SPEECH PRODUCTION AFTER TONGUE SURGERY AND TONGUE RECONSTRUCTION

Pascal Perrier†, Christophe Savariaux†, Jacques Lebeau‡ and Gustavo Magaña‡.
† Institut de la Communication Parlée, UPRESA CNRS 5009 INPG & Univ. Stendhal, 46 avenue Félix Viallet – 38031 Grenoble Cédex 01 – FRANCE
‡ Service de Chirurgie Plastique et Maxillo-Faciale, CHU Grenoble, BP 217, 38043 Grenoble Cédex 09 - FRANCE

ABSTRACT

This study is part of a larger project consisting in observing the ability of 14 French speaking patients to re-organize their speech production apparatus, after tongue surgery including partial or total glossectomy and tongue reconstruction or partial mandibulectomy. They were asked to pronounce a corpus including French vowels and VCV sequences. Data analysis was based on formant patterns for the vowels, and on the burst spectrum together with the VOT for the consonants. Different degrees of impairment were observed during the production of the vowels which led to the classification of the patients into 3 main groups. Consonant analysis was carried out in relation with these 3 main groups. The results were interpreted both in terms of the representation that the patients have of their speech production task and in terms of the articulatory correlates of the impediment of speech.

1. INTRODUCTION

In this paper, an analysis of the speech production of patients who had surgery in the vocal tract due to a cancer is presented. This cooperative work between speech scientists and surgeons had two kinds of objectives. We wanted to set up a procedure to quantitatively assess the capability of each patient to articulate speech sounds; this method differs from the usual techniques used by physicians, in that we do not intend to evaluate the communication skills of the patients [1], but their motor and articulatory skills. For the clinicians, such quantitative evaluation is useful because it allows them to infer, from the acoustic signal, and with a fair level of confidence, what the articulatory origins of the patient's speaking difficulties are. It should help speech therapists to determine the best way for helping the patient to learn to speak again. It should also give interesting information to the surgeons about the impact of the surgery, and could contribute to help them to enhance their techniques for the surgery as well as for the reconstruction. For speech scientists, in the line of the classical perturbation paradigms [2, 5, 6] studying the way the patients deal with the modifications of their speech apparatus should help understanding what the representation of the speech task in the speaker's brain is.

2. METHOD AND PROCEDURE

2.1. Corpus

The corpus used for this experiment consisted of a set of sequences, without linguistic meaning, based on the following objectives:

*Evaluating the capability of each patient to produce steady-state tongue positions without too strong accuracy requirements; for that, the ten French vowels [i, a, o, ɔ, u, y, e, ɛ, œ and ø] were recorded in isolation.

*Evaluating the capability of the patients to move their articulators without strong temporal requirements; a set of V-V sequences were recorded such as [i-a, a-i, i-e, e-i, i-ɛ, ɛ-i, u-y, y-u, ø-y and y-ø].

*Evaluating at the same time, the control of precise articulatory configurations such as the constriction (position and degree of closure) or the occlusion during consonant production, and the time control of the movement associated with the co-ordination between vocal folds and tongue; a set of VCV sequences were recorded, where

V={a and i} and C={z, ʒ, s, ʃ, d, g, t, k, r, l}.

The degree of complexity of the production task increased gradually: first, consonants were produced only during /aCa/ sequences, in which the [a] vowel does not need a high level of articulatory control, and second, consonants were produced in /aCi/ and /iCa/ sequences involving larger amplitude movement of the tongue.

2.2. Procedure

A set of 14 French speaking patients was recorded at the Grenoble Hospital on a Digital Audio Tape with a microphone. No sound-treated room was used for this recording, but the quality of the signal is good enough for analysis.

The patients have undergone a surgery including partial or total glossectomy and tongue reconstruction, or partial mandibulectomy. They were recorded only once, and the time elapsed between the day of their surgery and the day of the recording varies from 3 weeks to 3 years.

They were asked to pronounce the corpus, which was written on 6 separated pages, in the following manner: the 10 French vowels pronounced in isolation, the vowel-to-vowel transitions, and the VCV sequences. No time constraint or delay was given to the patients.

2.3. Data analysis

The acoustic signal was digitized at 20 kHz. For the vowels, data analysis was based on formant patterns measured after an LPC analysis; the classical F1-F2 and F2-F3 planes were used to plot the data. The variation of F1 will be classically interpreted as a measure of the vocal tract aperture; while the variation of F2 will give information about the front/back tongue positioning. As concerns F3, only its range of variation will be taken into consideration as a way to evaluate the speaker's capability to produce extreme palatal articulations (high F3 for the front vowel [i] *versus* low F3 for the back vowel [u]).

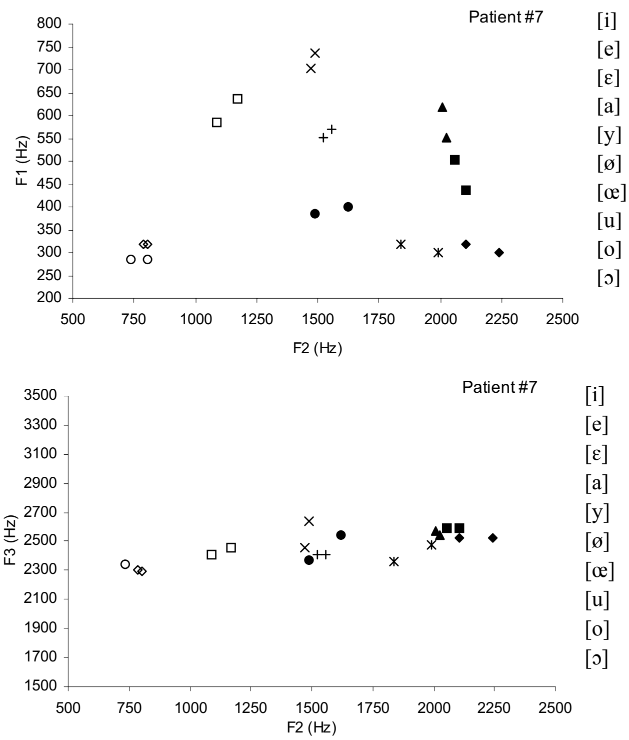
The consonants were analyzed in both temporal and spectral domains: in the temporal domain, the VOT and the consonantal hold duration were extracted; in the spectral domain, the spectrum of the burst (frequency range, slope, minima and maxima) and the F1, F2, F3 transitions were analyzed. These data will be interpreted through a comparison with data measured on two normal subjects and with data published in the literature. In this paper, only the unvoiced plosives [t, k] are presented.

3. ANALYSIS OF STEADY-STATE TONGUE POSITIONS

3.1. Observations

Starting from the representation of vowel productions in the F1-F2 and F2-F3 planes, the patients were classified into 3 main groups, depending on the level of distinction that they produced between vowels.

The first group includes 5 patients who managed, in spite of their surgery, to produce, for the majority of vowels, the classical vowel distribution within the acoustic space. The vowels pronounced in dynamic context (i.e. during V-V sequences) were correctly located within the acoustic vowel space. From these observations, it can be concluded that for this group of patients the surgery had no or very little impact on the speakers' capability to place and shape their tongue in the whole vocal tract.

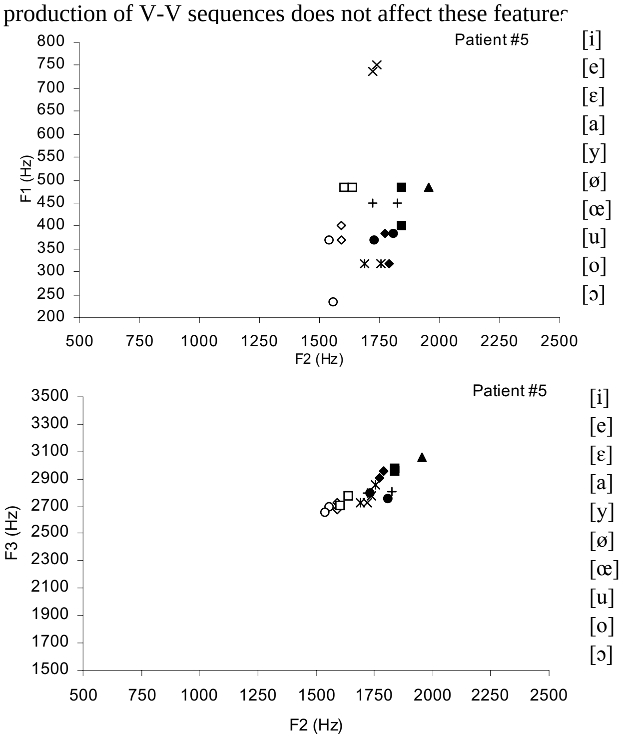


Figures 1A and 1B: Vowels of patient #7 plotted in the F1/F2 and F2/F3 planes respectively.

The second group, constituted by 6 patients, reveals some difficulties to produce simultaneously open/closed and front/back gestures of the tongue and/or to discriminate close vowels. This can be shown in Figure 1A, where the vowels produced by patient #7, are plotted in the F1-F2 plane. According to this figure, the patient has no difficulty in producing the extreme vowels [i, a, u]: F1 and F2 ranges are similar to canonical values. However the distribution of the vowels within these limits is peculiar: the patient has some difficulties producing the intermediate vowels (such as [ɛ] and [o]), which require a coordination of the front/back and open/closed gestures.

The data plotted in the F2/F3 plane (cf. Figure 1B) give more precision to these statements. The range of variation for F3 is very small: the 10 values are ranged between 2.2 and 2.6 kHz. This suggests that extreme front or back tongue positions are difficult to reach. The production of V-V sequences does not affect these features.

The 3 remaining patients, constituting the third group, showed great difficulties to produce a correct distinction between the 10 vowels. This is attested in Figures 2A & 2B for patient #5. The particularity of this patient is the poor variability of F2 ranged between 1.5 and 2 kHz. This suggests that the amplitude of the front/back gesture is very limited. It seems also that, except for the vowel [a], this patient has difficulty producing large differences in the degree of aperture to separate vowels. However, it should be noted that the hierarchy in the degree of aperture between vowels is preserved. Moreover, the variability is small in the F2/F3 space, suggesting that the patient has great difficulty moving his tongue in extreme positions. The production of V-V sequences does not affect these features.



Figures 2A and 2B: Vowels of patient #5 plotted in the F1/F2 and F2/F3 planes respectively.

3.2. Links with the nature of the surgery.

It should be reminded here that the classification into three groups was only based on acoustic measurements. These groups clearly depict differences among subjects in the importance of the impact of their handicap. It is thus interesting to observe whether or not this classification is consistent with the nature of the surgery undergone by the patients. For that we will focus on the three subjects considered above.

Patient #2 underwent two mandibulectomies, spaced by 5 years, and a radiotherapy. The bone reconstruction was not totally successful so that the left part of the mandible is not solid. The tongue remained anatomically intact.

Patient #7 underwent a surgery that removed a substantial part of the mouth floor and a little piece of the tongue root. The bottom of the mouth was reconstructed with a piece of the peroneal muscle. Surgery is also likely to have induced a sclerosis of the tongue muscles.

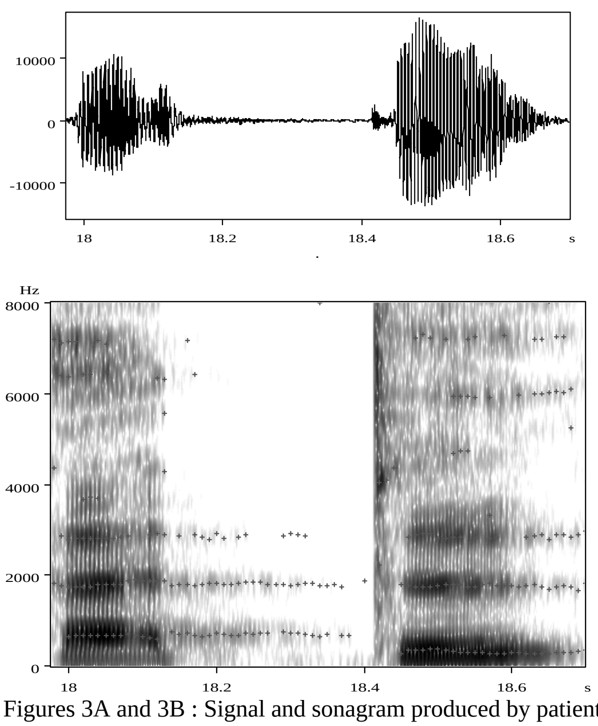
Patient #5 first underwent a Curietherapy that induced a fibrosis of the tongue muscles, and then a hemi-glossectomy of the linking part of the tongue, from the tongue root to the tongue tip. A reconstruction was also made with a piece of the dorsi muscle.

Thus from patient #2 to patient #5 via patient #7, the impact of the surgery on the tongue increases, from an impairment of the carrier articulator of the tongue, up to a removal of a substantial part of the tongue. It is interesting to note that the capability of the patients to properly produce vowels decreases exactly in the same order. Consequently studying vowel production acoustically seems to be a good tool for a first global evaluation of the degree of handicap.

4. COORDINATION OF ARTICULATORY MOVEMENTS

The analysis of the consonants is also exemplified with the 3 patients representative of the three groups determined from the analysis of the vowel production. Only the analysis of the unvoiced plosives [t, k] are presented in this paper.

For patient #5, who has great difficulties producing vowels, the production of these consonants is also extremely difficult. For the production of the [t] in [ita] or [ati], the burst spectrum is quite similar to those obtained in normal condition: a large frequency range with a maximum between 5 and 7 kHz. However, after the burst a long and intense frication noise is observed especially in [ati] context (cf. Figures 3A & 3B).



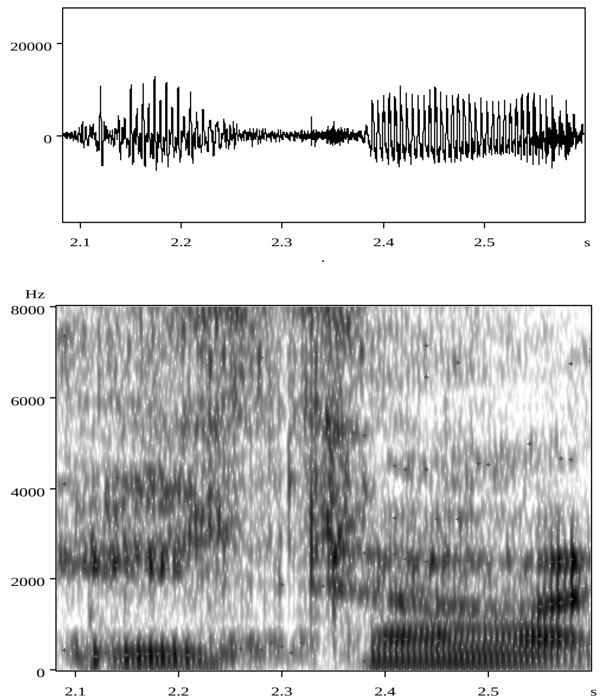
Figures 3A and 3B : Signal and sonagram produced by patient #5 during the [ati] sequence. The crosses on the sonagram represent the formant obtained after an LPC analysis.

In [ata], the burst spectrum is different from a normal one: it is more flat and zeros can be observed around 1.5 and 4 kHz. No abnormal frication noise is noticed. These observations are interpreted as resulting from the great difficulties for the patient to move the body of the tongue. When the vowel context is symmetrical and implies large amplitude movements ([ata]), the patient cannot reach the canonical [t] articulation with enough accuracy. In presence of [i], the global tongue movement is of smaller amplitude and the [t] is better produced: the sluggishness of the tongue could explain the important frication noise following the burst. In the temporal domain, the consonantal hold duration was twice longer in [ati] and [ata] for this patient than for a normal speaker (300 *vs*. 150 ms). In [ita], the consonantal hold duration is also significantly longer than for a normal speaker (220 ms). Concerning the VOT, the measures show the same duration for the 3 sequences (around 30 ms) whereas for a normal speaker the VOT is shorter in [ata] and [ita] than in [ati] (70 *vs.* 30 ms). This phenomenon in normal speech, could be due to the devoicing of the vowel [4]. The particularities observed for patient #5 suggested that the sluggishness of the tongue induced a re-organization of the glottis-tongue coordination.

The [k] produced by patient #5 and normal speakers show great differences. The burst spectrum obtained for the 3 sequences is quite flat whereas in normal speech 2 maxima are observed around 4 kHz and 7 kHz with a zero between these two maxima. In fact, the burst spectrum produced by this patient for [k] is quite similar to those produced for [t]. Moreover, a long (more than 100 ms) and very intense frication noise is observed after the plosion. This spectrum depicts a low frequency component around 0.5 kHz and 2 main maxima around 2 and 5 kHz. In the temporal domain, the consonantal hold durations obtained for the 3 sequences are all significantly longer than those measured on normal speakers. The same results as for [t] are observed: a shorter hold duration for [ika] than for [aka] or [aki] (300 *vs*. 350 ms). Concerning the VOT, they were twice longer than for normal speakers for [ika] and [aka] (85 *vs*. 40 ms) whereas they were twice shorter for [aki] (27 *vs*. 75 ms). Thus, in a right context [a], a longer VOT is observed for [k] than for [t]; in right context [i], the VOT are similar for the 2 consonants. This difference is in accordance with data published in the literature [4].

Thus, from the observations of acoustic signals it can be inferred that the hemi-glossectomy induced strong difficulties articulating consonants in the velar region implying a strong arching of the tongue. On the contrary, articulation in the dental-alveolar region seems to be possible so that, apparently, both [t] and [k] are articulated in this region. The distinction between these two consonants is produced by patient #5, first, in producing a more intensive frication noise after the [k] and, second, in keeping the VOT hierarchy between these two consonants.

For patient #7, the burst spectrum for [t] is different depending on the sequence: in [ata], 3 maxima are observed around 3, 5 and 7 kHz; in [ati], the envelop is bell-shaped with a maximum around 3.5 kHz; in [ita], 2 maxima around 3 and 6 kHz. From a temporal point of view, similar patterns are observed for the 3 sequences. As observed in Figures 4A & 4B, the burst was preceded and followed by noise. The frication noise before the burst has a quasi-periodic structure. After the burst, the noise is more energetic and consists of both frication and bursts. The frication part has a spectrum that is similar to the burst spectrum but a little bit more energetic. The results obtained in the temporal domain (consonantal hold and VOT) are quite similar to those of normal speakers, given the inter-speaker variability: the VOT is longer for [ati] than for [ata] and [ita] (70 *vs*. 45 ms).



Figures 4A and 4B : Signal and sonagram produced by patient #7 during the [ita] sequence.

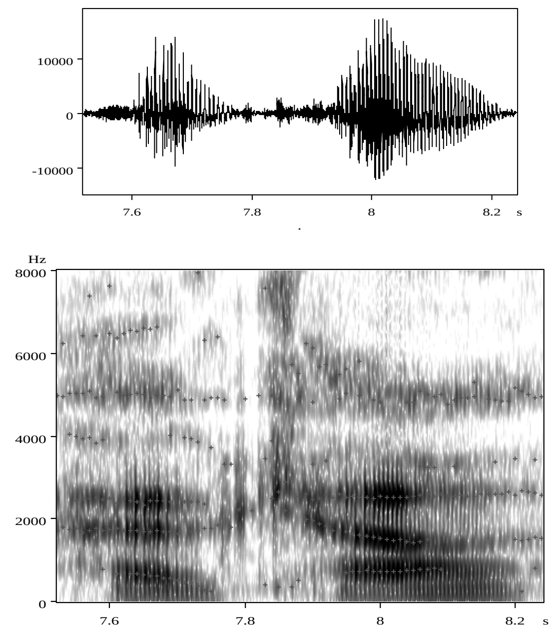
As concerns [k] for patient #7, the burst spectrum is similar to those of normal speakers: three maxima are observed around 3, 5 and 7 kHz. The same temporal pattern as for [t] is observed with the same kind of noise before and after the burst (see Figures 5A and 5B). The consonantal hold duration together with the VOT for the 3 sequences are in the same range as for normal speakers.

For this patient, our observations suggest that he had no difficulty articulating the consonants at the right place in the vocal tract but he had trouble shaping correctly his tongue and moving it with the appropriate timing before and after the burst.

As expected from the analysis of vowel production, the results obtained for patient #2 during the production of [t] and [k] both in the spectral and temporal domains show that this patient can produce these consonants like a normal speaker does. Indeed, all the measures of the burst and VOT are similar to those obtained for normal speakers. In two cases ([ita] and [ata]) the consonantal hold duration is longer than the normal one (200 *vs*. 150 ms). This last point suggests that general movement could be slower for this patient than for a normal speaker.

5. CONCLUSIONS

The 3 patients analyzed showed different degrees of difficulty during production which are consistent with the degree of impediment that was induced by the surgery. Hence the procedure that was set up, including a corpus with simple non-sense sequences and classical speech signal analysis based on spectral and temporal parameters, seems to be efficient to evaluate the degrees of impairment of the patients. It was also possible to propose some origins, in articulatory terms, for this impairment. In parallel, our observation was also helpful to understand the representation of the speech task in the speaker mind: it is certainly partially articulatory since patient #5 still respects the hierarchy of aperture between vowels although it has no acoustical impact. It is also clearly perceptual since patient #7 provides a distinction between [t] and [k] by the means of an intense frication noise and a longer VOT for [k].



Figures 5A and 5B : Signal and sonagram produced by patient #7 during the [aka] sequence.

ACKNOWLEDGMENTS

This work was supported by a grant from the *Ligue contre le Cancer*.

REFERENCES

[1] 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.

[2] 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.

[3] Ladefoged, P. and Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

[4] Neagu, A. 1998. Représentations phonétiques et identification des syllabes occlusive-voyelle en français. *Ph D thesis of the Institut National Polytechnique of Grenoble*, France.

[5] 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.

[6] 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*. (in press).