Methodology for Acoustic Characterization of a Labial Constraint in Speech Production

Leila Falek*, Hocine Teffahi, and Amar Djeradi

USTHB, Electronics and Computer science Faculty, Algiers, Algeria
lfalek@hotmail.fr

Abstract. We propose in this study, a method allowing us to characterize a speech occurred in a stress situation. For this, we created an artificial disturbance (stress lip) and then, we analyzed the effects of stress on the acoustic parameters of the signals produced. We have developed a methodology allowing us to analyze the timing, the fundamental frequency, formants frequencies (calculated with wavelet transform methodology) and the coarticulation between consonant and vowel. These parameters are generally used to produce vector descriptors in communication systems. Articulatory interpretation of results typifies well the constraint used.

Keywords: speech disturbance, acoustics parameters, labial constraint, formants frequencies, fundamental frequency, coarticulation, locus equation.

1 Introduction

At present, in an effort to get as close as possible to a natural speech, recognition systems, coding or speech synthesis are in search of invariants in the acoustic cues of natural speech taking into account all the constraints under which useful information is produced. Also, much research has emerged in the study of speech artificially disturbed [1]. These studies provide a better understanding of motor control in a constraint situation and thus demonstrate the effect of stress on these acoustic parameters to better use in communication systems.

Moreover, it is possible to disturb speech production in several ways: either directly or indirectly by modifying the vocal tract geometry, or by interrupting some kind of feedback (tactile, auditory or other). We were interested here at the vocal tract geometry constraint [1]. (The geometric perturbation can mean an oral handicap, or when a person speaking with an object in the mouth).

A tube inserted between the lips of four speakers served as a perturbation of the lips protrusion, during the production of 7 French vowels. The aim is then to analyze the effects of this disturbance on the acoustic parameters of recorded sounds, by analyzing the compensatory strategies adopted by the speakers for better adaptation to the stress, for reach their target sound.

* Corresponding author.

A. Elmoataz et al. (Eds.): ICISP 2012, LNCS 7340, pp. 131–141, 2012.

With Using the experiments of artificial perturbation of speech production, McFarland et al. (1996)[8] have revealed a significant difference : the vowels production is more affected by a change of oral function. The behavior of the vowels under stress

then is revealing information about motor control of speech. This justifies vowels analysis of in this study.

2 Experimental Setup for Labial Constraint Realization

The aim is to constrain the lip area, so we chose to use a lip ring diameter 3.4 cm and 1 cm in length (rigid), so as to compel the production of all vowels: it will allow for increase the lips area in the production of rounded french vowels [u] and [ou], and lower lip area in the production of [i], [a], [e] [é]. Although the variation in the lip area is the desired result, the insertion of a tube between the labial lips of the speaker also has the effect of causing a slight increase in the lips protrusion, as the speaker must use his lips to completely surround the tube, so that it can't fall and there is no air flow around the tube.

3 The Sentences Corpus

We conducted a corpus of french sentences composed of sequences consonant-vowel-consonant (CViC), which were incorporated into an interrogative carrier sentence for preserve the natural speech: nVina est la? With Vi = [a], [i], [ou], [o] ,[u], [e] and [é]. The consonant /n/ is chosen because it is a voiced consonant, to facilitate the acoustic analysis. The sentences of the corpus are then: “Nana est la ? “; “ Nina est la ?” ; “ Nouna est la ? “; “ Nena est la ?” ; “ Néna est la ?”; “ Nona est la ?”; “Nuna est la ?”. Each sentence is repeated seven times. These sentences were recorded initially without constraint and with constraint. The number of speakers is 4 (3 women and 1 man). What makes a corpus of: 7x10x4x2 = 560 sentences. The recording of the corpus was conducted in a calm, using a desktop computer with a sound card "Sound Blaster" , the Praat software and a microphone.

4 The Units Analyzed

The units in question are /CViC/, Vi is one of the seven french vowels (/a/, /i/, /ou/, /o/, /u/, /e/, /é/), in consonantic context /C/= /n/, for each speaker : /nan/, /nin/, /nen/, /nun/, /nén/, /non/, /noun/. to obtain these units, we manually segmented the corpus of sentences using Praat software. The analysis was done in the time domain and frequency domain.

In the time domain: we analyzed the durations (CViC) depending with the duration of the sentence, (d(CViC): average length of the unit CViC and d(sentence): average duration of the sentence , corresponding to the unit: CViC). The aim is to study the constraint impact on the temporal distribution in the analyzed sentence.

In the frequency domain: we analyzed the formant frequencies F1 and F2 and the fundamental frequency F0 of the vowel "Vi" (taken in the vowels middle).

We then analyzed the influenceof the constraint on vowel space and on the coarticulation (using the slope of the straight locus equations)

5 Analysis of Labial Constraint Influence on the Timing of "VCiV"

The comparison between the normal condition and in constraint condition is made on dispersions ellipses basis of average durations (d(CVC), d(sentence)) of each vowel , for each speaker, to take account of all samples. The abscissas of dispersion ellipses correspond to the average durations of sentences and the ordinates to length units VCiV. In blue color we have the situation without constraint and in red color with stress.

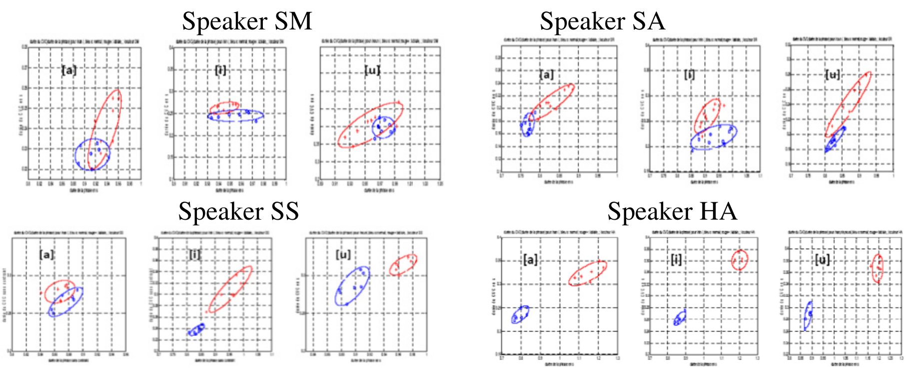

Fig. 1. Dispersion ellipses of average lengths (d (CViC), d(sentence)) of each vowel for each speaker

Figures 1 show an increase in the duration of VCiV in the sentence, for four speakers. So, the constraint has the effect of slowing the speech rate

6 Constraint Influence Analysis on the Vocalic Space

We calculated the formants frequencies F1 and F2 in the middle of the vowel, using a calculation method of formant trajectories, based on the complex continuous Morlet wavelet transform that we have developed.

6.1 Methodology for Calculating Formants Based on Complex Continuous Wavelets Transform

The wavelet transform is to decompose a signal x (t) in a family of functions ψ (t), localized in time and frequency called wavelets. It is defined by the relation [5 13 19]

$$\left(W_{\psi}x\right)(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\psi^{*}\left(\frac{t-b}{a}\right)dt \quad (1)$$

It is used to describe the frequencial content of a signal x(t), locally near a point (a, b), in the time-scale ("a" scale and "b" is the time). It indicates us the relative importance of frequency "1/a" around time "b" for the signal x (t).

The complex continuous wavelet transform is used to define the notion of instantaneous frequency (and formants), when using an analytic wavelet [17]. The analytic signal is a complex signal associated with a real signal, which provides access to the signal phase, therefore, at its frequency.

It was proposed by VILLE in 1948 **[20]**. This latter has interesting properties, particularly in terms of its Fourier transform, which is zero for negative frequencies.

An analytic signal z (t) can be calculated from the Hilbert transform [10] of a real signal x (t) such that:

$$H\left(x(t)\right) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x(s)}{t-s}ds \quad (2)$$

$$z_x(t) = x(t) + iH(x(t)) \quad (3)$$

H(x(t)) is the Hilbert transform of x(t) .

It is possible to define [6] from an analytic signal, an instantaneous frequency $f_x(t)$ [7], where $z_x(t)$ is an analytic signal

$$f_x(t) = \frac{1}{2\pi}\frac{d\,arg(z)}{dt}(t) \quad (4)$$

6.1.1 The Morlet wavelet

It is an analyzing wavelet **[15]** for small oscillations (a center frequency fc: around 1 Hz). It is very well localized in time. It is inspired by the Gabor elementary signal. It is obtained by modulation of a Gaussian. It is given by the following equation [3]:

$$\Psi(t) = \frac{1}{c}e^{j\omega_c t}\left[e^{-\frac{t^2}{2\sigma_t^2}} - \sqrt{2}\; e^{\frac{\omega_c^2\sigma_t^2}{4}} \quad e^{-\left(\frac{t^2}{\sigma_t^2}\right)}\right] \quad (5)$$

The product $\omega_c*\sigma_t$ fixed the link between the width of the Gaussian envelope of the wavelet and its oscillation frequency fc. For the Morlet wavelet $\omega_c = 2\pi fc$: $5 \leq \omega_c \leq 6$, so, $\omega_c*\sigma_t \geq 5$, $0.8 \leq Fc \leq 1$) Hz [3]

6.1.2 Presentation of the Method

The method we developed was made from some properties of time-frequency distribution and wavelet transform, namely: For a complex time-frequency distribution, there is information in the module and phase of the distribution.

Near the center frequency fc of wavelets, whose cyclicality corresponds to the nature of the signal: the wavelet transform amplitude has a maximum [4]. The phase coefficients of the time-frequency distribution varies cyclically at a frequency close to the instantaneous frequency of the signal, therefore the derivative of the phase coefficient is close to the instantaneous frequency of the signal. So, we will say that for a given time, the instantaneous frequency can be estimated by: the maximum of the distribution module, the fixed point of the time derivative of the phase of time-frequency distribution [6] [19]

For formants calculation, it is important to add that in this case, the phase derivative alone is not sufficient since the formant frequencies correspond to the instantaneous maximum power. From this information we developed an algorithm for calculating the instantaneous frequencies of formants contained in a speech signal [10].

The following figures show the different steps of the method developed, applied to a signal x(t) composed of a sum of four sinusoids:

$$x(t)=5*\sin(2\pi*300t)+10\sin(2\pi*1000t)+ 15\sin(2\pi*3000t)+ 10\sin(2\pi*5000t)$$

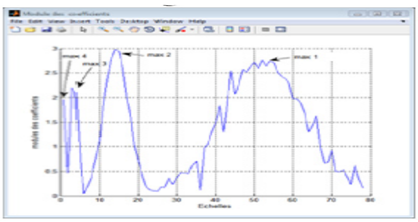

Fig. 2. modulate coefficients of the wavelet transform according to the scales

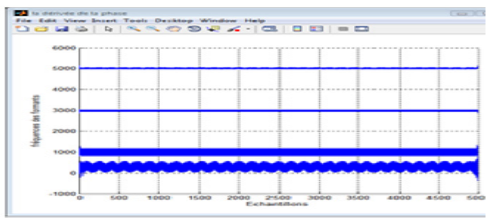

Fig. 3. Phase derivative of wavelet transform corresponding for each maximum

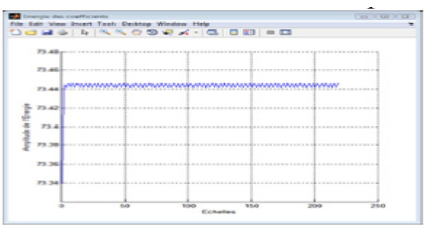

Fig. 4. Coefficients energy of the wavelet transform, for one frequency

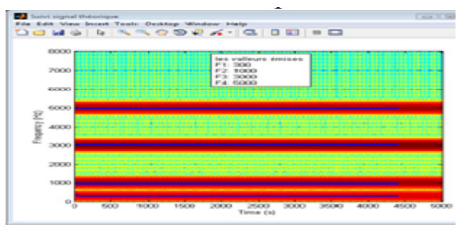

Fig. 5. Maximum energy for the 4 instantaneous frequencies

The results obtained in Figures 2, 3, 4 and 5 show that the method used to calculate the instantaneous frequencies allows to find the frequencies of the analyzed signal x(t).

Using this method, we have determined the formant values F1 and F2 in order to trace the vocalic spaces corresponding, in normal condition and in labial constraint, for the 10 samples, for each vowel and for each speaker. We then represented each vowel by its dispersion ellipse in the vowel space. We then plot surfaces of the vowel triangles [7] in the vocalic space (F1, F2) for each speaker in two conditions: normal condition and in labial constraint. Figure 6 shows results obtained for each speaker, and table 1 shows effects of stress on areas of vocalic triangles.

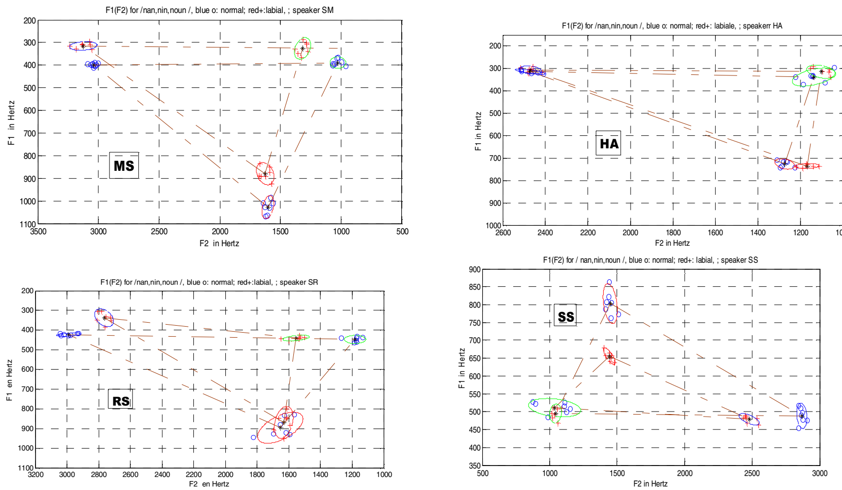

Fig. 6. Dispersion ellipses and vocalic triangle corresponding to vowels [a], [i] and [u] for each speaker HA, SS, RS and MS. In normal condition: in blue color and labial constraint: in red color.

Table 1. Areas of vocalic triangles by speaker
(In normal condition and in labial constraint)

speaker	Area(Hz^2) Normal condition	area(Hz^2) With constraint
HA	259960	288350
SS	282060	110570
RS	411330	263430
MS	639750	502090

Results obtained in Table 1 show a clear vowel reduction caused by the stress for all speakers. This is consistent with results obtained by Lane et al. [13] and other authors who have noticed that the size of the vocalic space was reduced in case of disturbance of speech. However, according Gendrot et al. [12], in case of a good motor control, reduction of the vowel space is carried out according to a centralization of vowels in the vowel space. This would mean that speakers reach theirs acoustics targets (thus remains intelligible despite the constraint).

Figure 7 shows a centralization case. The vowels are represented by their dispersion ellipses (corresponding to several realizations of the same vowel, by the same speaker) in the vocalic space (F2, F1). Results show that in articulatory terms, the vowel [i] become less anterior and more open, while the vowel [a] would be slightly less posterior and a little less open. These changes would influence the formant values: for the [i] vowel, an increase in the formant F1 and a decrease in the formant F2; for the vowel [a], an increased of F2 and decreased of F1. This is called centralization. The ellipses move toward to the center of the vocalic triangle in constraint situation. It would be interesting for us to see if our speakers centralize their vowels.

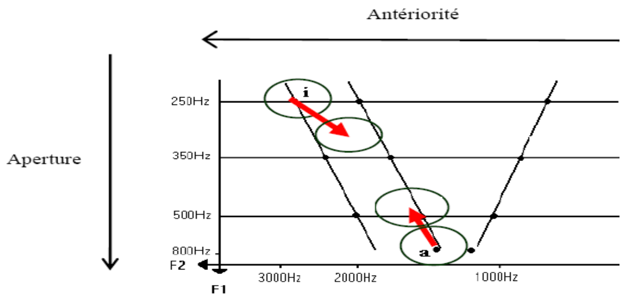

Fig. 7. Case of vowels [i] and centralization in vocalic space [21]

So, we studied the variation direction **of** formants F1 and F2 with the constraint for the three vowels [a], [i] [u], for four speakers. Results obtained showed that most speakers do not centralize. However, centralization is a sign of good motor control. This would mean that the stress is very important and the speakers have their motor control which has decreased. So the acoustic targets are not well met. (This means that the sounds are barely understandable).

7 Labial Constraint Influence Analysis on Intonation F0 and on Mouth Aperture

We have determinate the fundamental frequency and the formant F1 in the middle of the three vowels, in the normal case and in labial constraint, for the seven samples for each speaker. As F1 correspond to the mouth opening and F_0 to the intonation, we chose to represent the intonation according to the aperture to see the behavior of the speaker facing the stress in F_0 (F1). As we have 7 samples per vowel we have represented the F_0 and F1 by their ellipses of dispersion. The results obtained are regrouped in Figures 8.

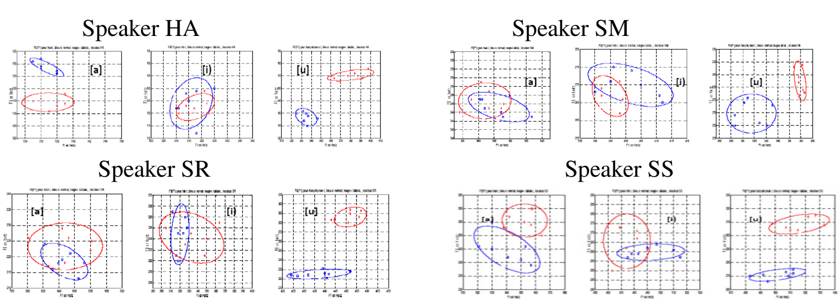

Fig. 8. F_0(F1) variation for each vowel and for each speaker

Figures 4 show in overall a small increase of F0 for vowels [a] and [i] with very little variation of F1 (since the lips are already stretched, the speaker has no difficulties to pronounce these two vowels). However, for the vowel [u], there is a considerable increase in F_0 and F1 simultaneously with the constraint. These results appear right, as the formant frequency F1 is related to the aperture and F_0 for the intonation: so, the speaker seems to make a major effort by increasing its F_0 for trying to reduce its aperture to say the vowel [u], since this one requires almost closing the mouth to be uttered (which justifies the increase in F1 since the constraint prevents mouth from closing). Vowel [u] is then pronounced as a vowel [o], thus increasing the F1).

8 Analysis Effects of the Constraint on the Coarticulation Degree

The coarticulation occurs when a vowel sees its characteristics change because of its phonetic environment. The experiences of disruption of speech have shown that in presence of constraint, the coarticulation increases. This is manifested by a vocalic reduction. An example of coarticulation can be observed when the consonant is followed by a rounded vowel: there is anticipation of the borough of the vowel during the consonant (see Figure 9). What changes acoustically consonant

Spectral characteristics of [k] vary depending on the nature of the following vowel. The main information of coarticulation is carried by the transition between the consonant and the vowel (or vis versa), surrounding the value of the form F2. The F2onset variation (in the beginning of the vowel, in the CV transition) compared to the F2stable (middle of the vowel) is then a source of information for coarticulation. Thus, in type sequences VCV or CVC, for example, the vowel or consonant of the middle is imbued with the surrounding phonemes. The degree of coarticulation is usually estimated by the slope of the straight locus. The concept of the locus equation has been developed by Lindblom in 1963[15]. Lindblom has found that there is a linearity in the second formant frequency variation at limit time of the beginning transition of the consonant to the post-consonantal vowel (F2o or F2onset), according to the second formant frequency target variation of this vowel. Lindblom has established the locus equations from a study of sequences CV.

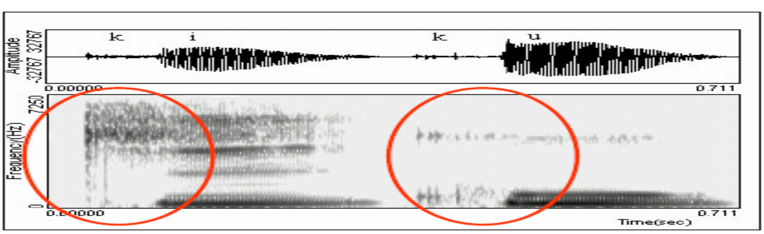

Fig. 9. Anticipation example in the production of CV / ki / and / ku /

The locus equation is written as: F2o = k * F2s + c , where "c" and "k" are constants. In case k = 0, the equation is constant, F2o = c, which would imply that the production of sequence CV is a simple articulatory chaining where the coarticulation between the consonant and the vowel is virtually non-existent. In case where k =1 and *c = 0, the* coarticulation with the vowel context is considered maximal.

In this study, we have plotted the straight lines of locus equations: F2onset(F2satble), with considering seven French vowels, for all CVi (C: consonant [n] and Vi: one of the seven French vowels considered), for all speakers, in normal condition and with constraint, on the same figure. Each vowel is represented by its dispersion ellipse. Results obtained are illustrated by figures 10.

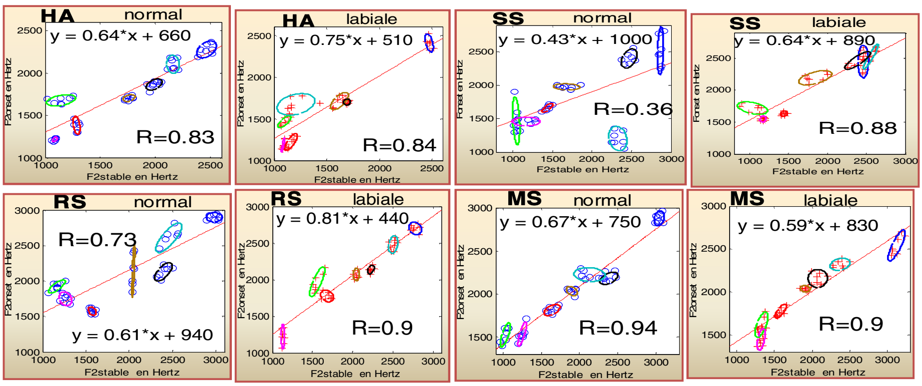

Fig. 10. Straight lines of locus by speaker (in normal condition and with labial constraint)

Locus equation is written: $F_{2onset} = m * F_{2stabl} + b$ where m and b are the slope ande intercept respectively . R is the regression coefficient. The results obtained are summarized in the following table:

Table 2. Parameters of straight lines of locus equations obtained

speakers	Parameters of straight lines of locus equations					
	Without constraint			With constraint		
	m	b	R	m	b	R
HA	0,64	660	0,83	0,75	510	0,84
SS	0,43	1000	0,36	0,64	890	0,88
RS	0,61	940	0,73	0,81	440	0,9
MS	0,67	750	0,95	0,59	830	0,9

This table shows that in labial constraint, we have a net increase in slopes of the straight lines of locus, so of the coarticulation

9 Conclusion

We studied in this work the influence of a labial constraint in speech production. We acoustically analyzed the behavior of the vowels in constraint, then the influence of the constraint on the coarticulation. The results obtained show a great influence of the constraint on the acoustic parameters in particular, the frequential parameters. They can be summarized such as:

- In the temporal field: - temporal increase, therefore the speaker tends to slow down his speech flow for better controlling his elocution rate with the constraint.
- In the frequential field: vocalic spaces are reduced but without vowels centralization. Thus the speakers reach with difficulty their acoustic targets with the constraint or then of the whole.
- In variation of F0(F1) : There is a small increase of F0 for vowels [a] and [i] with very little variation of F1 but for the vowel [u], there is a considerable increase in F0 and F1 simultaneously with the constraint
- With straight lines of locus equations, results shows that we have a net increase of coarticulation with the constraint.

All these results show that, the evaluation methods of the acoustic parameters we propose, accurately characterize the constraint used.

References

[1] Aubin, J.: Effets acoustiques et articulatoires des perturbations labiales sur la parole des enfants et des adultes. Université du Quebec à Montréal (Décembre 2005)
[2] Castellengo, M., Dubois, M.: Timbre ou timbres? Propriété du signal, de l'instrument ou construction cognitive (2005) 88
[3] Chui, C.: An introduction to wavelets. Academic Press (1995)
[4] Cnockaert, L.: Analyse du tremblement vocal et application à des locuteurs parkinsoniens. Thèse de doctorat en sciences de l'ingénieur. Université libre de Bruxelles ULB, Bruxelles (Décembre 2007)
[5] Cohen, L., Lee, C.: Instantaneous frequency and time-frequency distributions. In: IEEE International Symposium on Circuits and Systems, pp. 1231–1234 (1989) 11
[6] Emiya, V.: Spectrogramme d'amplitude et de fréquences instantanées (safi). Master's thesis, Aix-Marseille II (2004) 52
[7] Hirsch, F., Ferbach-Hecker, V., Fauvet, F., Vaxelaire, B.: Étude de la structure formantique des voyelles produites par des locuteurs bègues en vitesses d'élocution normale et rapide.Jep (2006)
[8] Mc Farland, D.H., Baum, S.R., Chabot, C.: Speech compensation to structural modifications of the oral cavity. Journal of the Acoustical Society of America 100(2), 1093–1104 (1996)
[9] Ferbach-Hecker, V.: La résistivité de la qualité des voyelles orales du français. In: SCOLIA, vol. 20, pp. 115–134 (2005)
[10] Flandrin, P.: Temps-fréquence. Hermès (1993) 11, 27, 29, 3.

[11] Jones, J.A., Munhall, K.G.: Learning to produce speech with an altered vocal tract: The role of auditory feedback. Journal of the Acoustical Society of America 113(1), 532–543 (2003)
[12] Gendrot, C., Adda-Deker, M.: Analyses formantiques automatiques de voyelles orales : évidences de la réduction vocalique en langues française et allemande. In: MIDL (2004)
[13] Lane, H., Denny, M., Guenther, F.H., Matthies, M., Ménard, L., Perkell, J.S., Stockmann, E., Tiede, M., Vick, J., Zandipour, M.: Effects of bite blocks and hearing status on vowel production. Journal of the Acoustical Society of America 118(3), 1636–1646 (2005)
[14] Lardiès, J., Ta, M.N., Berthillier, M.: Modal parameter estimation from output-only data using the wavelet transform. Archive of Applied Mechanics 73, 718–733 (2004)
[15] Lindblom, B.: Economy of Speech Gestures. The Production of Speech. Springer, New York (1983)
[16] Mallat, S.: Une exploration des signaux en ondelettes, Editions de l'école Polytechnique (2000)
[17] Navarro, L.: Analyse temps-fréquence de signaux vibratoires issus d'un réacteur de culture osseuse. Journée de la Recherche de l'EDSE (2007a) 143
[18] Rihaczek, A., Bedrosian, E.: Hilbert transforms and the complex representation of real signals. Proceedings of the IEEE 54(3), 434–435 (1966) 11
[19] Torrésani, B.: Analyse continue par ondelettes, Editions du CNRS, Paris (1995)
[20] Ville, J.: Théorie et applications de la notion de signal analytique. Câbles et Transmissions 1, 61–74 (1948) 11,13
[21] Calabrino, A.: Effets acoustiques du débit sur la production de la parole chez des locuteurs enfants et adultes. Actes du Xe Colloque des étudiants en Sciences du Langage 61–83 (2006)
[22] Gendrot, C., Adda-Deker, M.: Analyses formantiques automatiques de voyelles orales: évidences de la réduction vocalique en langues française et allemande. In: MIDL (2004)