Production de parole après traitements de cancers de la cavité endobuccale.

† Christophe Savariaux, Pascal Perrier,
‡ Jacques Lebeau, Gustavo Magaña & Chloé Dorange-Pattoret

† Institut de la Communication Parlée, UMR CNRS 5009 – INPG & Université Stendhal
46 avenue Félix Viallet – 38031 Grenoble Cedex 01 – France
‡ Service de Chirurgie Plastique et Maxillo-Faciale, CHU Grenoble, BP 217
38043 Grenoble Cedex 09 – France

ABSTRACT

The ability to speak after tongue surgery including partial glossectomy or mandibulectomy and tongue reconstruction is studied. Fourteen French speaking patients pronounced a set of French vowels and VCV sequences. Data analysis was based on formant frequencies for the vowels, and on the burst spectrum, the VOT and the hold duration for the consonants. Different degrees of impairment were observed for the vowels, and they were used to classify the patients into 3 groups. Consonant analysis was then carried out separately for each group. Results are interpreted in terms of speech representations and vocal tract perturbations and speech quality after surgery.

1. INTRODUCTION

Cet article présente les résultats d'une évaluation de la production de parole chez des patients ayant subi une opération de la cavité endobuccale avec perte de substance puis reconstruction. Notre démarche diffère de celles qui sont classiquement publiées dans la littérature [Del97], en ce sens que nos analyses sont centrées non pas sur la capacité des patients à communiquer par la parole, mais sur leur habileté à produire les configurations géométriques du conduit vocal nécessaires à l'articulation de la parole. Pour cela nous nous sommes appuyés sur une évaluation quantitative fondée sur la mesure de paramètres pertinents extraits du signal acoustique de parole. Deux objectifs ont guidé cette démarche, celui de la compréhension des mécanismes de compensation mis en jeu par les locuteurs, et celui de l'étude de l'impact sur la parole des différentes techniques chirurgicales et de reconstruction appliquées aux patients.

2. PROTOCOLE EXPÉRIMENTAL

2.1 Corpus

Le corpus utilisé pour cette expérience a été élaboré pour évaluer les capacités des patients à :

- produire des voyelles tenues sans mouvement ; les 10 voyelles du français ont été étudiées.
- produire des transitions sans contrainte temporelle forte ; des séquences V-V ont été étudiées.
- contrôler précisément leur articulation tout en assurant une coordination temporelle entre la langue et les cordes vocales ; une série de séquences VCV a été étudiée, où le degré de complexité de la tâche augmentait depuis des séquences du type /aCa/ vers des séquences du type /aCi/ et /iCa/ qui impliquent un large mouvement de la langue.

2.2 Procédure

Quatorze patients de langue maternelle française ont été enregistrés au CHU de Grenoble. Ces patients ont subi une glossectomie partielle ou totale avec reconstruction, ou une mandibulectomie partielle. Les enregistrements ont été faits pour chaque patient lors d'une unique session, et dans un délai allant de 3 semaines à 3 ans après le jour de l'intervention chirurgicale. Le corpus a été présenté, sans consigne temporelle, dans l'ordre suivant : les voyelles isolées, les transitions V-V puis les séquences VCV.

2.3 Analyse des données

Le signal acoustique a été numérisé à 20 kHz. Pour les voyelles, l'analyse a été fondée sur les fréquences des 3 premiers formants déterminées par une analyse LPC. Conformément à une démarche classique, la variation de F1 a été interprétée comme une mesure du degré d'aperture du conduit vocal, alors que la variation de F2 a permis d'estimer la position relative avant/arrière de la langue. L'ordre de grandeur de F3 a été exploité pour évaluer la capacité du sujet à produire des articulations extrêmes dans la région palatale : un F3 élevé correspond à une articulation très antérieure (type [i]) et un F3 bas est associé à une articulation très postérieure (type [u]). Les consonnes ont été analysées dans le domaine temporel par la mesure du VOT et de la durée de la tenue consonantique, et dans le domaine spectral via le spectre de l'explosion (plage de fréquence dominante, forme de l'enveloppe spectrale) et les transitions formantiques. Ces données ont été comparées à des mesures effectuées dans les mêmes conditions sur 2 locuteurs normaux, ainsi qu'à des données publiées dans la littérature. Ici, l'étude est limitée aux plosives non-voisées [t, k].

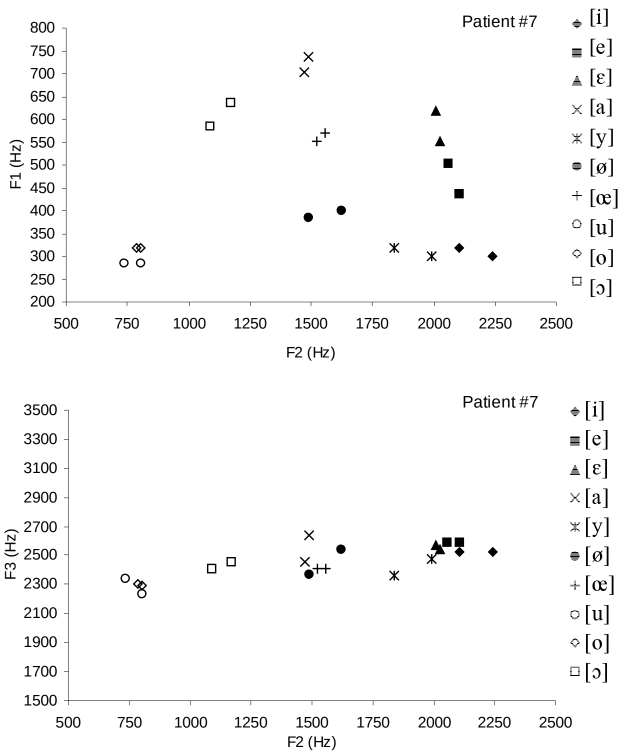
3. ANALYSE DES VOYELLES

3.1 Observations

L'analyse des productions vocaliques des 14 patients a permis de les classer en 3 groupes.

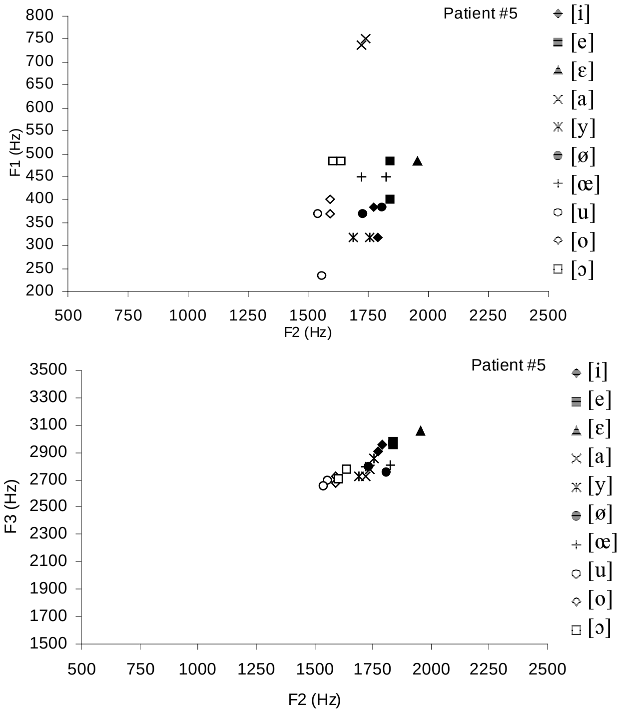
Le groupe 1, représenté par le locuteur #2, comprend 5 patients qui ont conservé la distribution classique des voyelles au sein du triangle vocalique, que ce soit en condition isolée ou en contexte. Il semble donc que, pour ces locuteurs, la chirurgie n'ait pas généré de gêne significative de la mobilité des articulateurs.

Le groupe 2 correspond à 6 patients qui semblent éprouver des difficultés à coordonner des mouvements simultanés avant/arrière et haut/bas de la langue et/ou à discriminer les voyelles fermées. C'est le cas du patient #7 dont la figure 1A montre les réalisations vocaliques dans le plan F1/F2 : on n'observe pas de difficultés notoires à produire les voyelles extrêmes [i, a et u], mais la position des autres voyelles à l'intérieur du triangle vocalique n'est pas conforme aux observations classiques. On note ainsi une position inhabituelle des voyelles centrales ([ɛ], [o]) dont l'articulation nécessite une coordination des gestes haut/bas et avant/arrière de la langue. La figure 1B confirme ces difficultés dans le plan F2/F3 : la plage de variation de F3 est très faible, se limitant, toutes voyelles confondues, à l'intervalle 2,2 – 2,6 kHz, suggérant que les positions extrêmes de la langue ne sont pas atteintes. Les séquences V-V sont correctes.



Figures 1A et 1B : Représentation des voyelles du patient #7 dans les plans F1/F2 et F2/F3.

Les 3 autres locuteurs, formant le groupe 3, montrent d'importantes difficultés pour produire des distinctions entre les 10 voyelles, comme en attestent les figures 2A et 2B pour le patient #5. F2 est cantonné entre 1,5 et 2 kHz, ce qui suggère une faible amplitude du geste avant/arrière de la langue. De plus, les voyelles sont mal séparées dans la dimension du premier formant, si on excepte la voyelle extrême [a]. Nous interprétons cela comme la conséquence de difficultés à contrôler différents degrés d'aperture. Toutefois il est intéressant de noter que l'ordre dans le degré d'aperture a été préservé. La faible dispersion des mesures dans le plan F2/F3 confirme que le patient ne parvient pas à déplacer sa langue dans des positions extrêmes (figure 2B).



Figures 2A et 2B : Représentation des voyelles du patient #5 dans les plans F1/F2 et F2/F3.

3.2 La nature de la chirurgie et ses effets

Cette classification est-elle cohérente avec la nature des interventions chirurgicales ? Pour répondre nous allons détailler les interventions subies par les sujets.

Le patient #2 (groupe 1) a subi 2 mandibulectomies à 5 ans d'intervalle, puis une radiothérapie. La reconstruction de l'os mandibulaire n'a pas été complète et sa partie gauche n'est pas solide. La langue est restée anatomiquement inchangée.

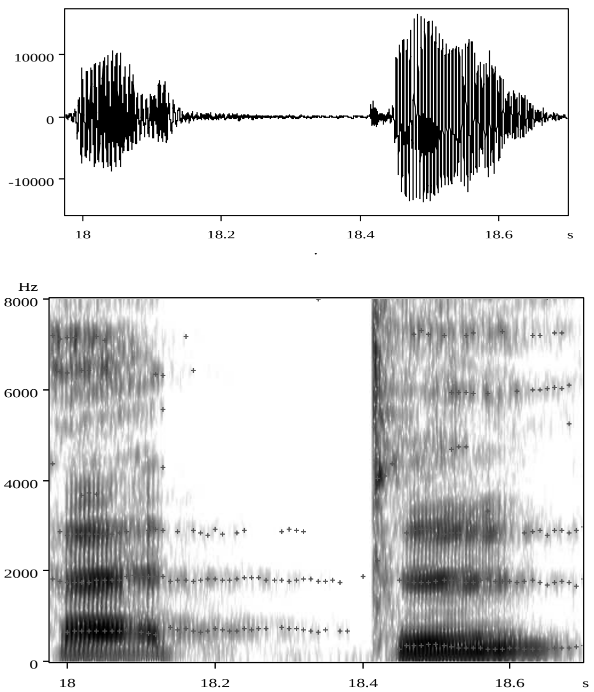
Le patient #7 (groupe 2) a subi l'ablation d'une large partie du plancher de la bouche ainsi que d'une petite partie de la racine de la langue. Le plancher de la bouche a été reconstruit avec un lambeau libre du muscle péroné. Ceci est associé à une sclérose des muscles de la langue.

Le patient #5 (groupe 3) a suivi une Curiethérapie qui a induit une fibrose des muscles de la langue. Trois ans plus tard, il a subi une hémi-glossectomie gauche, depuis la racine jusqu'au bout de la langue. Une reconstruction a été faite à partir d'un lambeau du muscle grand dorsal.

Ainsi il apparaît que l'impact de la chirurgie sur la langue va en croissant depuis le patient du groupe 1 à celui du groupe 3. La capacité qu'ont les patients à produire les voyelles décroît dans le même ordre. Ceci confirme le bien fondé de notre stratégie qui consiste à faire une première évaluation du degré de handicap sur la base des productions vocaliques.

4. COORDINATION DES MOUVEMENTS ARTICULATOIRES

Rappelons que seule l'analyse des plosives non-voisées [t, k] est faite dans ce papier.
L'étude du patient #5 témoigne de grosses difficultés à produire les consonnes. Pour le [t] de [ita] ou [ati], le spectre du burst est quasi identique à celui des locuteurs normaux, couvrant une large plage de fréquence avec un maximum entre 5 et 7 kHz. Après l'explosion apparaît un bruit de friction long et intense, surtout en contexte [ati] (figures 3A et 3B)



Figures 3A et 3B : Signal et sonagramme produit par le patient #5 durant la séquence [ati].

Pour [ata] le spectre du burst est différent de celui des productions normales : il est plus plat et on observe des zéros à 1,5 et 4 kHz. En revanche, aucun bruit de friction n'est observé. Ces résultats, dans leur ensemble, sont cohérents avec l'hypothèse d'une insuffisante mobilité de la langue, émise à partir de l'analyse des productions vocaliques. En effet, c'est lorsqu'un [i] précède ou suit la plosive [t], que la production spectrale est quasi normale. Or c'est justement dans ce cas que l'amplitude globale du mouvement de la langue est la plus faible. Le long bruit de friction observé dans ce contexte peut, lui aussi, être imputé à une trop grande lenteur de la langue qui, dans ce mouvement de faible amplitude, reste trop longtemps très proche du palais.

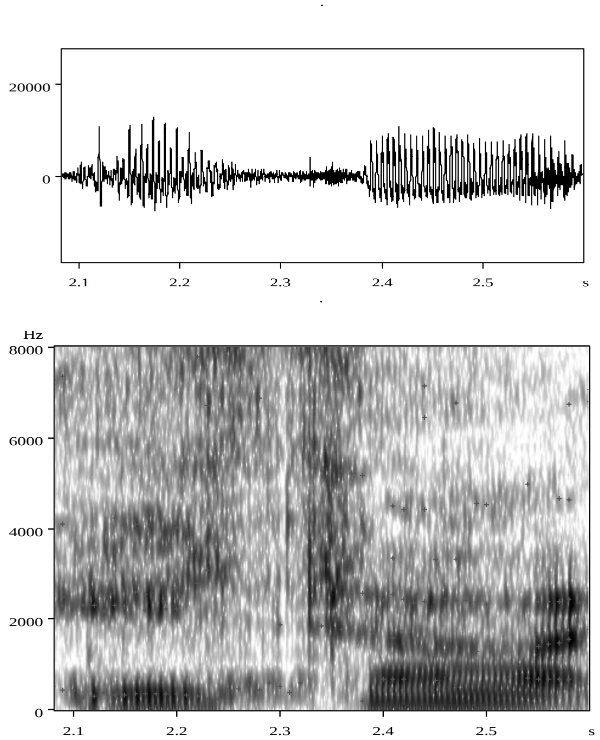
Cette hypothèse est confirmée aussi par les mesures de la tenue consonantique qui, pour [ati] et [ata], est deux fois plus longue que pour un locuteur normal (300 *vs*. 150 ms) et est significativement plus longue pour [ita] (220 *vs*. 150 ms). Les durées du VOT (30 ms) ne dépendent pas du contexte alors que pour les locuteurs normaux le VOT est plus court dans [ata] et [ita] que dans [ati] (30 *vs*. 70 ms). Or, il est admis généralement que, pour les locuteurs normaux, cette différence serait due au dévoisement de la voyelle [i] en contexte gauche [t] [Nea98]. Si on ne retrouve pas cet effet de coarticulation chez le patient, c'est probablement un signe que la stratégie globale de coordination glotte – langue a été altérée.

Le [k] du patient #5 est très différent de celui des locuteurs normaux. Le spectre du burst est quasi plat dans les 3 contextes, très similaire à celui du [t] de [ati] pour ce même patient, alors que, pour nos locuteurs de référence, on observe des maxima vers 4 et 7 kHz ainsi qu'un zéro entre ces 2 pics. Ce burst est suivi d'un bruit de friction intense et très long (plus de 100 ms), qui a de l'énergie autour 0,5 kHz et 2 maxima principaux vers 2 et 5 kHz. Les tenues consonantiques sont significativement supérieures à celles des sujets normaux. Cette durée est plus courte pour [ika] que pour [aka] et [aki] (300 *vs*. 350 ms). Le VOT de [ika] et [aka] est 2 fois plus long que pour les sujets normaux (85 *vs*. 40 ms), mais il est 2 fois plus court pour [aki] (27 *vs*. 75 ms). En contexte droit [a], le VOT est plus long pour [k] que pour [t] alors qu'en contexte droit [i], les VOT sont similaires pour les 2 consonnes. Cette différence est en accord avec les données issues de la littérature [Nea98].

À partir de ces mesures, il est possible de déduire que l'hémi-glossectomie subie par le patient #5 génère une forte gêne pour produire des consonnes articulées dans la région vélaire. Au contraire, l'articulation dans la région alvéo-dentale est possible, et nos observations laissent même penser que le [t] et le [k] sont tous les deux produits dans cette région. Malgré cette unicité du lieu d'articulation, le patient #5 préserve la distinction entre ces 2 consonnes, en générant pour le [k] un bruit de friction post-burst plus intense et un VOT plus long.

Pour le patient #7, le spectre de l'explosion du [t] varie selon le contexte : pour [ata], on note 3 maxima vers 3, 5 et 7 kHz ; pour [ati], l'allure générale du spectre est en cloche avec un maximum vers 3,5 kHz ; pour [ita], 2 maxima sont visibles vers 3 et 6 kHz. On observe les mêmes patrons temporels pour les 3 séquences : l'explosion est précédée d'un bruit de friction quasi-périodique, et elle est suivie d'un bruit très énergétique, mélange de friction et de plosion (figures 4A et 4B). Les mesures dans le domaine temporel sont similaires à celles des locuteurs normaux. Pour [k], le spectre du burst est similaire à celui des locuteurs normaux, mais le patron temporel est différent et très comparable à celui du [t], avec les mêmes types de bruit avant et après le burst. Les tenues consonantiques et les VOT sont comparables à ceux des locuteurs normaux.

Ces résultats suggèrent que ce patient arrive à articuler les consonnes au bon endroit, mais éprouve des difficultés pour déplacer sa langue avec les bonnes exigences temporelles, avant et après l'explosion.



Figures 4A et 4B : Signal et sonagramme produit par le patient #7 durant la séquence [ita].

Les résultats obtenus pour le patient #2 sont similaires à ceux des locuteurs normaux aussi bien dans le domaine temporel que spectral. Cependant, en contexte [ita] et [ata] les valeurs de tenue consonantique sont plus longues que les normales (200 *vs*. 150 ms). Ce dernier point semble suggérer que les mouvements articulatoires produits sont plus lents que pour un sujet normal.

5. CONCLUSIONS

L'analyse acoustique des productions vocaliques et consonantiques montre que la parole produite est de qualité différente pour les 3 patients analysés. Cette différence semble en rapport avec l'importance des modifications anatomiques qu'ont causées les interventions chirurgicales. La procédure proposée, fondée sur un corpus composé de logatomes sans signification linguistique et utilisant une analyse classique du signal dans les domaines temporel et fréquentiel, s'avère donc pertinente pour évaluer le degré de gêne des patients. On peut donc envisager de l'utiliser pour des évaluations de l'impact d'interventions chirurgicales d'importances similaires, mais faisant appel à des techniques différentes. Il nous a été aussi possible d'affiner les conséquences des perturbations anatomiques sur les gestes articulatoires utilisés lors de la production de la parole.

L'étude de locuteurs pathologiques, dans la lignée des travaux sur les perturbations artificielles de la parole [Gay81, Sav95], permet aussi d'avoir une meilleure idée de la représentation qu'a un locuteur de la tâche parole. Les conclusions que nous pouvons tirer des résultats exposés ci-dessus confirment celles de nos travaux sur les tubes labiaux : la représentation est clairement dans un espace perceptuel [Sav99] puisque le patient #5, empêché par sa pathologie de produire 2 lieux d'articulation différents pour [t] et [k], maintient une distinction par le biais d'un bruit de friction plus intense et d'un VOT plus long pour le [k]. Mais elle possède probablement des ancrages articulatoires [Sav95], puisque ce même patient respecte la hiérarchie dans le degré d'aperture des voyelles en dépit de l'absence de conséquence acoustique significative.

REMERCIEMENTS

Ces travaux ont été soutenus par une bourse de *La Ligue contre le Cancer*.

BIBLIOGRAPHIE

[Del97] Deleyiannis, FWB., Weymuller, Jr A. and Coltrera MD. (1997), “Quality of life of disease-free survivors of advanced (stage III or IV) oral pharyngeal cancer”, *Head & Neck*, 19, 466-473.

[Gay81] Gay T., Lindblom B. and Lubker J. (1981), “Production of bite-block vowels: Acoustic equivalence by selective compensation”, *J. Acoust. Soc. Am*., 69, 802-810.

[Nea98] Neagu A. (1998), “Représentations phonétiques et identification des syllabes occlusive-voyelle en français”, *Thèse de l'Institut National Polytechnique de Grenoble*, France.

[Sav95] Savariaux C., Perrier P. and Orliaguet J.P. (1995), “Compensation strategies for the perturbation of the rounded vowel [u] using lip-tube: A study of the control space in speech production”, *J. Acoust. Soc. Am*., 98, 2428-2442.

[Sav99] Savariaux C., Perrier P., Orliaguet J.P. and Schwartz J.L. (1999), “Compensation strategies for the perturbation of French [u] using a lip tube: II. Perceptual Analysis”, *J. Acoust. Soc. Am*., 106 , 381-393.