Speech production after glossectomy and reconstructive lingual surgery: a longitudinal study.

Christophe Savariaux[1], Pascal Perrier[1], Daniel Pape[12] & Jacques Lebeau[3].

[1]Institut de la Communication Parlée, UMR CNRS 5009, INPG & Univ. Stendhal,
46 Avenue Félix Viallet – 38031 Grenoble Cédex 01 – France
[2]ZAS – Research Centre for General Linguistics, Jägerstr. 10/11, 10117 Berlin, Germany
[3]Service de Chirurgie Plastique et Maxillo-Faciale, CHU Grenoble, BP 217,
38043 Grenoble Cédex 09 - France

Abstract

In this paper, the first results from a longitudinal study of speech production in French are presented for 9 patients who underwent vocal tract surgery, including partial or total glossectomy or pelvi-glossectomy and tongue reconstruction, or partial mandibulectomy. For each patient, two recordings were made of the acoustic speech signal, the first one a few days before surgery, and the second between 3 weeks and 12 months after surgery. The analysis of the data aimed at two main objectives. First we wanted to quantitatively assess the degree of speech production impairment induced by the surgery and to propose explanations in terms of articulation. Second, we were interested to observe and to understand how some of the subjects with a strong impairment were able to modify their speech production control strategies, in order to deal with the dramatic modifications of their oral cavity. Data analysis was based on formant patterns for the vowels, on Lisker and Abramson's VOT, and on temporal and spectral properties of the burst for the consonants. Speech production after surgery was evaluated on the basis of a comparison with the pre-surgery recording, which was considered as a reference describing the patients' speech production under normal condition. Special attention was paid to the patients who showed the largest post-surgery difficulties, in order to find the origins of their impairment and to study the strategy adopted to deal with this impairment.

Keywords : Speech Production, Glossectomy, Speech Pathology, Speech Perturbation

1. Introduction

It was often shown that the study of pathological speech could provide interesting information that helps to understand the control of speech production under normal condition. As a matter of fact, most of the recent studies focused on neural disorders such as stuttering ([1]), aphasia or apraxia of speech ([2]), neglecting the aspects associated with perturbations of the peripheral system. However, to study the impact of dramatic modifications of the peripheral speech production apparatus also permit to observe how the speakers can deal with the perturbation and elaborate new strategies in order to produce perceptually correct speech. It also offers a way to study the physical correlates of the perceptual task and to have some ideas about the representation of this task in the speakers mind.

At the same time, new tools have been recently developed in speech communication laboratories which permit to analyze speech signals and to relate the acoustical domain of speech production with the articulatory one. These tools can be very helpful for maxillo-facial surgeons and speech therapists, who like to evaluate reliably the quality of speech after the treatment of a pathology.

This paper is part of collaborative work between the Institut de la Communication Parlée and the University Hospital of Grenoble, which aims at both objectives : the evaluation of speech after surgery and a better understanding of the control of speech production from the study of pathological speech. We present the first results of a longitudinal study of patients who underwent vocal tract surgery due to a cancer of the oral cavity. We used an experimental protocol and a signal analysis procedure ([3], [4]) that differs from classical evaluation protocols proposed in the literature ([5]). Indeed, we do not evaluate the communication capability of the patients, but focus on the assessment of their capability to properly articulate the speech sounds under different conditions. This approach is likely to offer an quantitative method to compare the physical impacts of different surgical techniques, and into their consequences for speech production. A preliminary test of this approach was made in the past years on a set of 14 patients in post-surgery condition. The results showed that it was quite efficient for a quantitative assessment of the degree of speech production impairment of the patients. We could, indeed, propose some hypotheses about the articulatory causes for this impairment, and relate them to the nature of the surgery ([6]).

The new study differs from the preceding one, because all patients were recorded twice, once before and once after surgery. For each subject, speech production after surgery was then evaluated on the basis of the comparison with the pre-surgery recording. This offers a clearer view into the real impact of the surgery and about the re-learning process of speech articulation.

2. Method and Procedure

2.1. Corpus

The corpus consisted of a set of meaningless sequences, pronounced in isolation without any carrier sentence. Each kind of sequence was designed in order to evaluate a specific aspect of speech articulation. First, French vowels

/i, a, o, ɔ, u, y, e, ɛ, œ, ø/ were recorded with the aim of evaluating the capability to produce steady-state tongue positions without too strong accuracy requirements and without time constraints. Second, some Vowel-to-Vowel sequences, /i-a, a-i, i-e, e-i, i-ɛ, ɛ-i, u-y, y-u, ø-y, y-ø/ were recorded. It permits to test the capability of the patients to move their tongue from different locations in the vocal tract, in the absence of strong temporal constraints. In order to assess at the same time the control of precise articulatory configurations such as a consonantal constriction (position and degree of closure) or an occlusion, and the temporal control of the movement (including the glottis/vocal tract coordination), a set of VCV sequences were recorded, where V was /a/ or /i/ and C was one of the consonants /z, ʒ, s, ʃ, d, g, t, k, r, l/.

2.2. Subjects

Nine French speaking patients were recorded at Grenoble University Hospital on a Digital Audio Tape. No sound-treated room could be used for these recordings, but the quality of the signal was in general good enough for analysis. Patients were systematically recorded a first time within a few days before surgery and a second time after surgery within a period going from 3 weeks to a few months after surgery. The delay between surgery and the second recording depended on the agreement of the subject to do the recording, and could not be controlled better.

The patients were asked to pronounce the corpus as many times as possible, but due to their general physical and psychological condition, they often pronounced it only once or twice in the post-surgery condition. No time constraint or delay was given to the patients.

The patients underwent variable kinds of surgery including: 1) partial or total glossectomy with tongue reconstruction; 2) partial mandibulectomy or pelvi mandibulectomy; 3) partial or total pelvi glossectomy with reconstruction. A patient had an hemi-gingivectomy of the upper maxillary with extension to the sinus.

In order to respect the French law requiring the confidentiality of the medical records of a person, the patients will be labeled in the rest of the paper with anonymous numbers.

2.3. Data analysis

The acoustic signal was sampled at 20 kHz. For the vowels, data analysis was based on formant patterns measured with an LPC analysis of a 25 ms analysis window. The usual F1-F2 and F2-F3 planes were used to plot and to interpret the data. According to a classical approach, the variation of F1 is interpreted as a measure of the vocal tract aperture; while the variation of F2 gives information about the front/back tongue positioning. Concerning F3, it is only used to study high vowels, because its range of variation for these vowels is an index of the speaker's capability to produce extreme palatal articulations: a high F3 value is associated to a front articulation, like /i/, while a low F3 value correspond to a back vowel, like /u/ (see Fant's vocalic nomograms in [7] for more details).

The consonants were analyzed both in temporal and spectral domains. In this paper, only the voiced /d, g/ and the unvoiced /t, k/ stop consonants are studied. For them, in the temporal domain Lisker and Abramson's VOT ([8]) and the consonant hold duration were measured. In the spectral domain, characteristics of the burst spectrum, namely the frequency range, the slope, and the frequencies of the minima and maxima of amplitude, were measured. The burst spectrum was computed using an FFT for a 5 ms signal window that was located on the signal so that no trace of vocalic segment was included. In some cases, where the burst spectrum did not bring enough information, the F1-F2-F3 transitions of the VC sequence was also taken into consideration.

The acoustic data collected before surgery were considered for each speaker as the reference describing the speaker-dependent acoustical and temporal properties of speech under normal conditions. Consequently, the impact of the surgery as well as the compensation strategies developed by the speakers were analyzed from the data recorded after surgery through a comparison with the data measured before surgery. We are aware that this approach is not without any drawback, since tumor or pain do represent a potential perturbation of the speech before surgery. In addition, for some of the speakers, an extraction of all the teeth was made before the first recording session. For this reason, these "reference" patterns were systematically compared to other patterns measured on two normal subjects. In all cases, the range of the differences between data for the normal subjects and data for the patients in pre-surgery condition did not exceed the usual range of variation associated to the inter-speaker variability. This tends to validate our approach.

3. Production of vowels

The analysis of vowel production is primarily based on the vowels pronounced in isolation. Vowels in V-V or VCV contexts are examined afterward to see whether their characteristics are significantly different.

3.1. Before surgery

The representations in the F1-F2 and F2-F3 planes of the vowels recorded in isolation before surgery conformed, for the majority of patients, to the classical data published in the literature. However, it should be noted that for two of them an unusual variability could be observed for some vowels: /i, e, œ, a/ for patient #20 (see Figure 1, data in black) and front /i, y, ø/ vowels for patient #26. This suggests that for these speakers the production of some vowels could require more effort than for normal subjects.

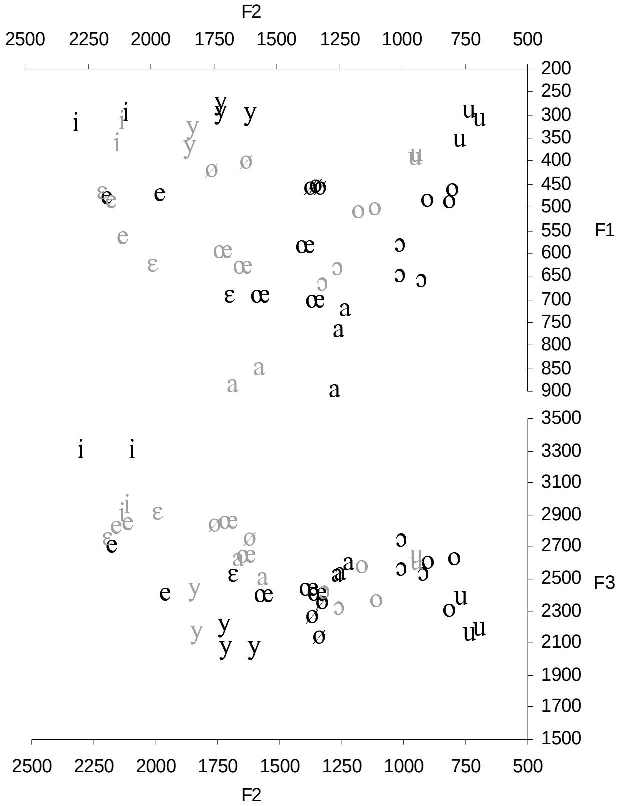

Fig. 1. Vowels produced by patient #20 plotted in the F1/F2 (top panel) and F2/F3 (low panel) planes respectively. Vowels produced before surgery are plotted in black and vowels produced after surgery are plotted in grey.

This phenomenon is, yet, not very strong and it can be concluded that the production of the vowels before surgery seems, for all speakers, to be very close to the natural one in spite of the presence of the tumor or of a possible pain.

3.2. After surgery

In line with our preceding studies, the patients were classified into 3 groups, based on the representation in the F1-F2 and F2-F3 planes of vowels after surgery. The first group includes 4 patients (#18, #24, #27 and #28), who managed to maintain, for the vowels pronounced in isolation, the classical distribution in the acoustic space. The vowels pronounced in the V-V and VCV sequences were also correctly located. For this group of patients, called G1 in the rest of the paper, it was considered that the surgery had no or very little impact on the speakers' capability to place and to shape their tongue in the whole vocal tract.

The second group, called G2, contains 4 patients (#20, #25, #26 and #30). These subjects failed to maintain the maximal frequency range of the vowel distribution in the acoustic space, suggesting that they encountered some difficulties either in the open/close or in the front/back direction. This phenomenon can be illustrated by the data measured for patient #20 in the F1-F2 plane (Figure 1 – top panel). In post-surgery condition the back vowels /u, o, ɔ, a/ show a higher F2 value as compared with the pre-surgery data. This trend is valid both for rounded and unrounded back vowels, and this suggests that patient #20 had some difficulties after surgery to move his tongue rear enough in the vocal tract. In addition, /u/ features a high F1 value, which could be the consequence of a difficulty to arch the tongue in velar region of the vocal tract. The observation of the data in the F1-F2 plane suggests that the patient has no difficulty to produce the front vowels /i, y, e, ɛ/, since F1 and F2 frequencies are very similar to the canonical values. However, in the F2-F3 plane (Figure 1 – low panel), it can be observed that, for vowel /i/, F3 has a lower value in the post-surgery condition. This suggests that the forward and upward tongue movement is somewhat limited for this patient after the surgery. These statements are confirmed by the data analyzed for V-V sequences.

The last group is actually limited to a unique patient (#19) who had obviously a big amount of difficulties to produce a distinction between the 10 vowels. This is attested in Figure 2 for patient #19.

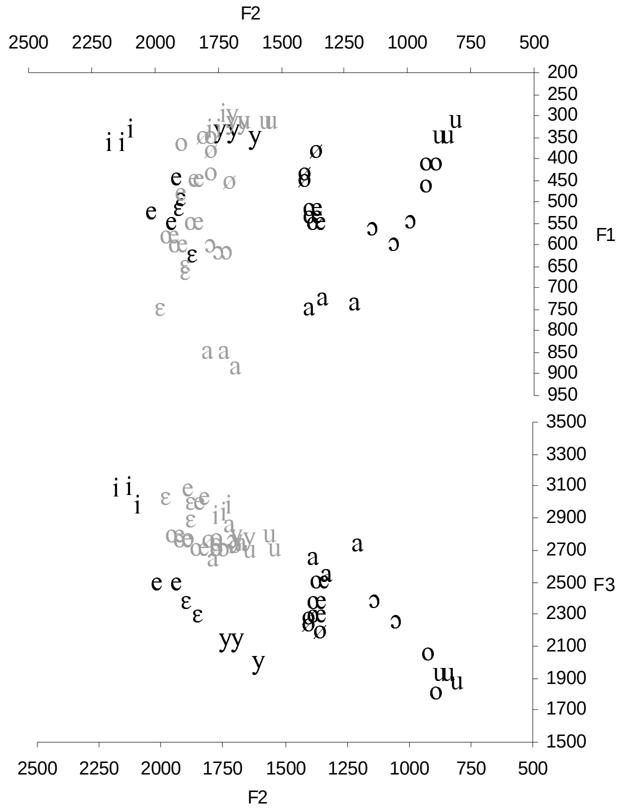

Fig. 2. Vowels produced by patient #19 plotted in the F1/F2 (top panel) and F2/F3 (low panel) planes respectively. Vowels produced before surgery are plotted in black and vowels produced after surgery are plotted in grey.

The specificity of the vowel production of this patient after tongue surgery lies in the very small range of variation in the F2 direction. It ranged from 1.5 to 2 kHz. This is an evidence that the amplitude of the front/back gesture is dramatically reduced. It can be confirmed by the representation in the F2/F3 space (Figure 2 – low panel), where all the data are concentrated in a small area. The limitation in the front/back direction is not observed in the open/close direction. Thus, it can be noted that among all the vowels the hierarchy in the degree of aperture is almost preserved: the range of variation of F1 is, indeed, correct and the distribution of the vowels along the F1 direction conforms to the standard one. The production of V-V sequences is in agreement with these observations.

4. Production of stop consonants

Since our preceding studies ([4], [6]) showed that the classification made on the basis of the skill in vowel production was a good predictor of the degree of impairment in the production of stop consonants, the rest of our analysis was made for each group separately.

4.1. Group G1

As could be expected, three of the four speakers who belong to this group did not show significant difficulties to produce the stop consonants after surgery. For these speakers, hold durations were in the range of 100 to 150ms, both for the pre-surgery and the post-surgery conditions, the unvoiced consonants being longer than their voiced counterparts. Quite a large inter-speaker variability could be observed in the VOT values, but their global ranges of variation (10 to 70ms) conform to data collected on normal speakers, and each speaker individually showed consistent durations from the pre-surgery to the post-surgery condition. Again inter-speaker variability could be observed in the temporal and spectral characteristics of the burst, but no significant impact of the surgery could be found.

For the fourth speaker (#24) of this group, the consonant hold duration was not affected by the surgery. However, noticeable differences could be observed for some sequences in VOT values and in the temporal description of the burst between the pre-surgery and the post-surgery condition.

In the pre-surgery condition, the VOT was around 70ms for /ati/, 50ms for /aki/, and 25ms for /ita/, /ata/, /ika/ and /aka/. After surgery, we found VOT values around 150ms for /ati/, 100ms for /ita/, 70 ms for /aki/, 60ms for /ika/ and /aka/; and 50ms for /ata/. Therefore, except for /aki/, the VOT after surgery was between two and four times longer than before surgery.

Concerning the temporal characteristics of the burst, we observed that for /aka/ and /ika/ the frication was longer. In addition, for /ika/, a noticeable frication noise was produced during the phase preceding the closure. For /t/, the burst is in some cases reduced to a frication noise. Its spectral characteristics conform to the standard properties of a /t/ if the plosion exists, or to the ones of a /s/ if there is only a frication. The same main tendencies where found for /g/ and /d/.

Since the spectral characteristics of the burst are correct, we can conclude that speaker #24 had no difficulty to articulate the consonants at the right place. However, the presence of a noticeable long frication noise at the consonant release, and in some cases during the closing gesture, suggests that this patient has difficulty to move quickly his tongue, at least when it reaches a high position in the vocal tract. The opening phase and the closing phase are thus longer, inducing airflow turbulences. This hypothesis is coherent with the lengthening of the VOT: since the release lasts longer, the pressure drop in the vocal tract takes a longer time, and the pressure drop through the vocal folds reaches later the value required to launch vocal folds vibration.

4.2. Group G2

Remember that in this group the speakers have some difficulties to produce either one of the extreme high vowels, /i/ or /u/ or both. Two of them (#25 and #30) did not present any significant change in the articulation of the stop consonants between the pre-surgery and the post-surgery condition. The hold durations, the VOT values, and the temporal and spectral characteristics of the burst conform to normal values (see Group G1 for values) and did not change with the condition.

For the two remaining subjects of this group, observed changes of the stop consonant production are coherent with the predictions that we could make from the distribution of the vowels in the acoustical space. This will be illustrated with the case of patient #20. Remember that for this patient our main conclusions were that the amplitude of the backward movement of the tongue was limited, and that he could not arch his tongue in the velopalatal region. In addition, we noted a small inaccuracy in the production of vowel /i/.

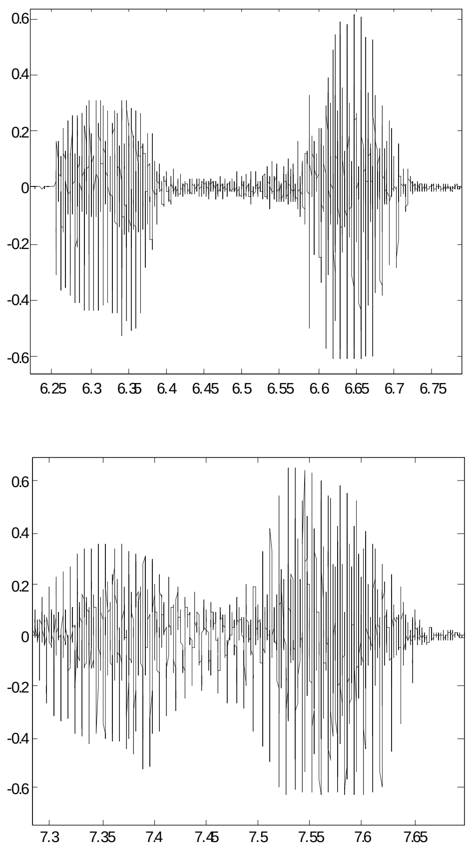

Fig. 3. Sequences /ika/ (top panel) and /iga/ (low panel) pronounced by patient #20 in post-surgery condition.

In the pre-surgery condition, the hold durations (around 170ms for /t/, 120ms for /d/, 140ms for /k/ and 130ms for

/g/) are in agreement with data for normal speech. The same statement can be made for the VOT values : 60ms for /ati/, 10ms for /ita/ and /ata/, 40ms for /aki/ and 20ms for /ika/ and /aka. The temporal and spectral properties of the burst are also normal.

In the post-surgery conditions, no change was observed for the alveo-dental /t/ and /d/. However, noticeable differences were observed for the velar consonants /g/ and /k/. The hold duration was as long as 200ms for /k/, and during this time, in all the vocalic contexts, a frication noise was produced (see Figure 3 – top panel). For this reason, we did not measure the VOT. For the voiced /g/ the hold duration was similar to the one in the pre-surgery condition, but the amplitude of the voiced signal during the closure phase was high and not compatible with the existence of an occlusion of the vocal tract (see Figure 3 – low panel). No burst was observed.

The correct spectral characteristics of the burst, when it exists, suggest that articulation location was correct for all the consonants (see [9] for experimental justifications). The patient articulated properly the alveo-dental consonants. However, the limitation that we observed for the vowels (that was in his incapacity to arch the tongue in the velopalatal region) seems to dramatically influence the production of the velar consonants. The patient was not able to produce a full occlusion in the vocal tract. The resulting inaccuracy of the closure gesture generated a frication noise in the case of the unvoiced consonant, and a semi-vowel-like signal in the case of the voiced one.

4.3. Patient #19

The analysis of the vowel production of this speaker revealed that he could not produce the required movements in the front/back direction, while the control of the open/close direction was mostly preserved.

In pre-surgery condition, the characteristics of the consonants were in agreement with the data measured for normal speech. In post-surgery condition, the hold duration was systematically longer. The temporal patterns of the unvoiced consonants were very variable. For /ata/ and /aka/ a low energy and short burst was observed. In one case a low energy frication noise was also produced during the beginning of the closure. In all cases, the VOT was shorter than 20ms. For /ita/, there was no noise at all, while for /ika/ a low energy frication noise was observed during the first part of the closure, followed by a short burst of low energy. When it was possible to measure it, the VOT was found to be somewhat longer (between 50 and 110ms) than the data classically measured on normal speech. For /aki/ and /ati/ a frication noise of fair amplitude was generated during the whole duration of the closure. For /g/ and /d/, in all the vocalic contexts, no burst was produced and the amplitude of the signal during the closure was very large, similar to the production of a semi-vowel.

Interestingly, in the majority of cases, the temporal signals were similar for the alveo-dental and the velar consonants (see examples on Figure 4). This suggests that the articulation location was the same for all these consonants. This observation is coherent with the one made for the vowels in the F2 direction. It seems that the patient had only very little freedom to move his tongue in the front/back direction. Consequently, he articulated all the sounds in the same front region of the vocal tract.

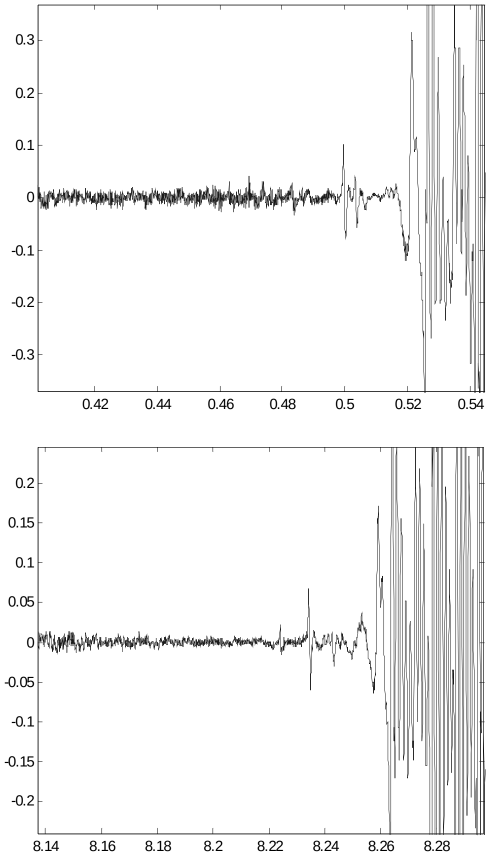

Fig. 4: Examples of the acoustic signal at the aperture of /t/ in /ata/ (top panel) and /k/ in /aka/ (low panel) for patient #19 in post-surgery condition.

The above described variability in the time domain would, then, be due to differences in the degree of constriction. The data for /g/ and /d/ show that these consonants were not produced with a full occlusion of the vocal tract. Similarly the small amplitude of the burst, if it existed, associated with the low amplitude frication noise observed during the closure of the unvoiced consonants, suggest that the speaker was again unable to produce a full occlusion, and that he also did not maintain the subglottal pressure during the closure duration of these consonants.

5. Interpretation

5.1. Interpretation in the light of the nature of the surgery.

The capacity of the patients to recover their skill in the articulation of speech is strongly subjects dependent. This variability can have many potential explanations, since the speakers did not undergo the same surgery, were recorded at different times after the surgery and went through different processes in order to adapt their articulation to the modifications of their oral cavity. However, for clinical purposes, it is interesting to check to which extent the degrees of articulatory skill that we found in the analysis of our data could be related to the nature of the surgery. For that, we will focus on four subjects: patient #18, who had no difficulty to produce proper vowels and consonants, and patients #24, #20 and #19, who were specifically analyzed in this paper and who were shown to have different degrees of speech production impairment.

Patient #18 underwent a hemi pelvi-mandibulectomy on the right side with a bone reconstruction. His tongue remained anatomically intact. Patient #24 had a resection of the front part of the tongue on the right side, without reconstruction. The tongue body itself was not modified. For patient #20, a part of the anterior mouth floor was removed and a reconstruction was made. Patient #19 underwent a full glossectomy including the removing of the posterior mouth floor. A reconstruction was made with a piece of the left dorsal muscle.

The comparison of the medical records of the patients with their articulatory skill in speech production leads to the following statements. The first one, quite obvious, is that the less the tongue body is modified during the surgery, the less the speech production is altered. Second, a resection of the tongue preserving the tongue body and the mouth floor does not strongly affect the capacity of the patient to displace his/her tongue in the vocal tract, preserving thus the quality of the vowel production. However, it probably stiffens the tongue in its upper part, explaining why patient #24 could not generate the fast movements required for the proper articulation of the stop consonants. Third, when the mouth floor is modified, the sounds articulated in the rear part of the vocal tract are more difficult to produce, again probably because of a stiffening of the tongue root. The actions of the styloglossus and the hyoglossus in particular are then less efficient. However, it is likely that a specific training could provide a noticeable improvement of the articulation of these sounds. Finally, removing the tongue body and replacing it with a piece of muscle has a dramatic impact on speech production. Tongue movements are then extremely limited, and it is very likely that most of the articulatory changes are produced with the mandible, essentially in the open/close direction.

5.2. Interpretation in terms of compensation strategies.

To study the compensation strategies developed by the patients after the surgery, we will specifically focus on patients #24, #20 and #19.

Patient #24 could articulate the consonants at the right place, but the durations were much longer than in normal speech. It induced the generation of an undesirable frication before and after the consonants. It is, so far, not possible to say whether the decrease of the speaking rate is a consequence of the possible stiffening of the tongue or whether it corresponds to a strategy elaborated by the speaker to reach the appropriate articulation locations. Further records should help clarifying this point.

Patient #20 could not produce an occlusion in the velopalatal region. He did not seem to have elaborated any strategy to facilitate the perception of the voiced stop /g/ which sounds like a semi-vowel. However, for /k/ he generated a strong frication noise to compensate for the absence of the burst, associated with the absence of a full occlusion.

Finally, patient #19 produced significant differences among vowels only in the open/close direction. No distinction could be made between vowels /u/ and /y/ or /ɛ/ and /ɔ/. Similarly, alveo-dental and velar stop consonants are very similar. However, it is interesting to observe that for unvoiced consonant this patient probably compensated for his incapacity to produce the occlusion: he reduced the subglottal pressure stopping the vocal fold vibrations and, therefore, reducing the velocity of the airflow in the vocal tract.

Acknowledgements.

This project was supported by the “Programme Cognitique” of the French Ministry of Education and Research. We also like to warmly thank our subjects for their participation to the study and for their availability.

REFERENCES

[1] De Nil, L.F., Kroll, R.M., Kapur, S. & Houle, S. (2000). A positron emission tomography study of silent and oral reading of single words in stuttering and nonstuttering adults. *Journal of Speech, Language & Hearing Research,* 43(3), 1038-1053.

[2] McNeil, M.R. & Kent, R.D. (1990). Motoric characteristics of adult aphasic and apraxic speakers. In. G.R. Hammond (Ed.), *Cerebral Control of Speech and Limb Movements* (pp. 349-386). Amsterdam, The Netherlands: North-Holland.

[3] Lebeau J., Perrier P., Magaña G., Bettega G., Sadek H., Dardart B., Mazaud F. & Raphaël B. (1998). La Parole après reconstruction des pertes de substance endobuccales induites par exérèse carcinologique : Quels outils d’évaluation ? Pour quelle fiabilité et quelle reproductibilité ? In A. Jortay (Ed.) *Réhabilitation anatomique et fonctionnelle après les traitements des cancers bucco-pharyngés.* (pp. 161-170). Éditions EDK: Paris

[4] J. Lebeau, C. Savariaux, P. Perrier, G. Bettega & B. Raphael (2000). Evaluation fonctionnelle des reconstitutions endobuccales. *Revue de Stomatologie et de Chirurgie Maxillo-Faciale, Vol. 101(2),* 60-64.

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

[6] Perrier P., Savariaux C., Lebeau J. & Magaña G. (1999). Speech production after tongue surgery and tongue reconstruction. *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 3, pp. 1805-1808). San Francisco, USA.

[7] Fant, G. (1960). *Acoustic Theory of Speech Production*. La Hague, The Netherlands : Mouton.

[8] Lisker, L. & Abramson, A.S. (1964). A cross language study of voicing in initial stops : acoustical measurements. *Word 20,* 384-422.

[9] Bonneau A., Djezzar L. & Laprie Y. (1996). Perception of place of articulation of French stop bursts. *Journal of the Acoustical Society of America, 100(1),* 555-564.