M.-J. Corsten. *Socioeconomic factors in head and neck cancer*. J Otolaryngol Head Neck Surg, vol. 37, no. 4, pages 597–601, August 2008. (Cité page 6.)
- [Joshi 2004] S Joshi, Brad Davis, Matthieu Jomier et Guido Gerig. *Unbiased diffeomorphic atlas construction for computational anatomy*. Neuroimage, vol. 23 Suppl 1, 2004. (Cité page 53.)
- [Kam 2003] Michael K. M. Kam, Ricky M. C. Chau, Joyce Suen, Peter H. K. Choi et Peter M. L. Teo. *Intensity-modulated radiotherapy in nasopharyngeal carcinoma : dosimetric advantage over conventional plans and feasibility of dose escalation*. Int J Radiat Oncol Biol Phys, vol. 56, no. 1, pages 145–157, 2003. (Cité page 20.)
- [Kass 1987] M. Kass, A. Witkin et D. Terzopoulos. *Snakes : Active Contour Models*. International Journal of Computer Vision, vol. 1, no. 4, pages 321–331, 1987. (Cité page 34.)
- [Keyhaninejad 2006] S. Keyhaninejad, R.-A. Zoroofi, S.-K. Setarehdan et G. Shirani. *Automated segmentation of teeth in multi-slice CT images*. IET Conference Publications, vol. 2006, no. CP522, pages 339–344, 2006. (Cité pages 185 et 186.)
- [Khotanlou 2007] H. Khotanlou, J. Atif, E. Angelini, H. Duffau et I. Bloch. *Adaptive segmentation of internal brain structures in pathological MR images depending on tumor types*. In Proceedings of the 4th IEEE International Sym-

- posium on Biomedical Imaging (ISBI'07), pages 588–591, april 2007. (Cité page 219.)
- [Kittler 1998] J. Kittler, M. Hatef, R.P.W. Duin et J. Matas. *On Combining Classifiers*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pages 226–239, 1998. (Cité page 64.)
- [Klein 2008] S. Klein, U.-A. van der Heide, I.-M. Lips, M. van Vulpen, M. Staring et J.-P.-W. Pluim. *Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information*. Medical Physics, vol. 35, no. 4, pages 1407–1417, 2008. (Cité pages 55, 105, 106, 109, 110, 112 et 126.)
- [Klein 2010] S. Klein, M. Loog, F. van der Lijn, T. den Heijer, A. Hammers, M. de Bruijne, A. van der Lugt, R.P.W. Duin, M.M.B. Breteler et W.J. Niessen. *Early diagnosis of dementia based on intersubject whole-brain dissimilarities*. In Proceedings of the 7th IEEE International Symposium on Biomedical Imaging (ISBI'10), pages 249–252. IEEE Press, 2010. (Cité page 109.)
- [Kobashi 1995] M. Kobashi et L.G. Shapiro. *Knowledge-based organ identification from CT images*. Pattern Recognition, vol. 28, no. 4, pages 475–491, 1995. (Cité page 37.)
- [Kuncheva 2007] L.-I. Kuncheva. *Combining Pattern Classifiers : Methods and Algorithms*. Neural Networks, IEEE Transactions on, vol. 18, no. 3, page 964, Mai 2007. (Cité page 64.)
- [Kyriacou 1999] S. K. Kyriacou, C. Davatzikos, S. J. Zinreich et R. N. Bryan. *Non-linear elastic registration of brain images with tumor pathology using a biomechanical model*. IEEE Transactions on Medical Imaging, vol. 18, no. 7, pages 580–592, July 1999. (Cité page 42.)
- [Lamecker 2002] H. Lamecker, T. Lange et M. Seebass. *A Statistical Shape Model for the Liver*. In Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'02), Part II, volume 2489 of LNCS, pages 421–427. Springer Verlag, 2002. (Cité page 64.)
- [Landman 2010] B.A. Landman, J.A. Bogovic et J.L. Prince. *Simultaneous Truth and Performance Level Estimation with Incomplete, Over-complete, and Ancillary Data*. Proc Soc Photo Opt Instrum Eng, vol. 7623, page 76231N, March 2010. (Cité page 65.)
- [Langerak 2009] T.-R. Langerak, U.-A. van der Heide, I.-M. Lips, A.N.T.J. Kotte, M. van Vulpen et J.P.W. Pluim. *Label fusion using performance estimation with iterative label selection*. In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI'09), pages 1186–1189. IEEE Press, jun. 2009. (Cité pages 65, 107, 110, 112 et 113.)
- [Langerak 2010] T. Langerak, U. van der Heide, A. Kotte, M. Viergever, M. van Vulpen et J. Pluim. *Label Fusion in Atlas-Based Segmentation Using a Selective and Iterative Method for Performance Level Estimation (SIMPLE)*.

- IEEE Transactions on Medical Imaging, vol. 29, no. 12, pages 2000–2008, July 2010. (Cité pages 59, 65, 107, 112, 113 et 124.)
- [Le 2009] B.H. Le, Z. Deng, J. Xia, Y.-B. Chang et X. Zhou. *An Interactive Geometric Technique for Upper and Lower Teeth Segmentation*. In Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'09), Part II, volume 5762 of *LNCS*, pages 968–975. Springer Verlag, 2009. (Cité page 186.)
- [Leemput 1999] K. Van Leemput, F. Maes, D. Vandermeulen et P. Suetens. *Automated model-based tissue classification of MR images of the brain*. IEEE Transactions on Medical Imaging, vol. 18, no. 10, pages 897–908, 1999. (Cité pages 35 et 66.)
- [Leow 2005] A. Leow, S.-C. Huang, A. Geng, J. Becker, S. Davis, A. Toga et P. Thompson. *Inverse Consistent Mapping in 3D Deformable Image Registration : Its Construction and Statistical Properties*. In Proceedings of the 19th International Conference on Information Processing in Medical Imaging (IPMI'05), volume 19 of *LNCS*, pages 493–503. Springer Verlag, 2005. (Cité page 48.)
- [Leung 2010] K.K. Leung, J. Barnes, G.R. Ridgway, J.W. Bartlett, M.J. Clarkson, K. Macdonald, N. Schuff, N.C. Fox et S. Ourselin. *Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease*. Neuroimage, vol. 51, no. 4, pages 1345–1359, July 2010. (Cité pages 105, 106, 112, 126 et 217.)
- [Levendag 2004] P. Levendag, M. Braaksma, E. Coche, H. van Der Est, M. Hamoir, K. Muller, I. Noever, P. Nowak, J. van Sorensen De Koste et V. Grégoire. *Rotterdam and Brussels CT-based neck nodal delineation compared with the surgical levels as defined by the American Academy of Otolaryngology-Head and Neck Surgery*. Int J Radiat Oncol Biol Phys, vol. 58, no. 1, pages 113–123, Janvier 2004. (Cité page 25.)
- [Little 1997] J.A. Little, D.L.G. Hill et D.J. Hawkes. *Deformations incorporating rigid structures*. Computer Vision and Image Understanding, vol. 66, no. 2, pages 223–232, 1997. (Cité page 48.)
- [Lorenzen 2005] Peter Lorenzen, Brad Davis et Sarang C. Joshi. *Unbiased Atlas Formation Via Large Deformations Metric Mapping*. In Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'05), Part II, volume 3750 of *LNCS*, pages 411–418. Springer Verlag, 2005. (Cité page 53.)
- [Lötjönen 2009] J. Lötjönen, J. Koikkalainen, L. Thurfjell et D. Rueckert. *Atlas-based registration parameters in segmenting sub-cortical regions from brain MRI-images*. In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI'09), pages 21–24. IEEE Press, 2009. (Cité pages 105, 106, 110 et 217.)

- [Marsland 2003] Stephen Marsland, Carole J. Twining et Christopher J. Taylor. *Groupwise Non-rigid Registration Using Polyharmonic Clamped-Plate Splines*. In Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'03), Part II, volume 2879 of *LNCS*, pages 771–779. Springer Verlag, 2003. (Cité pages 50 et 51.)
- [Martin 2008] S. Martin, V. Daanen et J. Troccaz. *Atlas-based prostate segmentation using an hybrid registration*. *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, no. 6, pages 485–492, 2008. (Cité page 46.)
- [Martin 2010] S. Martin, J. Troccaz et V. Daanen. *Automated segmentation of the prostate in 3D MR images using a probabilistic atlas and a spatially constrained deformable model*. *Medical Physics*, vol. 37, no. 4, pages 1579–1590, 2010. (Cité page 34.)
- [Mazonakis 2001] M. Mazonakis, J. Damilakis, H. Varveris, P. Prassopoulos et N. Gourtsyiannis. *Image segmentation in treatment planning for prostate cancer using the region growing technique*. *Br J Radiol*, vol. 74, no. 879, pages 243–248, March 2001. (Cité page 33.)
- [Meller 2004] S. Meller et W.-A. Kalender. *Building a statistical shape model of the pelvis*. *International Congress Series*, vol. 1268, pages 561–566, 2004. Proceedings of the 18th International Congress and Exhibition of Computer Assisted Radiology and Surgery (CARS'04). (Cité page 64.)
- [Montagnat 1999] Johan Montagnat. *Modèles déformables pour la segmentation et la modélisation d'images médicales 3D et 4D*. Thèse de sciences, Université de Nice-Sophia Antipolis, Décembre 1999. (Cité page 34.)
- [Montagnat 2001] J. Montagnat, H. Delingette et N. Ayache. *A review of deformable surfaces : topology, geometry and deformation*. *Image and Vision Computing*, vol. 19, no. 14, pages 1023–1040, December 2001. (Cité page 34.)
- [Murgasova 2009] M. Murgasova, D. Rueckert, D. Edwards et J. Hajnal. *Robust segmentation of brain structures in MRI*. In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI'09), pages 17–20. IEEE Press, jun. 2009. (Cité page 217.)
- [Murgasova 2010] M. Murgasova, L. Srinivasan, I.S. Gousias, P. Aljabar, J.V. Hajnal, A.D. Edwards et D. Rueckert. *Construction of a dynamic 4D probabilistic atlas for the developing brain*. In Proceedings of the 7th IEEE International Symposium on Biomedical Imaging (ISBI'10), pages 952–955. IEEE Press, 2010. (Cité page 54.)
- [Ourselin 2000] S. Ourselin, A. Roche, S. Prima et N. Ayache. *Block Matching : A General Framework to Improve Robustness of Rigid Registration of Medical Images*. In Proceedings of the 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'00), volume 1935 of *LNCS*, pages 557–566. Springer Verlag, Octobre 2000. (Cité pages 46, 49 et 221.)

- [Ozen 2005] J. Ozen, B. Dirican, K. Oysul, M. Beyzadeoglu, O. Ucok et B. Beydemir. *Dosimetric evaluation of the effect of dental implants in head and neck radiotherapy*. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, vol. 99, no. 6, pages 743–747, June 2005. (Cité page 182.)
- [Park 2005] Hyunjin Park, Peyton H. Bland, Alfred O. Hero et Charles R. Meyer. *Least Biased Target Selection in Probabilistic Atlas Construction*. In Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'05), Part II, volume 3750 of *LNCS*, pages 419–426. Springer Verlag, 2005. (Cité page 50.)
- [Parraga 2007] A. Parraga, A. Susin, J. Pettersson, B. Macq et M. De Craene. *3D Atlas Building in the Context of Head and Neck Radiotherapy Based on Dense Deformation Fields*. In Proceedings of the 20th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'07), pages 321–328, oct. 2007. (Cité pages 50 et 51.)
- [Peroni 2009a] M. Peroni, M.F. Spadea, M. Riboldi, G. Baroni, G.T.Y. Chen et G.C. Sharp. *Validation of an automatic contour propagation method for lung cancer 4D adaptive radiation therapy*. In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI'09), pages 1071–1074. IEEE Press, 2009. (Cité page 41.)
- [Peroni 2009b] M. Peroni, J. Stancanello, A. Muacevic, M. Riboldi, M.-F. Spadea, G.-C. Sharp et G. Baroni. *Feasibility Study of B-Spline Deformable Registration between CT and MRI Atlas for Trigeminal Neuralgia Application to Radiosurgery*. In Proceedings of the 11th World Congress on Medical Physics and Biomedical Engineering (WC'09), volume 25/4, pages 1649–1652. Springer Verlag, 2009. (Cité page 42.)
- [Piccirillo 2007] J.F. Piccirillo, I. Costas et M.E. Reichman. *Cancers of the head and neck*. In SEER Survival Monograph : Cancer Survival Among Adults : U.S. SEER Program, 1988-2001, Patient and Tumor Characteristics, pages 7–22. 2007. (Cité page 8.)
- [Pollo 2005] C. Pollo, M. Bach Cuadra, O. Cuisenaire, J.-G. Villemure et J.-P. Thiran. *Segmentation of brain structures in presence of a space-occupying lesion*. *Neuroimage*, vol. 24, no. 4, pages 990–996, February 2005. (Cité page 42.)
- [Qazi 2010] A. Qazi, J. Kim, D. Jaffray et V. Pekar. *Probabilistic Refinement of Model-Based Segmentation : Application to Radiation Therapy Planning of the Head and Neck*. In Proceedings of the 5th International Workshop on Medical Imaging and Augmented Reality, volume 6326 of *LNCS*, pages 403–410. Springer Verlag, 2010. (Cité pages 65, 179 et 217.)
- [Ramus 2009] Liliane Ramus et Grégoire Malandain. *Using consensus measures for atlas construction*. In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI'09), pages 1190–1193. IEEE Press, July 2009. (Cité page 62.)

- [Ramus 2010a] L. Ramus et G. Malandain. *Multi-atlas based segmentation : Application to the head and neck region for radiotherapy planning*. In Proceedings of Head And Neck Auto-Segmentation Challenge Workshop, pages 281–288, September 2010. Satellite Workshop of MICCAI’10. (Cité page 102.)
- [Ramus 2010b] L. Ramus, J. Thariat, P.-Y. Marcy, Y. Pointreau, G. Bera, O. Commowick et G. Malandain. *Outils de contourage, utilisation et construction d’atlas anatomiques : exemples des cancers de la tête et du cou*. Cancer/Radiothérapie, vol. 14, no. 3, pages 206–212, 2010. (Cité page 31.)
- [Ramus 2010c] Liliane Ramus, Souha Aouadi, Yoann Pointreau, Guillaume Bera, Yungan Tao, Anne Beaudré, Vincent Grégoire et Grégoire Malandain. *Design and use of patient-specific atlases for radiotherapy planning of the head and neck region*. In Acts of the 29th European Society for Therapeutic Radiology and Oncology (ESTRO’10), volume 96, pages S–513, 2010. (Cité page 102.)
- [Ramus 2010d] Liliane Ramus, Olivier Commowick et Grégoire Malandain. *Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces*. In Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI’10), Part III, volume 6363 of LNCS, pages 155–162. Springer Verlag, September 2010. (Cité page 102.)
- [Ramus 2010e] Liliane Ramus et Grégoire Malandain. *Assessing Selection Methods in the Context of Multi-atlas Based Segmentation*. In Proceedings of the 7th IEEE International Symposium on Biomedical Imaging (ISBI’10), pages 1321–1324. IEEE Press, April 2010. (Cité page 102.)
- [Roche 1998] A. Roche, G. Malandain, X. Pennec et N. Ayache. *The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration*. In Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI’98), volume 1496 of LNCS, pages 1115–1124. Springer Verlag, Octobre 1998. (Cité page 45.)
- [Rohde 2003] G.K. Rohde, A. Aldroubi et B.M. Dawant. *The adaptive bases algorithm for intensity based nonrigid image registration*. IEEE Transactions on Medical Imaging. Special issue on image registration, vol. 22, pages 1470–1479, 2003. (Cité page 48.)
- [Rohlfing 2004] T. Rohlfing, R. Brandt, R. Menzel et C.R. Jr. Maurer. *Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains*. Neuroimage, vol. 21, no. 4, pages 1428–1442, 2004. (Cité pages 106, 110, 124, 125 et 126.)
- [Rohlfing 2005a] T. Rohlfing, R. Brandt, R. Menzel, D. Russakoff et C.R. Jr. Maurer. *Quo Vadis, Atlas-Based Segmentation ?* In The Handbook of Medical Image Analysis – Volume III : Registration Models, chapitre 11, pages 435–486. Kluwer Academic / Plenum Publishers, Août 2005. (Cité page 49.)
- [Rohlfing 2005b] T. Rohlfing et C.R. Jr. Maurer. *Shape-based averaging for combination of multiple segmentations*. In Proceedings of the 8th International

- Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'05), Part II, volume 3750 of *LNCS*, pages 838–845. Springer Verlag, 2005. (Cité page 64.)
- [Rohlfing 2007] T. Rohlfing et C.R. Jr. Maurer. *Shape-based averaging*. IEEE Trans Image Process, vol. 16, no. 1, pages 153–161, January 2007. (Cité pages 64 et 126.)
- [Rohlfing 2009a] T. Rohlfing, E.-V. Sullivan et A. Pfefferbaum. *Subject-Matched Templates for Spatial Normalization*. In Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'09), Part II, volume 5762 of *LNCS*, pages 224–231. Springer Verlag, 2009. (Cité page 54.)
- [Rohlfing 2009b] T. Rohlfing, E.V. Sullivan et A. Pfefferbaum. *Regression models of atlas appearance*. In Proceedings of the 21st International Conference on Information Processing in Medical Imaging (IPMI'09), volume 21 of *LNCS*, pages 151–162. Springer Verlag, 2009. (Cité page 54.)
- [Rouvière 1948] H. Rouvière. Anatomie humaine descriptive et topographique, pages 226–230. Masson, 6e édition, 1948. (Cité page 23.)
- [RTOG 2004] RTOG. *RTOG 0022 : Phase I/II study of conformal and intensity modulated irradiation for oropharyngeal cancer*, 2004. (Cité page 26.)
- [RTOG 2005] RTOG. *RTOG 0225 : A Phase II study of intensity modulated radiation therapy (IMRT) +/- chemotherapy for nasopharyngeal cancer*, 2005. (Cité page 26.)
- [RTOG 2010] RTOG. *RTOG 0920 : A Phase III study of postoperative radiation therapy (IMRT) +/- cetuximab for locally-advanced resected head and neck cancer*, 2010. (Cité page 26.)
- [Rueckert 1999] D. Rueckert, L. L. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach et D. J. Hawkes. *Nonrigid Registration Using Free-Form Deformations : Application to Breast MR Images*. IEEE Transactions on Medical Imaging, vol. 18, no. 8, pages 712–721, 1999. (Cité page 48.)
- [Rueckert 2001] D. Rueckert, A.F. Frangi et J.A. Schnabel. *Automatic Construction of 3D Statistical Deformation Models Using Non-rigid Registration*. In Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'01), volume 2208 of *LNCS*, pages 77–84. Springer Verlag, 2001. (Cité pages 52 et 54.)
- [Rueckert 2003] Daniel Rueckert, Alejandro F Frangi et Julia A Schnabel. *Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration*. IEEE Transactions on Medical Imaging, vol. 22, no. 8, pages 1014–1025, Août 2003. (Cité page 52.)
- [Sabuncu 2007] M.-R. Sabuncu, M.-E. Shenton et P. Golland. *Joint Registration and Clustering of Images*. In Proceedings of the Statistical Registration Workshop : Pair-wise and Group-wise Alignment and Atlas Formation, pages 47–54, 2007. Satellite Workshop of MICCAI'07. (Cité page 54.)

- [Sabuncu 2008] M.R. Sabuncu, S.K. Balci et P. Golland. *Discovering Modes of an Image Population through Mixture Modeling*. In Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'08), Part II, volume 5242 of *LNCS*, pages 381–389. Springer Verlag, 2008. (Cité page 54.)
- [Sabuncu 2009] M.-R. Sabuncu, S.-K. Balci, M.-E. Shenton et P. Golland. *Image-driven population analysis through mixture modeling*. IEEE Transactions on Medical Imaging, vol. 28, no. 9, pages 1473–1487, September 2009. (Cité page 54.)
- [Sammon 1969] J.W. Sammon. *A Nonlinear Mapping for Data Structure Analysis*. IEEE Trans. Comput., vol. 18, pages 401–409, May 1969. (Cité page 50.)
- [Sdika 2010] M. Sdika. *Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote*. Medical Image Analysis, vol. 14, no. 2, pages 219–226, April 2010. (Cité pages 111, 179 et 217.)
- [Shi 2010] F. Shi, P.-T. Yap, Y. Fan, J.-H. Gilmore, W. Lin et D. Shen. *Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation*. Neuroimage, vol. 51, no. 2, pages 684–693, June 2010. (Cité pages 105, 106, 110 et 118.)
- [Sotiras 2010] Aristeidis Sotiras, Yangming Ou, Ben Glocker, Christos Davatzikos et Nikos Paragios. *Simultaneous Geometric - Iconic Registration*. In Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'10), volume 6362 of *LNCS*, pages 676–683. Springer Verlag, 2010. (Cité page 46.)
- [Stadler 2008] M.E. Stadler, M.R. Patel, M.E. Couch et D.N. Hayes. *Molecular Biology of Head and Neck Cancer : Risks and Pathways*. Hematology/Oncology Clinics of North America, vol. 22, no. 6, pages 1099–1124, 2008. Head and Neck Cancer. (Cité page 6.)
- [Stapleford 2010] L.J. Stapleford, J.D. Lawson, C. Perkins, S. Edelman, L. Davis, M.W. McDonald, A. Waller, E. Schreibmann et T. Fox. *Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer*. Int J Radiat Oncol Biol Phys, vol. 77, no. 3, pages 959–966, July 2010. (Cité pages 30, 32 et 33.)
- [Stefanescu 2004] R. Stefanescu, O. Commowick, G. Malandain, P.-Y. Bondiaou, N. Ayache et X. Pennec. *Non-Rigid Atlas to Subject Registration with Pathologies for Conformal Brain Radiotherapy*. In Proceedings of the 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'04), volume 3216 of *LNCS*, pages 704–711. Springer Verlag, Septembre 2004. (Cité page 42.)
- [Studer 2006] G. Studer, S.-P. Studer, R.-A. Zwahlen, P. Huguenin, K.-W. Gratz, U.-M. Lutolf et C. Glanzmann. *Osteoradionecrosis of the mandible : minimized risk profile following intensity-modulated radiation therapy (IMRT)*.

- Strahlenther Onkol, vol. 182, no. 5, pages 283–288, May 2006. (Cité page 203.)
- [Studholme 1999] C. Studholme, D. L. G. Hill et D. J. Hawkes. *An overlap invariant entropy measure of 3D medical image alignment*. Pattern Recognition, vol. 1, no. 32, pages 71–86, 1999. (Cité page 45.)
- [Studholme 2003] Colin Studholme. *Simultaneous Population Based Image Alignment for Template Free Spatial Normalisation of Brain Anatomy*. In Proceedings of the 2nd International Workshop on Biomedical Image Registration (WBIR'03), volume 2717 of LNCS, pages 81–90, 2003. (Cité page 52.)
- [Subsol 1998] G. Subsol, J.-Ph. Thirion et N. Ayache. *A Scheme for Automatically Building 3D Morphometric Anatomical Atlases : application to a Skull Atlas*. Medical Image Analysis, vol. 2, no. 1, pages 37–60, 1998. (Cité page 43.)
- [Talairach 1988] J. Talairach et P. Tournoux. Co-planar stereotaxic atlas of the human brain. Thieme, 1988. (Cité page 50.)
- [Teguh 2010] D.N. Teguh, P.C. Levendag, P.W.J. Voet, A. Al-Mamgani, X. Han, T.K. Wolf, L.S. Hibbard, P. Nowak, H. Akhiat, M.L.P. Dirx, B.J.M. Heijmen et M.S. Hoogeman. *Clinical Validation of Atlas-Based Auto-Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck*. Int J Radiat Oncol Biol Phys, 2010. In Press. (Cité pages 30, 32, 56, 105, 106, 110, 124, 126 et 179.)
- [Teng 2006] C.-C. Teng, L.G. Shapiro et I. Kalet. *Automatic Segmentation of Neck CT Images*. In Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems (CBMS'06), pages 442–445, 2006. (Cité pages 33 et 37.)
- [Teng 2007] C.-C. Teng, L.G. Shapiro, I. Kalet, C. Rutter et R. Nurani. *Head and neck cancer patient similarity based on anatomical structural geometry*. In Proceedings of the 4th IEEE International Symposium on Biomedical Imaging (ISBI'07), pages 1140–1143. IEEE Press, apr. 2007. (Cité pages 107, 112 et 124.)
- [Terhaard 2005] C. Terhaard, P. Braam, J. Roesink, E. Astreinidou et N. Raaijmakers. *Significant improvement of parotid flow of IMRT compared to conventional radiotherapy in oropharyngeal cancer*. Int J Radiat Oncol Biol Phys, vol. 63, page S372, 2005. (Cité page 20.)
- [Terzopoulos 1988] D. Terzopoulos et K. Fleischer. *Deformable Models*. The Visual Computer, vol. 4, pages 306–331, 1988. (Cité page 34.)
- [Thariat 2010a] J. Thariat, E. de Mones, V. Darcourt, G. Poissonnet, P.-Y. Marcy, N. Guevara, A. Bozec, C. Ortholan, J. Santini, R.-J. Bensadoun et O. Dasonville. *Dent et irradiation : prévention et traitement des complications dentaires de la radiothérapie y compris l'ostéoradionécrose*. Cancer/Radiothérapie, vol. 14, no. 2, pages 137–144, 2010. (Cité pages 182 et 183.)

- [Thariat 2010b] J. Thariat, E. De Mones, V. Darcourt, G. Poissonnet, O. Dassonville, C. Savoldelli, P.-Y. Marcy, G. Odin, N. Guevara, A. Bozec, C. Ortholan, J. Santini et R.-J. Bensadoun. *Dent et irradiation : denture et conséquences sur la denture de la radiothérapie des cancers de la tête et du cou*. Cancer/Radiothérapie, vol. 14, no. 2, pages 128–136, 2010. (Cité pages 182 et 185.)
- [Thariat 2010c] Juliette Thariat, Liliane Ramus, Philippe Maingon, Vincent Grégoire, Vincent Darcourt, Serge Marcie et Grégoire Malandain. *Atlas de contourage automatique des dents pour l'optimisation des soins dentaires*. In Acts of the 21ème Congrès Annuel de la Société Française de Radiothérapie Oncologique (SFRO'10), 2010. (Cité page 182.)
- [Thariat 2010d] Juliette Thariat, Liliane Ramus, Philippe Maingon, Vincent Grégoire, Serge Marcie et Grégoire Malandain. *Construction of a dental atlas for automatic teeth segmentation to improve post-irradiation dental care management*. In Acts of the 52nd Annual Meeting of the American Society for Therapeutic Radiology and Oncology (ASTRO'10), 2010. (Cité page 182.)
- [Thariat 2011] J. Thariat, L. Ramus, P. Maingon, G. Odin, V. Grégoire, V. Darcourt, P.Y. Marcy, N. Guevara, M.H. Orlanducci, S. Marcie, G. Poissonnet, A. Bozec, O. Dassonville, F. Demard, L. Castillo, J. Santini et G. Malandain. *DENTALMAPS : Automatic Dental Delineation for Radiotherapy Planning in Head and Neck Cancer*. Int J Radiat Oncol Biol Phys, 2011. In press. (Cité page 182.)
- [Turaka 2010] A. Turaka, T. Li, N. Nicos, B. Burtness, M. Lango, J. Ridge, B. Weinberg et S. Feigenberg. *SU-GG-T-178 : Intensity-Modulated Radiation Therapy (IMRT) in the treatment of oropharyngeal carcinoma : clinical outcomes and relation of parotid gland volume with xerostomia*. In Proceedings of the 52nd of the Annual Meeting of the American Association of Physicists in Medicine (AAPM'10), volume 37, page 3226, 2010. (Cité page 20.)
- [van der Lijn 2008] F. van der Lijn, T. den Heijer, M.M.B. Breteler et W.J. Niesen. *Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts*. Neuroimage, vol. 43, no. 4, pages 708–720, December 2008. (Cité page 217.)
- [van Herk 1994] M. van Herk et H.-M. Kooy. *Automatic three-dimensional correlation of CT-CT, CT-MRI, and CT-SPECT using chamfer matching*. Medical Physics, vol. 21, no. 7, pages 1163–1178, Juillet 1994. (Cité page 43.)
- [van Rikxoort 2008] E.M. van Rikxoort, I. Isgum, M. Staring, S. Klein et B. van Ginneken. *Adaptive local multi-atlas segmentation : application to heart segmentation in chest CT scans*. In Proceedings of SPIE Medical Imaging, volume 6914, pages 691407.1–691407.6. SPIE, 2008. (Cité pages 106, 110, 112, 113, 124 et 126.)
- [van Rikxoort 2009] E.M. van Rikxoort, M. Prokop, B. de Hoop, M.A. Viergever, J.P.W. Pluim et B. van Ginneken. *Automatic segmentation of the pulmo-*

- nary lobes from fissures, airways, and lung borders : evaluation of robustness against missing data.* In Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'09), Part I, volume 5761 of *LNCS*, pages 263–271. Springer Verlag, 2009. (Cit  pages 55, 56, 105, 107, 110 et 124.)
- [van Rikxoort 2010] E.M. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M.A. Viergever, J.P.W. Pluim et B. van Ginneken. *Adaptive local multi-atlas segmentation : application to the heart and the caudate nucleus.* Medical Image Analysis, vol. 14, no. 1, pages 39–49, February 2010. (Cit  pages 59, 105, 110, 112, 113, 124 et 126.)
- [Vercauteren 2009] T. Vercauteren, X. Pennec, A. Perchant et N. Ayache. *Diffeomorphic Demons : Efficient Non-parametric Image Registration.* NeuroImage, vol. 45, no. 1, Supp.1, pages S61–S72, 2009. PMID : 19041946. (Cit  page 48.)
- [Vincent 1991] L. Vincent et P. Soille. *Watersheds in digital spaces : an efficient algorithm based on immersion simulations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, pages 583–598, Juin 1991. (Cit  pages 34 et 83.)
- [Viola 1997] P. Viola et W.M. Wells. *Alignment by maximization of mutual information.* International Journal of Computer Vision, vol. 24, no. 2, pages 137–154, 1997. (Cit  page 45.)
- [Wang 2010a] H. Wang, S. Das, J. Pluta, C. Craige, M. Altinay, B. Avants, M. Weiner, S. Mueller et P. Yushkevich. *Standing on the Shoulders of Giants : Improving Medical Image Segmentation via Bias Correction.* In Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'10), volume 6363 of *LNCS*, pages 105–112. Springer Verlag, 2010. (Cit  page 217.)
- [Wang 2010b] Q. Wang, L. Chen, P.-T. Yap, G. Wu et D. Shen. *Groupwise registration based on hierarchical image clustering and atlas synthesis.* Human Brain Mapping, vol. 31, no. 8, pages 1128–1140, 2010. (Cit  page 53.)
- [Warfield 2004] Simon K Warfield, Kelly H Zou et William M Wells. *Simultaneous truth and performance level estimation (STAPLE) : an algorithm for the validation of image segmentation.* IEEE Transactions on Medical Imaging, vol. 23, no. 7, pages 903–921, Juillet 2004. (Cit  pages 3, 65, 66, 100, 107 et 208.)
- [Wolz 2009] R. Wolz, P. Aljabar, D. Rueckert, R.A. Heckemann et A. Hammers. *Segmentation of subcortical structures and the hippocampus in brain MRI using graph-cuts and subject-specific a-priori information.* In Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI'09), pages 470–473. IEEE Press, jun. 2009. (Cit  page 217.)

- [Wu 2007] M. Wu, C. Rosano, P. Lopez-Garcia, C.-S. Carter et H.-J. Aizenstein. *Optimum template selection for atlas-based segmentation*. Neuroimage, vol. 34, no. 4, pages 1612–1618, 2007. (Cit  pages 55, 56, 110, 112 et 124.)
- [Wu 2009] D. Wu, X.-M Li, M.-S Yan et al. *The influence on the parotid function after IMRT in nasopharyngeal cancer*. Heilong Jiang Medical Journal, vol. 33, pages 168–170, march 2009. (Cit  page 20.)
- [Wu 2010] G. Wu, H. Jia, Q. Wang et D. Shen. *Groupwise Registration with Sharp Mean*. In Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'10), volume 6362 of LNCS, pages 570–577. Springer Verlag, 2010. (Cit  page 53.)
- [Yang 2010] J. Yang, Y. Zhang, L. Zhang et L. Dong. *Automatic segmentation of parotids from CT scans using multiple atlases*. In Proceedings of Head And Neck Auto-Segmentation Challenge Workshop, pages 323–330, September 2010. Satellite Workshop of MICCAI'10. (Cit  pages 55 et 105.)
- [Yu 1995] C.-X Yu. *Intensity-modulated arc therapy with dynamic multileaf collimation : an alternative to tomotherapy*. Physics in Medicine and Biology, vol. 40, no. 9, pages 1435–1449, 1995. (Cit  page 19.)
- [Yu 2002] C.-X Yu, X.-A Li, L. Ma, D. Chen, S. Naqvi, D. Shepard, M. Sarfaraz, T.-W Holmes, M. Suntharalingam et C.-M Mansfield. *Clinical implementation of intensity-modulated arc therapy*. Int J Radiat Oncol Biol Phys, vol. 53, no. 2, pages 453–463, June 2002. (Cit  page 19.)
- [Zhang 2007] Tiezhi Zhang, Yuwei Chi, Elisa Meldolesi et Di Yan. *Automatic Delineation of On-Line Head-And-Neck Computed Tomography Images : Toward On-Line Adaptive Radiotherapy*. Int J Radiat Oncol Biol Phys, vol. 68, no. 2, pages 522–530, 2007. (Cit  page 41.)
- [Zhang 2009] X. Zhang, J. Tian, Y. Wu, J. Zheng et K. Deng. *Segmentation of head and neck CT scans using atlas-based level set method*. In Proceedings of Head And Neck Auto-Segmentation Challenge Workshop, September 2009. Satellite Workshop of MICCAI'09. (Cit  pages 56, 105, 106 et 109.)
- [Zhuang 2010a] X. Zhuang, K. Leung, K. Rhode, R. Razavi, D. Hawkes et S. Ourselin. *Whole Heart Segmentation of Cardiac MRI Using Multiple Path Propagation Strategy*. In Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'10), volume 6361 of LNCS, pages 435–443. Springer Verlag, 2010. (Cit  pages 105, 106, 110, 112, 125 et 126.)
- [Zhuang 2010b] X. Zhuang, K.S. Rhode, R.S. Razavi, D.J. Hawkes et S. Ourselin. *A registration based propagation framework for automatic whole heart segmentation of cardiac MRI*. IEEE Transactions on Medical Imaging, vol. 29, no. 9, pages 1612–1625, 2010. (Cit  page 51.)
- [Zimmer 2005] C. Zimmer et J.-C. Olivo-Marin. *Coupled parametric active contours*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 11, pages 1838–1842, nov. 2005. (Cit  page 38.)

- [Zitová 2003] B. Zitová et J. Flusser. *Image registration methods : a survey*. Image and Vision Computing, vol. 21, no. 11, pages 977–1000, 2003. (Cité page 42.)

Résumé : L'objectif de cette thèse est de fournir des outils de segmentation automatique basés sur l'utilisation d'atlas anatomiques. Les algorithmes proposés sont appliqués et évalués dans le cadre de la radiothérapie de la région ORL : segmentation des organes à risque et des aires ganglionnaires pour la planification du traitement, et segmentation des dents pour l'évaluation des risques liés aux soins dentaires post-irradiation.

Nous abordons d'abord la construction d'un atlas moyen à partir d'une base d'images manuellement segmentées considérées comme atlas de référence. Nous proposons en particulier plusieurs algorithmes permettant d'estimer une segmentation moyenne à partir de segmentations incomplètes comportant des contours manquants.

Afin de dépasser les limites inhérentes à l'atlas moyen unique, et notamment ses difficultés à gérer la variabilité anatomique inter-individuelle, nous proposons de construire plusieurs atlas moyens à partir d'une stratification de la base d'atlas, et d'utiliser pour chaque patient l'atlas moyen le plus adapté parmi ceux disponibles. Nous présentons ensuite plusieurs approches permettant de sélectionner puis de fusionner pour chaque patient les atlas de la base les plus pertinents. Nous proposons des approches globales, régionales et locales de sélection d'atlas et un protocole non biaisé pour les évaluer. Nous proposons aussi différents schémas pour la fusion d'atlas, notamment la construction d'un atlas moyen composite adapté au patient. Nous montrons que la sélection et la fusion d'atlas apportent une amélioration significative par rapport à l'atlas moyen.

La plupart de ces travaux ont été intégrés dans le logiciel ISOgray de DOSIsoft.

Mots clés : segmentation par atlas, sélection d'atlas, conception d'atlas adaptés au patient, stratification d'atlas, radiothérapie, cancers ORL.

Design and use of anatomical atlases for automatic segmentation : application to radiotherapy of the head and neck

Abstract : The goal of this thesis is to provide automatic segmentation tools based on the use of anatomical atlases. The proposed algorithms are applied and evaluated in the context of radiotherapy of head and neck cancers : segmentation of the organs at risk and the lymph node levels of the neck for treatment planning, and segmentation of the teeth to assess the risk level associated to a given post-irradiation dental care.

We first address the construction of an average atlas from a database of manually delineated images considered as reference atlases. In particular, we propose several algorithms to estimate an average segmentation from incomplete segmentations that comprise of missing contours.

In order to get over the limits inherent in the average atlas, and in particular its difficulty to cope with the inter-individual anatomical variability, we propose to build several average atlases after stratification of the database into several sub-groups, and to use for each patient the most appropriate average atlas.

Then, we present several approaches to select and fuse the most relevant atlases of the database for each patient. We propose global, regional, and local approaches for atlas selection, and an unbiased protocol to assess them. We also propose several methods for the atlas fusion step, and in particular the construction of a patient-specific average atlas from the most similar atlases selected on each region. We show that the selection and fusion of atlases enable a significant improvement compared to the use of the average atlas.

Most of this work was integrated into the software ISOgray commercialised by DOSIsoft.

Keywords : atlas-based segmentation, atlas selection, patient-specific atlas, atlas stratification, radiotherapy, head and neck cancers.
