



HAL
open science

Low-Frequency Vocal Modulations in Vowels produced by Parkinsonian Subjects

L. Cnockaert, J. Schoentgen, P. Auzou, C. Ozsancak, L. Defebvre, F. Grenez

► **To cite this version:**

L. Cnockaert, J. Schoentgen, P. Auzou, C. Ozsancak, L. Defebvre, et al.. Low-Frequency Vocal Modulations in Vowels produced by Parkinsonian Subjects. *Speech Communication*, 2008, 50 (4), pp.288. 10.1016/j.specom.2007.10.003 . hal-00499199

HAL Id: hal-00499199

<https://hal.science/hal-00499199>

Submitted on 9 Jul 2010

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

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

Accepted Manuscript

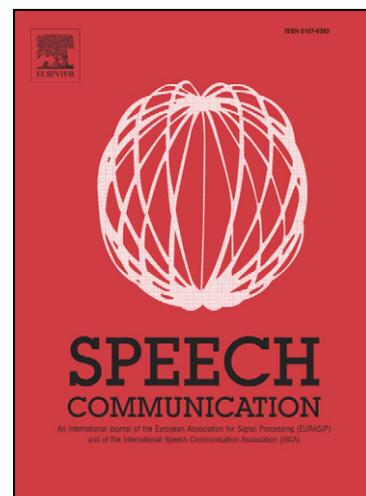
Low-Frequency Vocal Modulations in Vowels produced by Parkinsonian Subjects

L. Cnockaert, J. Schoentgen, P. Auzou, C. Ozsancak, L. Defebvre, F. Grenz

PII: S0167-6393(07)00174-4
DOI: [10.1016/j.specom.2007.10.003](https://doi.org/10.1016/j.specom.2007.10.003)
Reference: SPECOM 1671

To appear in: *Speech Communication*

Received Date: 30 January 2007
Revised Date: 15 September 2007
Accepted Date: 22 October 2007



Please cite this article as: Cnockaert, L., Schoentgen, J., Auzou, P., Ozsancak, C., Defebvre, L., Grenz, F., Low-Frequency Vocal Modulations in Vowels produced by Parkinsonian Subjects, *Speech Communication* (2007), doi: [10.1016/j.specom.2007.10.003](https://doi.org/10.1016/j.specom.2007.10.003)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Low-Frequency Vocal Modulations in Vowels produced by Parkinsonian Subjects

L. Cnockaert^{a,*1}, J. Schoentgen^{a,2}, P. Auzou^{b,c},
C. Ozsancak^{b,d}, L. Defebvre^b, F. Grenez^a

^a*Laboratoire d'Images, Signaux et Dispositifs de Télécommunications, Faculté des Sciences Appliquées, Université Libre de Bruxelles, CP 165/51, av. F.D. Roosevelt 50, 1050 Brussels, Belgium*

^b*Service de Neurologie A et Pathologie du Mouvement, CHRU de Lille, Faculté de médecine H. Warenbourg, EA 6283 IFR 114, France*

^c*Service d'Explorations Fonctionnelles Neurologiques, Etablissement Hélios Marin - Groupe Hopale, rue du Dr Calot 47, 62600 Berck sur Mer, France*

^d*Service de Neurologie, CH de Boulogne sur Mer, 62200 Boulogne sur Mer, France*

Abstract

Low-frequency vocal modulations here designate slow disturbances of the phonatory frequency F_0 . They are present in all voiced speech sounds, but their properties may be affected by neurological disease. An analysis method, based on continuous wavelet transforms, is proposed to extract the phonatory frequency trace and low-frequency vocal modulation in sustained speech sounds. The method is used to analyze a corpus of vowels uttered by male and female speakers, some of whom are healthy and some of whom suffer from Parkinson's disease. The latter present general speech problems but their voice is not perceived as tremulous. The objective is to discover differences between speaker groups in F_0 low-frequency modulations. Results show that Parkinson's disease has different effects on the voice of male and female speakers. The average phonatory frequency is significantly higher for male parkinsonian speakers. The modulation amplitude is significantly higher for female parkinsonian speakers. The modulation frequency is significantly higher and the ratio between the modulation energies in the frequency-bands $[3Hz, 7Hz]$ and $[7Hz, 15Hz]$ is significantly lower for parkinsonian speakers of both genders.

Key words: speech analysis, vocal modulations, parkinsonian speech

31 The aim of this article is to propose an analysis of low-frequency modula-
32 tions of the phonatory frequency F_0 in parkinsonian and control speakers.
33 Low-frequency modulations designates slow disturbances of the phonatory fre-
34 quency, which is the frequency of the voice source signal generated by the pul-
35 satile air flow at the glottis. In this article, properties of the vocal disturbances
36 of healthy speakers and speakers with Parkinson's disease, whose voices are
37 not perceived as tremulous, are compared. The properties that are considered
38 are the *modulation amplitude* (MA), *modulation frequency* (MF) and *modu-*
39 *lation energy ratio* (MER). The low-frequency modulations of F_0 described
40 here differ from pathological narrow-band vocal tremor that is occasionally
41 observed in speakers suffering from neurological disorders.

42 Cycle lengths in voiced speech sounds vary stochastically. Conventionally, one
43 distinguishes perturbations of the cycle lengths due to vocal jitter from per-
44 turbations due to vocal tremor (Titze, 1995; Schoentgen, 2002). Jitter refers
45 to short-term cycle-to-cycle perturbation of the duration of the speech cy-
46 cles (Titze, 1994). Vocal tremor is a slow supra-cycle fluctuation, character-
47 ized by frequencies less than 15 Hz. For some authors, the word *tremor* has a
48 connotation of narrow-band modulation at a single frequency (Yair and Gath,
49 1988; Holmes et al., 2000). Therefore, the expression *low-frequency modulation*
50 is preferred here to describe F_0 fluctuations between 3 Hz and 15 Hz.

51 Tremor of the limbs is defined as involuntary, rhythmic movements that are
52 seen better distally as in the fingers of hands but may sometimes be seen in the
53 eyelids, tongue, face or other body parts (Fucci and Petrosino, 1984). Freund
54 (1987) distinguishes between physiological and pathological tremor. Physi-
55 ological tremor accompanies any muscle activity. Major factors determining
56 physiological tremor are rhythmic changes due to neurally or mechanically de-
57 termined oscillations, pulsatile blood flow, and breathing. Physiological tremor
58 is therefore not expected to present a single modulation peak. On the other
59 hand, rhythmic pathological tremor is characterized by strong motor unit
60 synchronization. This may lead to large-amplitude narrow-band tremor that
61 becomes disturbing to the patient and interferes with limb motion.

* Corresponding author. Tel:+32-2-6503089, fax:+32-2-6504713

Email addresses: lcnockae@ulb.ac.be (L. Cnockaert), jschoent@ulb.ac.be
(J. Schoentgen), pauzou@yahoo.fr (P. Auzou), c_ozsancak@yahoo.fr (C.
Ozsancak), fgrenez@ulb.ac.be (F. Grenez).

¹ L. Cnockaert is a fellow with the *Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture* (Belgium).

² J. Schoentgen is a Senior Research Associate with the *Fonds National de la Recherche Scientifique* (Belgium).

62 The previous paragraph refers to tremor of the limbs or other body parts.
63 The low-frequency modulation of the phonatory frequency is a property of the
64 cycle lengths of voiced speech sounds. The production of voiced speech sounds
65 involves the vibration of the vocal folds, which is controlled via laryngeal
66 and respiratory muscles. One may assume that physiological tremor causes
67 low frequency modulations of F_0 , which has been called vocal microtremor in
68 normophonic speakers (Schoentgen, 2002). Given that part of the physiological
69 tremor has neurological causes, speakers with neurological diseases who do
70 not present marked narrow-band vocal tremor may still present low-frequency
71 voice modulations that differ from those of healthy speakers.

72 Parkinson's disease is a progressive, degenerative disease of the central nervous
73 system resulting in rigidity, resting ([3 Hz, 7 Hz]) and posture ([8 Hz, 12 Hz])
74 tremor, and reduced range of movement of the limbs, neck, and head. Speech
75 and swallowing difficulties are frequent: Hartelius and Svensson (1994) have
76 observed that 70% of the patients have experienced impairment of speech and
77 voice after the onset of the disease. Speech disorders have been regarded as one
78 of their greatest problems by 29% of the patients. Laryngeal dysfunctions have
79 also been observed (Logemann et al., 1978; Hanson et al., 1984). The effect of
80 the disease on the laryngeal muscle may reduce the patient's ability to initiate
81 phonation, produce adequate loudness, and vary pitch. Perez et al. (1996) have
82 observed that a majority of patients exhibit tremor of the larynx at rest, at
83 normal pitch or in loud voice. Audible symptoms are breathiness, roughness,
84 hoarseness, and tremulous voice. Narrow-band tremor may be heard in the
85 voice as well as seen in the extremities (Logemann et al., 1978). Holmes et al.
86 (2000) have observed narrow-band tremor only for patients with advanced
87 Parkinson's disease. Ziegler and Hoole (1999) have mentioned gender-specific
88 voice dysfunction in Parkinson's disease, the male speakers showing predom-
89 inantly increased pitch and breathiness, and the female speakers showing a
90 quivering voice with a strained quality.

91 Some studies have investigated the acoustic characteristics of the voices of
92 patients with Parkinson's disease (Kent et al., 1994; King et al., 1994; Zwirner
93 and Barnes, 1992). Most have reported an average phonatory frequency within
94 the normal range, but a larger than normal variation of phonatory frequency
95 in sustained vowels (Zwirner and Barnes, 1992), which may reflect a loss of
96 control of motor activity. Hirose et al. (1995) have studied the voice quality of
97 patients with neurological disorders, including Parkinson's disease. They have
98 observed a higher variability in the phonatory frequency F_0 in parkinsonian
99 speakers, compared to control speakers, as well as for fast [16 Hz, $F_0/2$] as for
100 slow fluctuations [0.1 Hz, 16 Hz].

101 Few studies have specifically described the characteristics of vocal tremor or
102 low-frequency modulation of F_0 for tremor patients or normophonic speak-
103 ers. Schoentgen (2002) has studied vocal microtremor in normophonic and

104 mildly hoarse speakers, for vowels [a], [i] and [u]. He has calculated two mod-
105 ulation level and two modulation frequency cues. Winholtz and Ramig (1992)
106 have studied vocal tremor in speakers with normal voice, vibrato and patholog-
107 ical tremor, by means of modulation frequency and modulation level cues. Yair
108 and Gath (1988) have reported data for vowels [a] for nine parkinsonian and
109 three control speakers. They have reported rhythmic variations between 4 Hz
110 and 6 Hz in the phonatory frequency of the parkinsonian speakers, correspond-
111 ing to sharp peaks in the phonatory frequency spectra. Yair and Gath (1988)
112 have also reported that for normal speakers, most of the tremor energy has
113 been concentrated below 3 Hz. At higher frequencies the energy has been fee-
114 ble and evenly spread throughout the frequency interval. Data published by
115 these authors are discussed later and compared to the results obtained in this
116 study.

117 The present work presents an analysis of the modulation amplitude, modula-
118 tion frequency and ratio of the modulation energy in low and high frequency
119 intervals. These cues are examined for speakers with Parkinson's disease as
120 well as healthy control speakers. These studies are motivated by the following
121 reasons.

122 First, to the authors' knowledge, data are scarce with regard to the low-
123 frequency modulation of F_0 in normophonic speakers, and in speakers with
124 Parkinson's disease or other neurological disorders. Parkinson's disease is the
125 neurological disease that is the most common (Defebvre, 2005). Data that
126 may contribute to improving the patients' voice quality are therefore worth
127 recording. Also, one may wish to investigate whether acoustic cues exist that
128 distinguish normophonic control and parkinsonian speakers.

129 Second, existing studies about Parkinson's disease (Yair and Gath, 1988; Win-
130 holtz and Ramig, 1992; Hirose et al., 1995) are difficult to compare. Indeed,
131 most data pertain to speakers who are under treatment. Treatment of the dis-
132 ease as well as severity of the vocal symptoms are expected to vary between
133 studies. Hence, the availability of additional data is conducive to the discovery
134 of general rules about the effects of the disease.

135 Third, the acoustic assessment of low-frequency modulations of F_0 is not
136 well documented. Data about low-frequency vocal modulations have been
137 previously obtained by demodulation (Winholtz and Ramig, 1992), or by
138 analysing point processes (Yair and Gath, 1988) or vocal cycle length time
139 series (Schoentgen, 2002). Here, an alternative is investigated that is based on
140 continuous wavelet transforms. It enables tracking instantaneously the low-
141 frequency disturbances of the phonatory frequency.

142 To analyse low-frequency vocal modulation, the phonatory frequency estima-
143 tor must be able to track small frequency perturbations and handle disordered

144 speech signals. Algorithms that estimate the phonatory frequency (Hess, 1983;
145 Mitev and Hadjitodorov, 2003) fall into different categories, which involve the
146 measurement of the length of each vocal cycle (Kadambe and Boudreaux-
147 Bartels, 1992; Schoentgen, 2002), the estimation of the average period in an
148 analysis frame (Medan et al., 1991; Boersma and Weenink, 2004), or the esti-
149 mation of the instantaneous frequency of the fundamental spectral component
150 of the speech signal (Winholtz and Ramig, 1992). The last category presents
151 advantages when tracking small frequency perturbations: Firstly, the phona-
152 tory frequency must not be considered stationary over an analysis frame. Sec-
153 ondly, the instantaneous phonatory frequency trace is easier to handle math-
154 ematically than the unevenly sampled cycle duration time series (Schoentgen,
155 2002), or cycle event point processes (Yair and Gath, 1988).

156 In the method that is presented here, the phonatory frequency is assigned to
157 the instantaneous frequency, which is defined as the rate of change of the phase
158 of the estimated fundamental of the speech signal. Existing methods differ in
159 the way the phase trace is obtained. The conventional method consists in low-
160 pass-filtering the speech signal in the vicinity of the fundamental and using
161 the phase of the analytical signal (Winholtz and Ramig, 1992). This requires,
162 however, a prior estimate of the phonatory frequency and the assumption that
163 variations around this estimate are small. Therefore, more than one phase trace
164 is computed, and the phonatory frequency is chosen for each time-step, based
165 on criteria such as the following. Qiu et al. (1995) perform Hilbert transforms
166 on the speech signal filtered by two different low-pass filters. Nakatani and
167 Irino (2004) employ the short-time Fourier-transform phase and dominant
168 harmonic components. Kawahara et al. (1999) exploit a continuous wavelet
169 transform with an analytical wavelet, and detect the fixed points in the wavelet
170 central frequency to instantaneous frequency map.

171 In this article, a method is presented that obtains the instantaneous frequency
172 via two continuous wavelet transforms. The first guaranties reliability and the
173 second sensitivity to frequency perturbations. The adequacy of this technique
174 for the estimation of low-frequency vocal modulations is investigated by com-
175 parison with one analysis frame-based method (Boersma and Weenink, 2004)
176 and two instantaneous frequency-based methods (Winholtz and Ramig, 1992;
177 Kawahara et al., 1999).

178 The size and characteristic frequency of the modulation are summarized by
179 means of three cues, which are the modulation amplitude, the modulation
180 frequency and the ratio of the modulation energies in low [3 Hz, 7 Hz] and
181 high frequency [7 Hz, 15 Hz] intervals.

182 The lower frequency limit of the analyses is set to 3 Hz. The aim is to discard
183 effects of the heart beat, which is expected in the vicinity of 1-2 Hz (Orlikoff
184 and Baken, 1989). Moreover, variations slower than 3 Hz are difficult to esti-

185 mate because their cycle length becomes long compared to the signal length.
186 This low-frequency limit is considered acceptable, inasmuch as the effects of
187 Parkinson's disease, such as resting and posture tremor, are typically observed
188 at frequencies higher than 3 Hz.

189 Statistical analyses are carried out to investigate whether significant differ-
190 ences existed between frequency modulation cues of parkinsonian and control
191 and male and female speakers. Discrimination analyses are carried out to test
192 whether a separation between parkinsonian and control speakers is possible
193 on the base of selected cues.

194 This work is organized as follows: Section 2 presents the F_0 estimation and
195 the acoustic cues of low-frequency vocal modulation. In Section 3, the F_0 es-
196 timation method is evaluated on synthetic and real data, and compared to
197 existing F_0 estimation methods. In Section 4 the low-frequency vocal modu-
198 lation analysis is applied to recorded vowels. Section 4.1 presents the corpora
199 and statistical methods used in this work. Section 4.2 gives the results and
200 their statistical analyses, which are then discussed in Section 4.3.

201 2 Methods

202 Here we present the F_0 estimation and the acoustic cues indicating low-
203 frequency vocal modulation.

204 2.1 Estimation of F_0

205 Estimation of F_0 is based on the instantaneous frequency obtained via a *con-*
206 *tinuous wavelet transform* (CWT) (Addison, 2002). In a first stage, a CWT
207 of the speech signal is computed and an approximate F_0 estimate is obtained
208 via the central frequency of the wavelet with maximal CWT modulus. In a
209 second stage, another CWT is carried out, using a shorter mother wavelet.
210 The F_0 estimate is assigned to the instantaneous frequency corresponding to
211 the wavelet whose central frequency is equal to the F_0 estimate obtained dur-
212 ing the first stage. The first stage enables an easy detection of the maximum
213 in the CWT modulus and the second stage provides a precise instantaneous
214 frequency.

215 The evaluation of this F_0 estimation method is carried out by means of syn-
216 thetic vowels [a], generated via a source-filter model. The vocal tract is sim-
217 ulated by an all-pole filter (Rabiner and Schafer, 1978) and the source signal
218 by means of Fant's model (Fant et al., 1985). The phonatory frequency is

219 75 Hz, and the four formant frequencies and bandwidths are 700 Hz, 1200 Hz,
 220 2500 Hz and 3500 Hz, and 140 Hz, 180 Hz, 55 Hz and 200 Hz respectively.
 221 The source parameters are the following : *relative open phase* = 0.4, *reciprocal*
 222 *of the negative peak value* = 0.1, and *closing phase / open phase* = 0.2. Fig.
 223 2 shows the time-evolving modulus of a CWT with parameter $\omega_c \sigma_t = 5$ for
 224 a synthetic vowel. A ridge of high modulus is seen at 75 Hz that is the syn-
 225 thetic phonatory frequency. During the CWT calculation, the wavelet central
 226 frequency is increased in steps of 1 Hz in the interval [50 Hz, 200 Hz].

227 The instantaneous frequency $IF(t)$ of a band-pass signal $s(t)$ may be defined
 228 by means of its Hilbert transform $H[s(t)]$ and associated analytical signal
 229 $s_a(t)$ (Boashash, 1992).

$$s_a(t) = s(t) + jH[s(t)], \quad (1)$$

$$IF(t) = \frac{d(\arg[s_a(t)])}{dt}. \quad (2)$$

230 The instantaneous frequency can also be defined by means of a continuous
 231 wavelet transform, using an analytical wavelet (Le-Tien, 1997). The CWT of
 232 a signal $x(t)$ and the IF are defined by

$$CWT(\lambda, t) = \int_{-\infty}^{+\infty} x(u) \frac{1}{\sqrt{\lambda}} \psi^* \left(\frac{u-t}{\lambda} \right) du, \quad (3)$$

$$IF(\lambda, t) = \frac{d(\arg[CWT(\lambda, t)])}{dt}. \quad (4)$$

233 In (3) the mother wavelet is denoted by $\psi(t)$ and $CWT(\lambda, t)$ denotes the
 234 wavelet transform at time t and scale λ . The CWT coefficients (3) are com-
 235 plex. Here, the mother wavelet is the complex Morlet wavelet (Percival and
 236 Walden, 2000) shown in Fig. 1, which can be expressed as

$$\psi_{\omega_c}(t) = C e^{-i\omega_c t} \left[e^{-\frac{t^2}{2\sigma_t^2}} - \sqrt{2} e^{-\frac{\omega_c^2 \sigma_t^2}{4}} e^{-\frac{t^2}{\sigma_t^2}} \right]. \quad (5)$$

The scale of the wavelet is fixed by central frequency $f_c = \frac{\omega_c}{2\pi}$, which is the
 frequency of oscillation of the wavelet. Parameter σ_t fixes its decay. The prod-
 uct $\omega_c \sigma_t$ is constant for a wavelet family. The normalization factor C is chosen
 such that

$$\int_{-\infty}^{+\infty} |\psi_{\omega_c}(t)|^2 dt = 1.$$

237 The effective duration of the wavelet is conventionally defined as $2\sigma_t$.

238 The Gaussian envelope of the complex Morlet wavelet minimizes the product
 239 of the wavelet's time and frequency spreads, and therefore optimizes the time-
 240 frequency resolution (Addison, 2002). The modulus and phase of the CWT
 241 estimate the envelope and instantaneous phase of the spectral components
 242 of the signal in the band centred on frequency f_c (Mallat, 1999). The time-
 243 derivative of the phase of the complex CWT is therefore an estimate of the
 244 instantaneous frequency of the signal in that band.

245 The possibility of estimating the phonatory frequency via CWTs is based on
 246 the following observation. In the neighbourhood of the wavelet central frequen-
 247 cies that best fit the cyclicity of the signal, the modulus of the complex CWT
 248 attains a maximum and the instantaneous frequency obtained by means of the
 249 phase of the CWT is close to the cyclicity of the signal. In the (f_c, IF) plane,
 250 the instantaneous frequency presents a plateau, in the vicinity of the actual
 251 value, over a large central frequency band (Carmona et al., 1997). This is illus-
 252 trated in Fig. 3, which shows as a function of the wavelet central frequency for
 253 time t the modulus and instantaneous frequency of the wavelet transform for
 254 the synthetic vowel in Fig. 2. For the wavelet central frequencies for which the
 255 CWT moduli are high, the IF displays a plateau at the phonatory frequency
 256 of the speech signal.

257 We thus propose to estimate F_0 via the IF of the CWT whose modulus is at
 258 a maximum in the interval [50 Hz, 500 Hz] (Cnockaert et al., 2005). When
 259 the plateau is flat, F_0 is easily obtained, even when the CWT is calculated
 260 with a large frequency step. The CWT using the wavelet family obtained by
 261 $\omega_c \sigma_t = 5$ is adequate to detect the position of the maximum of the modulus
 262 corresponding to the phonatory frequency.

263 For present purposes, variations in F_0 up to 15 Hz must be detected correctly.
 264 Earlier studies show that the amplitude of the modulation of the phonatory
 265 frequency may be underestimated by the method described above. Indeed,
 266 when the time resolution of the wavelet is not high enough, the IF trace is
 267 smoothed over the effective duration of the wavelet. This problem is all the
 268 more difficult the lower F_0 and the faster its variation. For example, for a
 269 wavelet with $\omega_c \sigma_t = 5$, the effective duration $2\sigma_t$ is equal to 16ms when the
 270 central frequency is 100 Hz. Small F_0 variations at a frequency above 10 Hz
 271 are therefore smoothed over the wavelet duration and attenuated.

272 A mother wavelet of shorter effective duration, and thus smaller $\omega_c \sigma_t$ value,
 273 is therefore expected to be better suited for tracking short-term variations of
 274 F_0 . Fig. 4 shows the modulus of the wavelet transform, for $\omega_c \sigma_t = 2.5$, of the
 275 previous synthetic vowel, and Fig. 5 shows the modulus and IF at two posi-
 276 tions in the vocal cycle. One sees that the plateau where the IF is equal to
 277 the phonatory frequency is shifted with regard to the wavelet central frequen-
 278 cies for which the CWT moduli are high. Also, one observes that the plateau

279 is shorter on real speech signals. It could thus still be used to estimate F_0 .
 280 Locating the corresponding modulus peak, however, has become more diffi-
 281 cult. Indeed, owing to the larger bandwidth and shorter effective length, the
 282 wavelet transform is influenced more by the second harmonic and intra-cycle
 283 instationarities, such as the instants of glottal closure. For some positions in
 284 the speech cycle, the modulus peak at the phonatory frequency may disappear,
 285 as illustrated by the dashed line in Fig. 5.

286 As expected, a longer wavelet provides a more reliable F_0 estimate, whereas
 287 a shorter wavelet is more sensitive to F_0 variations. Two wavelet transforms
 288 are therefore combined: one with a high frequency resolution to identify the
 289 maximum of the CWT moduli and a second with a high time resolution to
 290 estimate IF. As illustrated in Fig. 6, this procedure functions as follows:

- 291 (1) CWT of the speech signal, with $\omega_c \sigma_t = 5$, and recording, for each time-
 292 step, of the wavelet central frequency \hat{f}_c corresponding to the maximal
 293 modulus of the CWT.
- 294 (2) CWT of the speech signal, with $\omega_c \sigma_t = 2.5$, and calculation of the CWT
 295 instantaneous frequency $IF_{2.5}$ via the time-derivative of the CWT phase.
- 296 (3) Estimation of the instantaneous F_0 value via $IF_{2.5}$ of the wavelet whose
 297 central frequency \hat{f}_c is equal to the one recorded at the first stage.
- 298 (4) Filtering of the F_0 trace to eliminate residual oscillations owing to intra-
 299 cycle instationarities.

300 2.2 Acoustic Cues of Low-Frequency Vocal Modulation

301 In this work, three acoustic cues are considered: The modulation amplitude
 302 characterizes the size of the modulation compared to the average phonatory
 303 frequency. The modulation frequency and modulation energy ratio charac-
 304 terize the distribution of the modulation energy in the frequency interval,
 305 respectively by means of the centroid of the modulation spectrum and the
 306 ratio of the modulation energy at low and high frequencies.

307 To calculate the modulation cues, a continuous wavelet transform of the F_0
 308 trace is carried out, using the complex Morlet wavelet with $\omega_c \sigma_t = 5$. Before
 309 transforming the F_0 trace, the trend, which designates slow changes of F_0
 310 below the 3 Hz limit, is removed, using a method described by Yair and Gath
 311 (1988). The wavelet central frequency is increased by steps of 0.1 Hz in the
 312 interval [3 Hz, 15 Hz]. The CWT of the IF trace enables estimates of the
 313 modulation amplitude, frequency, and energy ratio to be obtained.

314 2.2.1 Modulation Amplitude

315 The instantaneous modulation energy is obtained by summing the square of
 316 the modulus of the wavelet transform over the frequency interval $[f_{min}, f_{max}]$.
 317 The square root, normalized by average \bar{F}_0 , gives an estimate of the relative
 318 instantaneous modulation amplitude,

$$MA(t) = \frac{\sqrt{\sum_{f_c=f_{min}}^{f_{max}} CWT^2(2\pi f_c, t)}}{\text{average}(F_0)}. \quad (6)$$

319 Here CWT denotes the wavelet transform of the trace $F_0(t)$. Parameters f_{min}
 320 and f_{max} are equal to 3 Hz and 15 Hz.

321 2.2.2 Modulation Frequency

322 One wishes to characterize the modulation frequency by means of a single
 323 value. The modulation frequency is therefore defined as the centroid of the
 324 modulation spectrum. The centroid is the sum, over the frequency interval
 325 $[f_{min}, f_{max}]$, of the instantaneous frequencies $IF(2\pi f_c, t)$ of the CWT of the
 326 F_0 trace, weighted by the wavelet transform energy. The weight is set to zero
 327 when the CWT modulus is lower than 1% of the maximal modulus. The
 328 instantaneous modulation frequency is thus defined as follows,

$$MF(t) = \frac{\sum_{f_c=f_{min}}^{f_{max}} [CWT^2(2\pi f_c, t)IF(2\pi f_c, t)]}{\sum_{f_c=f_{min}}^{f_{max}} [CWT^2(2\pi f_c, t)]}. \quad (7)$$

329 2.2.3 Modulation Energy Ratio

330 This cue is inspired by the observation of the spectral energy distribution of
 331 F_0 traces, which are different for parkinsonian and control speakers. The *mod-*
 332 *ulation energy ratio* (MER) in the frequency bands $[f_{min}, f_{mid}]$ and $[f_{mid}, f_{max}]$
 333 is calculated according to

$$MER(t) = \frac{\sum_{f_c=f_{min}}^{f_{mid}} CWT^2(2\pi f_c, t)}{\sum_{f_c=f_{mid}}^{f_{max}} CWT^2(2\pi f_c, t)}. \quad (8)$$

334 The value of the middle frequency f_{mid} has been fixed by observing the ratios
 335 obtained for parkinsonian and control speakers, with f_{mid} in the range 5 Hz
 336 to 10 Hz. The f_{mid} value that separates best both groups has been found to
 337 be 7 Hz.

339 This section reports the evaluation of the proposed F_0 estimation method on
 340 synthetic vowels, the comparison on synthetic vowels with three F_0 estimation
 341 methods, and the comparison on real vowels of the two best methods.

342 3.1 Evaluation on sustained synthetic vowels

343 The accuracy of the F_0 extraction is evaluated by means of synthetic vowels
 344 with modulated F_0 as described in Section 2.1. The synthetic fundamental
 345 frequency is modulated by a cosine such that

$$f_{\text{source}}(t) = f_{\text{av}} \cdot [1 + A_{T,\text{ref}} \cdot \cos(2\pi F_{T,\text{ref}} t)], \quad (9)$$

346 where f_{source} , f_{av} , $A_{T,\text{ref}}$ and $F_{T,\text{ref}}$ are the instantaneous phonatory frequency,
 347 the average phonatory frequency, the reference modulation amplitude and the
 348 reference modulation frequency, respectively.

349 The modulation frequency and amplitude are estimated by recording the fre-
 350 quency and peak-to-peak amplitude of the sinusoidal variation of the F_0 trace.
 351 The performance with regard to modulation amplitude tracking is numerically
 352 expressed by the ratio of the estimated and reference modulation amplitudes.
 353 This ratio characterizes the ability of the method to record the modulation of
 354 the phonatory frequency of the synthetic signal, and ideally is equal to one.

355 For the proposed CWT-based method, the modulation frequency is detected
 356 correctly, whereas the modulation amplitude is underestimated. This under-
 357 estimation depends on average F_0 and modulation frequency, but not on the
 358 reference modulation amplitude. The estimated-to-reference amplitude ratio
 359 is shown in Fig. 7, as a function of the modulation frequency, for synthetic
 360 vowels [a] with various average phonatory frequencies F_0 , and with a refer-
 361 ence modulation amplitude of 1%. The ratio decreases when the modulation
 362 frequency increases and the average F_0 decreases. The reason for this effect is
 363 that when the modulation frequency gets closer to the phonatory frequency,
 364 one perturbation period extends over a few vocal cycles only. The detection
 365 of fast F_0 variations is more difficult, because of the smoothing of the CWT
 366 over the effective duration of the wavelet.

367 The CWT-based F_0 estimation is compared to three other F_0 analysis meth-
 368 ods:

- 369 • A Hilbert transform-based method (Winholtz and Ramig, 1992). The speech
 370 signal is low-pass filtered above the phonatory frequency, the value of which

371 must be estimated first. The IF trace, obtained via the associated analytical
 372 signal, is low-pass-filtered at 25 Hz to obtain a smooth estimate of the time-
 373 evolution of F_0 .

- 374 • The *TEMPO* method (Kawahara et al., 1999), which is founded on a fixed-
 375 point analysis of a wavelet central frequency to IF mapping and on carrier-
 376 to-noise ratios.
- 377 • The CC-method in PRAAT (Boersma and Weenink, 2004), which is based
 378 on a forward cross-correlation analysis. The minimum phonatory frequency
 379 is fixed to 75 Hz.

380 Figure 8 shows the estimated-to-reference modulation amplitude ratio as a
 381 function of the modulation frequency for the four F_0 extraction methods syn-
 382 thetic vowels [a] with \bar{F}_0 equal to 100 Hz and 300 Hz. The modulation am-
 383 plitude is underestimated, except for the Hilbert transform-based method. In
 384 the case of the latter, a feeble attenuation is observed, which depends on the
 385 low-pass filter cut-off. For the other three methods, the underestimation of the
 386 modulation amplitude increases with the modulation frequency and decreases
 387 with average F_0 . The reason is the same as the one described above for the
 388 CWT-based method. Figure 8 suggests that the CWT-based method reports
 389 low-frequency F_0 variation more accurately than TEMPO and PRAAT. Rapid
 390 variations are underestimated, but this underestimation is small in the rele-
 391 vant frequency range, which is 3-15 Hz. For \bar{F}_0 equal to 100 Hz, the attenua-
 392 tion of a 15 Hz variation is less than 7.5%.

393 3.2 Evaluation on disordered speech

394 The application of F_0 modulation analysis is expected to be clinical mainly.
 395 The analysis must therefore be reliable for elderly and dysphonic speakers.
 396 The CWT-based method is preferred to the Hilbert-based method, for the
 397 following reasons.

398 F_0 traces obtained via CWT-based and Hilbert transform-based methods are
 399 quasi-identical for real speech signals, with a slightly better detection of high
 400 frequency perturbations with the second method, which has the following
 401 drawbacks, however.

402 First, it needs a prior estimation of average \bar{F}_0 . It relies thus on another F_0
 403 extraction algorithm, which must correctly handle disordered speech signals.

404 Second, it requests low-pass filtering that must correctly isolate the fundamen-
 405 tal from the harmonics as well as track F_0 trends and intonation. A trade-off
 406 exists between the selectivity of the filter, which implies a long impulse re-
 407 sponse, and its ability to track large F_0 variations, for which a short response
 408 would be more appropriate. Because of this trade-off, it may happen that the

409 F_0 trace obtained via the Hilbert transform becomes erroneous. An example is
410 shown in Figures 9 and 10, for a real speech uttered by a 64-year-old speaker
411 with diplophonia. Diplophonia is observed in Fig. 9 to occur during the time-
412 intervals 11 s – 12.55 s and 12.7 s – 13.6 s. The amplitude of the F_0 oscillations
413 due to diplophonia is small because the F_0 traces are smoothed by the final
414 low-pass filter. At times 12.3 s, 12.9 s and 13.2 s, the F_0 trace obtained via
415 the Hilbert transform presents aberrant peaks. These failures are explained by
416 the subharmonics at multiples of $F_0/2$ (Fig.9). Ideally subharmonics at $F_0/2$
417 and $3F_0/2$ should be eliminated by filtering, without altering the side-bands
418 of the fundamental at F_0 , which inform about the low-frequency modulation.
419 Both requests are difficult to satisfy simultaneously.

420 In the CWT-based method, the filtering is adapted instantaneously depending
421 on F_0 . This method is therefore retained.

422 4 Analysis of Sustained Vowels Uttered by Parkinsonian and Con- 423 trol Speakers

424 Here we begin our analysis of the effects of Parkinson’s disease on the articu-
425 lation of sustained vowels.

426 4.1 Corpora and Statistical Methods

427 The corpus is composed of speakers with Parkinson’s disease and control
428 speakers. The utterances are sustained vowel segments [a] obtained in the
429 framework of a maximum phonation time task.

430 The participants are 37 French-speaking subjects (28 males and 9 females)
431 who have Parkinson’s disease and have reported general speech problems, but
432 whose voices are not perceived as tremulous, and 35 French-speaking control
433 subjects (28 males and 7 females), who do not report any laryngeal problems.
434 The recordings of the parkinsonian speakers are made during treatment, by
435 implant or medication. The average intelligibility scores (Auzou et al., 1998)
436 are 17 and 19 for the male and female parkinsonian speakers. The intelligibility
437 score ranges from 0 (very disturbed speech) to 24 (normal speech). The age-
438 ranges are 44 to 75 and 42 to 75 years for the male, and 49 to 73 years for the
439 female parkinsonian and control speakers. The average maximum phonation
440 times are 11s and 19s for the male, and 11s and 14s for the female parkinsonian
441 and control speakers.

442 The recordings are carried out on EVA stations (SQLab, 2005) in the *Service*

443 *de Neurologie* of the *Centre Hospitalier Régional Universitaire de Lille*, or of
444 the *Centre Hospitalier Universitaire de Rouen*, both in France. The sampling
445 rate is 25 kHz for 67 speakers and 6.25 kHz for 5 speakers. A sampling rate
446 of 6.25 kHz is large enough for estimating low-frequency modulations of the
447 phonatory frequency. Including the recordings at 6.25 kHz enables analysing
448 more signals, and improving the reliability of the statistical analyses.

449 The cues that are analysed statistically are the average phonatory frequency,
450 average modulation amplitude, average modulation frequency and average
451 modulation energy ratio obtained for 5 sec-long signal fragments at the begin-
452 ning of each recording excluding onsets.

453 Two-factor analyses of variances are carried out for each cue, the factors being
454 gender and health status of the speakers. The null hypotheses are that the
455 means are the same for control and parkinsonian, as well as male and female
456 speakers. Subsequently, discriminant analyses are carried out, to test whether
457 a separation between parkinsonian and control speakers is possible on the basis
458 of selected cues. Three cues only (phonatory frequency, modulation amplitude
459 and modulation frequency) are used, because for a discriminant analysis, the
460 independent variables should not be strongly correlated.

461 4.2 Results

462 This subsection illustrates the instantaneous modulation cues for two exam-
463 ples: one parkinsonian and one control speaker. Fig. 11 shows the F_0 trace,
464 the CWT^2 coefficients, the modulation amplitude, frequency, and energy ra-
465 tio, for a control and a parkinsonian speaker. One sees that the F_0 of the
466 parkinsonian speaker presents stronger modulation and the cue values vary
467 rapidly. This observed lack of short-time stability has motivated averaging
468 the instantaneous cue values over the 5-second analysis interval.

469 Table 1 shows the quartiles of the averages of phonatory frequency, modulation
470 amplitude, modulation frequency and modulation energy ratio, for male and
471 female, control and parkinsonian speakers. Visual inspection confirms that
472 the median phonatory frequency is higher for female than for male speakers,
473 and for male parkinsonian than for male control speakers, while the median
474 phonatory frequency is lower for female parkinsonian than for female control
475 speakers.

476 For the modulation amplitude and frequency, the medians are higher for
477 parkinsonian than for control speakers. For the modulation energy ratio, the
478 medians are lower for parkinsonian than for control speakers.

479 Two-factor analyses of variance are carried out for each cue, the factors be-

480 ing gender and health status. Table 2 summarizes the results via the F -
481 statistic (Zar, 1996) and the probability that the null hypothesis is true.
482 Table 2 confirms the following. First, the interaction between health status
483 and gender is not significant for the modulation frequency and energy ra-
484 tio ($F = 3.1, p = .082$ and $F = 1.1, p = .302$, respectively). Males and
485 females may therefore be lumped together when studying these cues. Sec-
486 ond, the modulation frequency is significantly higher for parkinsonian speak-
487 ers ($F = 4.5, p = .038$), and the modulation energy ratio is significantly lower
488 ($F = 4.5, p = .038$). Third, the interaction between health status and gen-
489 der is significant for two cues: phonatory frequency ($F = 7.3, p = .009$) and
490 modulation amplitude ($F = 4.9, p = .030$).

491 Simple effects analyses via Student's t -test show that for male speakers, the
492 phonatory frequency is significantly higher for parkinsonian speakers ($t =$
493 $-3.3, p = .002$). No significant differences are found for female speakers ($t =$
494 $1.3, p = .197$). Simple effects analyses also show that, for female speakers, the
495 modulation amplitude is significantly higher for parkinsonian speakers ($t =$
496 $-2.7, p = .021$), while no significant differences are found for male speakers
497 ($t = .8, p = .401$).

498 For a discriminant analysis, the independent variables should not be strongly
499 correlated. Therefore, the correlation between cues is studied via Pearson's mo-
500 ment correlation (Zar, 1996), the results of which are presented in Table 3. A
501 statistically significant correlation is observed between modulation frequency
502 and modulation energy ratio for both genders. This correlation is expected
503 because both cues depend on the distribution of the modulation energy in
504 the [3 Hz, 15 Hz] frequency interval. A statistically significant correlation is
505 also observed between phonatory frequency and modulation frequency for the
506 male speakers.

507 The discrimination between parkinsonian and control speakers is studied via
508 Wilks' lambda (Leech et al., 2005). Because the modulation frequency and
509 modulation energy ratio cues are correlated, the latter is not included in the
510 discriminant analysis. The linear combination of the three predictor variables
511 (phonatory frequency, modulation amplitude and modulation frequency) that
512 discriminates best between parkinsonian and control groups is estimated. For
513 male speakers, Wilk's lambda is significant ($\Lambda = .689, \chi^2 = 19.6, p < .001$),
514 which indicates that a model including these cues is able to discriminate sta-
515 tistically between the parkinsonian and control speakers. Table 4 presents the
516 standardized function coefficients and the correlation of each cue with the dis-
517 criminant function, which suggest that phonatory frequency and modulation
518 frequency contribute most. Results show that the model correctly classifies
519 79% of the parkinsonian speakers and 82% of the control speakers. For female
520 speakers, Wilk's lambda is not significant ($\Lambda = .68, \chi^2 = 4.8, p = .19$), which
521 indicates that the model based on the same three cues cannot discriminate

522 statistically between the female parkinsonian and control speakers.

523 4.3 Discussion

524 In this section, the topics are the effects of Parkinson's disease on phona-
525 tory frequency and its low-frequency modulation, and the comparison of data
526 reported here and elsewhere (Winholtz and Ramig, 1992; Schoentgen, 2002;
527 Hirose et al., 1995; Yair and Gath, 1988).

528 4.3.1 Phonatory frequency and Parkinson's disease

529 The average phonatory frequency of males and females is affected differently by
530 Parkinson's disease : it increases for male speakers while it decreases for female
531 speakers. These inter-gender differences agree with most results reported in
532 the literature: Ziegler and Hoole (1999) mention increased pitch only for male
533 speakers with Parkinson's disease. Holmes et al. (2000) study speakers with
534 early and advanced Parkinson's disease: a higher mean F_0 in monologues is
535 associated with advanced disease in men only, with no differences for women.
536 Comparison of the maximum and minimum F_0 during scale singing reveals
537 that females in a later stage of Parkinson's disease have a reduced maximum
538 F_0 compared to early-stage females, while males in a later stage of Parkinson's
539 disease have higher minimum F_0 than early-stage males. However, some studies
540 do not observe any difference in the phonatory frequency of parkinsonian and
541 control speakers (Zwirner et al., 1991; Jimenez-Jimenez et al., 1997). Therefore
542 a controversy still exists in the literature about the effect of Parkinson's disease
543 on the phonatory frequency (Robert and Spezza, 2005). The explanation of
544 these discrepancies may be related to differences in the tasks, measurements
545 or health status of the speakers.

546 4.3.2 Low-frequency modulation of F_0 and Parkinson's disease

547 The modulation frequency increases significantly for parkinsonian speakers of
548 both genders. The observed increase is a consequence of the increase of the
549 spectral energy in the F_0 trace above 7 Hz. Indeed, for parkinsonian speakers,
550 the modulation energy decreases more slowly at high frequencies in the spec-
551 trum of the F_0 fluctuations. For some speakers an energy peak can be observed
552 in the interval [8 Hz, 12 Hz], as shown in Fig. 12. The position of this peak
553 in the vicinity of 10 Hz would suggest a presence of static posture tremor,
554 which occurs between 8 Hz and 12 Hz and which is observed for parkinsonian
555 patients (Gresty and Findley, 1984; Defebvre, 2005).

556 The modulation amplitude is affected differently by Parkinson's disease for

557 male and female speakers: it increases significantly for the females and does
558 not differ significantly for the males. This differs from the observation of nor-
559 mophonetic and mildly dysphonic speakers by Schoentgen (2002), *i.e.* that the
560 modulation amplitude of microtremor does not differ for male and female con-
561 trol speakers, suggesting that the relative cycle-to-cycle perturbations owing
562 to microtremor do not evolve proportionally to the vocal cycle length. In the
563 present study, the interquartile intervals of the modulation amplitude of the
564 male and female control speakers are indeed .55% – .95% and .47% – .73%,
565 as shown in Table 1.

566 The differences in the relative modulation amplitude observed here, however,
567 between control and parkinsonian speakers may be explained by differences in
568 the average phonatory frequencies. The values of the modulation amplitude is
569 significantly higher for female parkinsonian speakers only, compared to female
570 control speakers. Indeed, F_0 drops for the female parkinsonian speakers. The
571 amplitude modulation cue is defined as the standard deviation of the phona-
572 tory frequency divided by the average phonatory frequency (Eq. 6). For female
573 parkinsonian speakers, the decrease of the average phonatory frequency and
574 the increase of the modulation energy possibly contribute both to an increase
575 of the relative modulation amplitude. For male parkinsonian speakers, how-
576 ever, the increase of the average phonatory frequency and the increase of the
577 modulation energy have opposite effects on the relative modulation ampli-
578 tude, which does therefore not differ significantly for parkinsonian and control
579 speakers.

580 Also, one sees in Fig. 12 that most of the modulation energy is concentrated
581 below 7 Hz. In males, a modulation energy increase above 7 Hz would therefore
582 have to be large to counter the effect on the relative modulation amplitude of
583 the phonatory frequency increase. Accordingly, if the modulation amplitude
584 is calculated in the frequency interval [7 Hz, 15 Hz] instead of [3 Hz, 15 Hz],
585 a statistically significant increase of the relative modulation amplitude is also
586 observed for male parkinsonian speakers ($F = 14.69, p < .001$).

587 4.3.3 Comparison with former studies

588 In this subsection, results obtained here are compared to results obtained in
589 former studies. Table 5 shows the quartiles of the average phonatory frequency
590 and modulation cues obtained by Winholtz and Ramig (1992), and Schoentgen
591 (2002). The average phonatory frequency data for the control speakers that
592 are obtained here agree with these data.

593 The modulation frequency data of the control speakers slightly differ from the
594 data obtained by Winholtz and Ramig, and Schoentgen. This difference can
595 be explained by the different frequency intervals for which the analyses are

596 carried out. The frequency intervals are [3 Hz, 15 Hz] in this study, [2.5 Hz,
597 25 Hz] in Winholtz and Ramig's, and [0.1 Hz,25 Hz] in Schoentgen's.

598 In the present study, the relative modulation amplitudes for the control speak-
599 ers are lower than those observed by Winholtz and Ramig, and Schoentgen.
600 These differences can be explained by the dissimilar definitions of the rela-
601 tive modulation amplitude, which is given by the standard deviation of the
602 phonatory frequency normalized by the average phonatory frequency in this
603 study, and by the maximum deviation of the phonatory frequency normalized
604 by the average phonatory frequency in the study by Winholtz and Ramig.
605 Lower values are expected when the standard deviation is used. Schoentgen
606 has studied two modulation amplitude cues : maximum deviation (*Sch1*) or
607 standard deviation (*Sch2*) normalized by the average phonatory frequency.

608 The frequency intervals and components in which the analyses are carried
609 out, however, also differ. In our study, all the spectral energy between 3 Hz
610 and 15 Hz is taken into account, while Schoentgen includes all statistically
611 significant spectral peaks whose frequencies are strictly greater than zero and
612 less than 25 Hz.

613 Further, the low-frequency modulation cues in this study differ from the cues
614 studied by Hirose et al. (1995). They can thus not be compared directly. Both
615 studies arrive at similar conclusions, however. Hirose et al. have observed
616 high-energy fluctuations of the phonatory frequency in the interval [0.1 Hz,
617 16 Hz]. This agrees with the increase of the modulation energy for parkinsonian
618 speakers we observe. But the increase of the modulation energy is not always
619 reflected in the relative modulation amplitude for the males, because it is
620 compensated by the increase of the average phonatory frequency.

621 Finally, the low-frequency modulation data in this study differ from those
622 obtained by Yair and Gath (1988). Yair and Gath observe sharp peaks in
623 the vicinity of 5 Hz in the F_0 fluctuation spectra of parkinsonian speakers.
624 The positions of these peaks match those of the tremor of the limbs of the
625 speakers. This observation raises a question with regard to the comparison
626 of the resting tremor frequency of the limbs with the static posture tremor
627 frequency expected to be observed in sustained speech sounds. Nonetheless,
628 no narrow peaks are observed in the F_0 fluctuation spectra of the parkinsonian
629 speakers participating in the present study. This discrepancy may be explained
630 by differences with regard to the health status of the speakers. The speakers
631 analysed by Yair and Gath present narrow-band tremor of the limbs and
632 narrow-band vocal tremor. The speakers involved in the present study are
633 being treated to eliminate or decrease the symptoms of the disease. They
634 report general speech problems, but their voices are not perceived as tremulous
635 according to the clinical evaluation. The speech analyses confirm this absence
636 of vocal narrow-band tremor, while evidencing other differences in the F_0 low-

637 frequency fluctuations.

638 5 Conclusion

639 An analysis of the frequency modulation of F_0 in the [3 Hz, 15 Hz] interval
640 is presented. The phonatory frequency trace is obtained by means of the in-
641 stantaneous frequency estimated via two continuous wavelet transforms. The
642 first guarantees reliability and the second sensitivity to fast perturbations.
643 Comparison with existing F_0 estimation methods shows the adequacy of the
644 method for the analysis of the low-frequency modulation of F_0 .

645 An analysis of speech sounds sustained by parkinsonian and control speakers
646 shows the following.

- 647 (1) The effects of Parkinson's disease on F_0 low-frequency modulation cues
648 differ for male and female speakers: the phonatory frequency is signifi-
649 cantly higher for male speakers, and the relative modulation amplitude
650 significantly higher for female speakers.
- 651 (2) For parkinsonian speakers of both genders, the modulation frequency is
652 significantly higher and the modulation energy ratio significantly lower
653 than for control speakers.
- 654 (3) A statistically significant discrimination between parkinsonian and con-
655 trol speakers is observed for male speakers, based on modulation fre-
656 quency and phonatory frequency. For female speakers, no statistically
657 significant discrimination is observed, because of the small size of the
658 corpus.

659 Acknowledgements

660 The authors thank John McDonough for helpful comments regarding the
661 manuscript. They thank the anonymous reviewers for their constructive com-
662 ments and helpful suggestions. They also would like to acknowledge support
663 from *COST Action 2103* (Advanced Voice Function Assessment).

664 References

- 665 Addison, P., 2002. The illustrated wavelet transform handbook: introductory
666 theory and applications in science, engineering, medicine and finance. Insti-
667 tute of Physics Publishing.

- 668 Auzou, P., Ozsancak, C., Jan, M., Léonardon, S., Ménard, J., Gaillard, M.,
669 Eustache, F., Hannequin, D., July 1998. Clinical assessment of dysarthria:
670 presentation and validation of a method. *Rev. Neurol. (Paris)* 154 (6-7),
671 523–530.
- 672 Boashash, B., 1992. Estimation and interpreting the instantaneous frequency
673 of a signal - part i : Fundamentals. *Proceedings of the IEEE* 80 (4), 520 –
674 539.
- 675 Boersma, P., Weenink, D., 2004. Praat: doing phonetics by computer.
676 www.praat.org.
- 677 Carmona, R., Hwang, W., Torresani, B., 1997. Characterization of signals by
678 the ridges of their wavelet transform. *IEEE Trans. on Signal Processing*
679 45 (10), 2586 – 2590.
- 680 Cnockaert, L., Grenez, F., Schoentgen, J., 2005. Fundamental frequency esti-
681 mation and vocal tremor analysis by means of morlet wavelet transforms.
682 *Proc. ICASSP, Philadelphia (USA)*, 393–396.
- 683 Defebvre, L., September 2005. La maladie de parkinson. In: Ozsancak, C.,
684 Auzou, P. (Eds.), *Les troubles de la parole et de la déglutition dans la*
685 *maladie de Parkinson*. Solal, pp. 9–30.
- 686 Fant, G., Liljencrants, J., Lin, Q., 1985. A four-parameter model of glottal
687 flow. *STL-QSPR* 4, 1–13.
- 688 Freund, H., 1987. Central rhythmicities in motor control and its perturbances.
689 In: Rensing, L., an der Heiden, U., Mackey, M. (Eds.), *Temporal Disorder*
690 *in Human Oscillatory Systems*. Springer, Berlin, pp. 79–82.
- 691 Fucci, D., Petrosino, L., 1984. The practical applications of neuroanatomy for
692 the speech-language pathologist. In: N.J.Lass (Ed.), *Speech and Language,*
693 *Advances in Basic Research and Practice*. Vol. 11. Academic, New York,
694 pp. 249–317.
- 695 Gresty, M. A., Findley, L. J., 1984. Postural and resting tremor in parkinson’s
696 disease. *Adv. Neurol.* 40, 361–364.
- 697 Hanson, D., Gerratt, B., Ward, P., 1984. Cinegraphic observations of laryngeal
698 function in parkinson’s disease. *Laryngoscope* 94, 348–353.
- 699 Hartelius, L., Svensson, P., 1994. Speech and swallowing symptoms associated
700 with parkinson’s disease and multiple sclerosis: A survey. *Folia Phoniatr.*
701 *Logop.* 46, 9–17.
- 702 Hess, W., 1983. *Pitch Determination of Speech Signals*. Springer-Verlag,
703 Berlin.
- 704 Hirose, H., Imaizumi, S., Yamori, M., 1995. Voice quality in patients with neu-
705 rological disorders. In: *Vocal Fold Physiology*, O. Fujimura and M. Hirano.
706 Singular, San Diego, pp. 235–248.
- 707 Holmes, R. J., Oates, J. M., Phyland, D. J., Hughes, A. J., July 2000. Voice
708 characteristics in the progression of parkinson’s disease. *Int. J. Lang. Comm.*
709 *Dis.* 35 (3), 407–418.
- 710 Jimenez-Jimenez, F., Gamboa, J., Nieto, A., Guerrero, J., Orti-Pareja, M.,
711 J.A.Molina, Garcia-Albea, E., Cobeta, I., April 1997. Acoustic voice analysis
712 in untreated patients with parkinson’s disease. *Park. RElated Dis.* 3 (2),

- 713 111–116.
- 714 Kadambe, S., Boudreaux-Bartels, G., 1992. Application of the wavelet trans-
715 form for pitch detection of speech signals. *IEEE Trans. on Information The-*
716 *ory* 38 (2), 917–924.
- 717 Kawahara, H., Katayose, H., de Cheveigne, A., Patterson, R., 1999. Fixed
718 point analysis of frequency to instantaneous frequency mapping for accurate
719 estimation of f0 and periodicity. *Proc. Eurospeech*, 2781–2784.
- 720 Kent, R., Kim, H., Weismer, G., Kent, J., 1994. Laryngeal dysfunction in
721 neurological disease: Amyotrophic lateral sclerosis, parkinson disease, and
722 stroke. *J. Med. Speech Lang. Pathol.* 2, 157–175.
- 723 King, L., Ramig, L., Lemke, J., Horii, Y., 1994. Parkinson's disease: Longi-
724 tudinal changes in acoustic parameters of phonation. *J. Med. Speech Lang.*
725 *Pathol.* 2, 29–42.
- 726 Le-Tien, T., 1997. Some issues of wavelet functions for instantaneous frequency
727 extraction in speech signals. *Proc. IEEE Tencon 1997*, 31–34.
- 728 Leech, N. L., Barrett, K. C., Morgan, G. A., 2005. *SPSS for intermediate statis-*
729 *tics: use and interpretation*, 2nd Edition. Lawrence Erlbaum Associates.
- 730 Logemann, J., Fisher, H., Boshes, B., Blonsky, E., 1978. Frequency and cooc-
731 currence of vocal tract dysfunctions in the speech of a large sample of parkin-
732 son patients. *J. Speech Hear. Dis.* 43, 47–57.
- 733 Mallat, S., 1999. *A Wavelet Tour of Signal Processing*, 2nd Edition. Academic
734 Press, San Diego.
- 735 Medan, Y., Yair, E., Chazan, D., 1991. Super resolution pitch determination
736 of speech signals. *IEEE Trans. on Signal Processing* 39 (1), 40–48.
- 737 Mitev, P., Hadjitodorov, S., 2003. Fundamental frequency estimation of voice
738 of patients with laryngeal disorders. *Information Sciences* 156 (1-2), 3–19.
- 739 Nakatani, T., Irino, T., December 2004. Robust and accurate fundamental
740 frequency estimation based on dominant harmonic components. *J. Acoust.*
741 *Soc. Am.* 116 (6), 3690–3700.
- 742 Orlikoff, R., Baken, R., 1989. Fundamental frequency modulation of the human
743 voice by the heartbeat: preliminary results and possible mechanisms. *J.*
744 *Acoust. Soc. Am.* 85, 888–893.
- 745 Percival, D., Walden, A., 2000. *Wavelet methods for time series analysis*. Cam-
746 bridge University Press.
- 747 Perez, K., Ramig, L., Smith, M., Fromey, C., 1996. The parkinson larynx:
748 tremor and videostroboscopic findings. *J. Voice* 10, 354–361.
- 749 Qiu, L., Yang, H., Koh, S.-N., 1995. Fundamental frequency determination
750 based on instantaneous frequency estimation. *Signal Processing* 44, 233–
751 241.
- 752 Rabiner, L. R., Schafer, R. W., 1978. *Digital processinf of speech signals*.
753 Prentice Hall.
- 754 Robert, D., Spezza, C., 2005. La dysphonie parkinsonienne. In: Ozsancak, C.,
755 Auzou, P. (Eds.), *Les troubles de la parole et de la déglutition dans la*
756 *maladie de Parkinson*. Solal, pp. 131–143.
- 757 Schoentgen, J., 2002. Modulation frequency and modulation level owing to

- 758 vocal microtremor. *J. Acoust. Soc. Am.* 112 (2), 690–700.
- 759 SQLab, 2005. Eva 2 workstation, for voice and speech clinical assessment.
760 <http://aune.lpl.univ-aix.fr/sqlab/>.
- 761 Titze, I., 1994. Summary statement. Workshop on acoustic voice analysis,
762 National center for voice and speech.
- 763 Titze, I., 1995. Definitions and nomenclature related to voice quality. In: Fu-
764 jimura, O., Hirano, M. (Eds.), *Vocal Fold Physiology*. Singular, San Diego,
765 pp. 335–342.
- 766 Winholtz, W., Ramig, L., 1992. Vocal tremor analysis with the vocal demod-
767 ulator. *J. Speech Hear. Res.* 35, 562–573.
- 768 Yair, E., Gath, I., 1988. On the use of pitch power spectrum in the evaluation
769 of vocal tremor. *Proceedings of the IEEE* 76 (9), 1166–1175.
- 770 Zar, J. H., 1996. *Biostatistical Analysis*, 3rd Edition. Prentice Hall.
- 771 Ziegler, W., Hoole, P., 1999. Neurologic disease. In: Kent, R., Ball, M. (Eds.),
772 *Voice quality measurement*. Singular, p. 397.
- 773 Zwirner, P., Barnes, G., 1992. Vocal tract steadiness: A measure of phonatory
774 and upper airway motor control during phonation in dysarthria. *J. Speech*
775 *Hear. Res.* 35, 761–768.
- 776 Zwirner, P., Murry, T., Woodson, G., 1991. Phonatory function of neurologi-
777 cally impaired patients. *J. Comm. Disord.* 24, 287–300.



Fig. 1. Complex Morlet wavelet for $\omega_c \sigma_t = 5$.

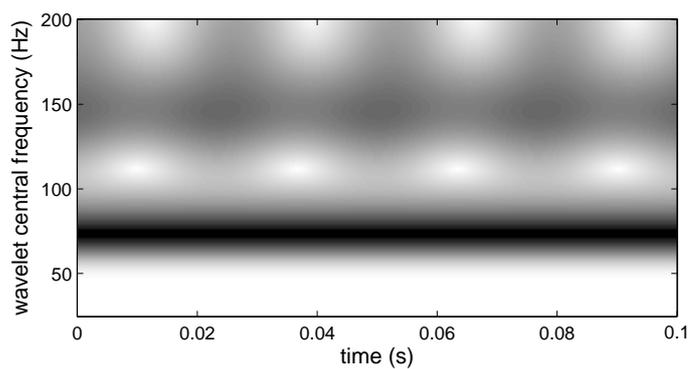


Fig. 2. CWT modulus with parameter $\omega_c \sigma_t = 5$, for a synthetic speech signal (high moduli are represented in black, low moduli in white).

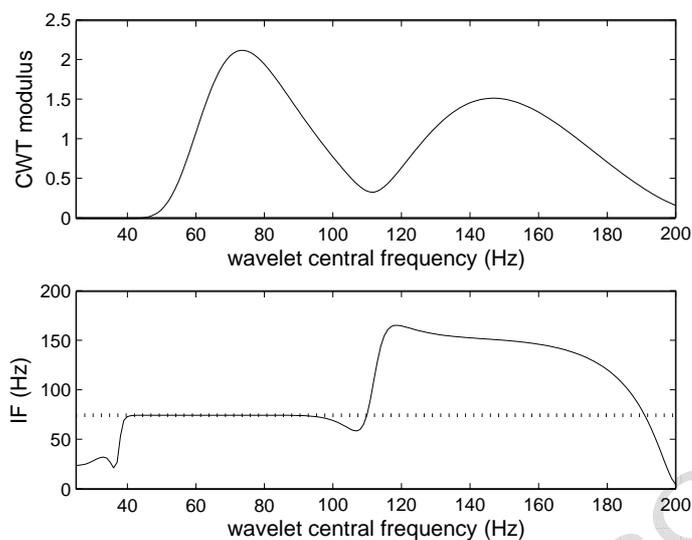


Fig. 3. CWT modulus (upper plot) and instantaneous frequency (lower plot) with parameter $\omega_c \sigma_t = 5$. The dotted line marks $IF = 75\text{Hz}$.

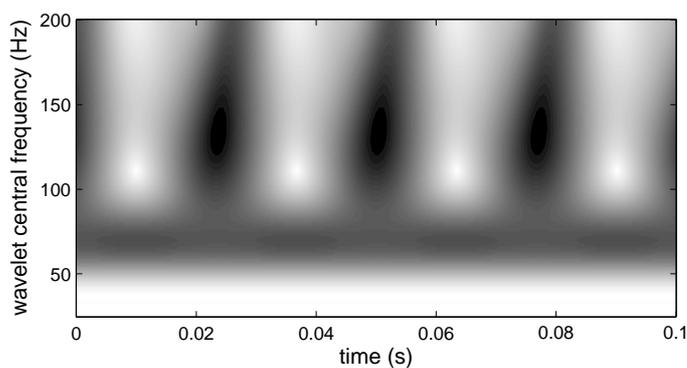


Fig. 4. CWT modulus with parameter $\omega_c \sigma_t = 2.5$, for a synthetic speech signal (high moduli are represented in black, low moduli in white).

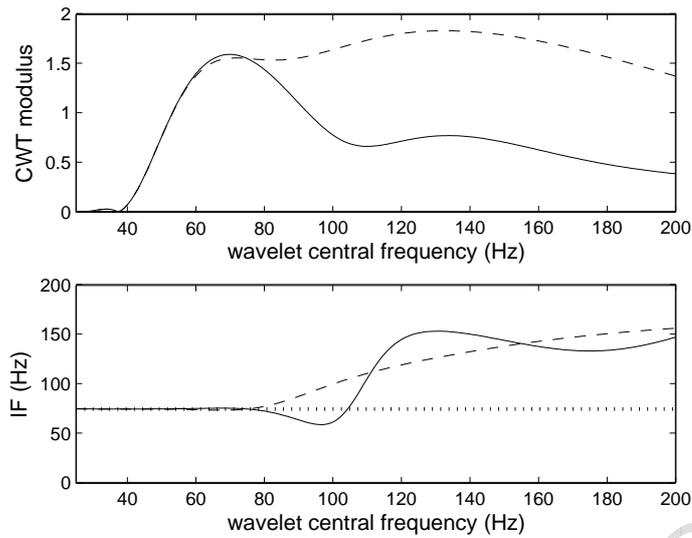


Fig. 5. CWT modulus and instantaneous frequency with parameter $\omega_c \sigma_t = 2.5$, at two different positions in the vocal cycle (plain and dashed lines). The dotted line marks $IF = 75\text{Hz}$.

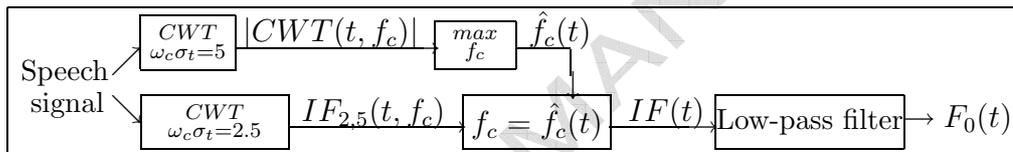


Fig. 6. Block diagram of the F_0 estimation method.

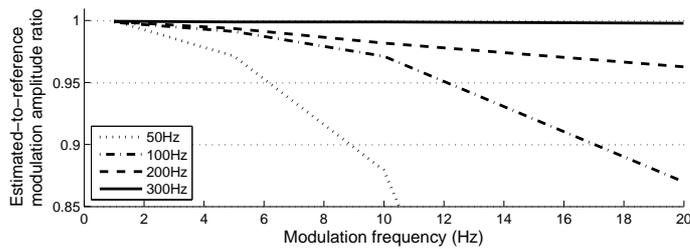


Fig. 7. Estimated-to-reference amplitude ratio of the CWT-based method as a function of modulation frequency, for synthetic vowels with different average F_0 values (50 Hz to 300 Hz).

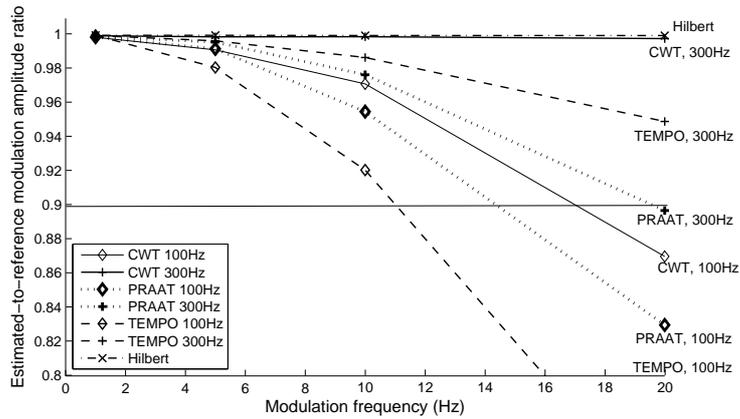


Fig. 8. Estimated-to-reference amplitude ratio as a function of modulation frequency, for synthetic vowels with average F_0 of 100Hz and 300Hz , for different F_0 estimation methods (CC method in PRAAT: dotted lines, TEMPO: dashed lines, CWT-based method: plain lines, Hilbert-based method: dash-dotted line). The plain line marks 90%.

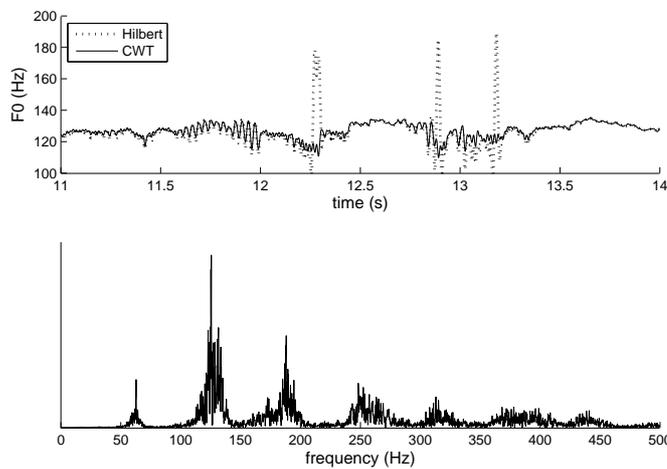


Fig. 9. F_0 traces obtained by the CWT-based method (plain line) and the Hilbert-based method (dashed line), and spectrum of a speech signal with diplophonia, for a 64-year old control speaker.

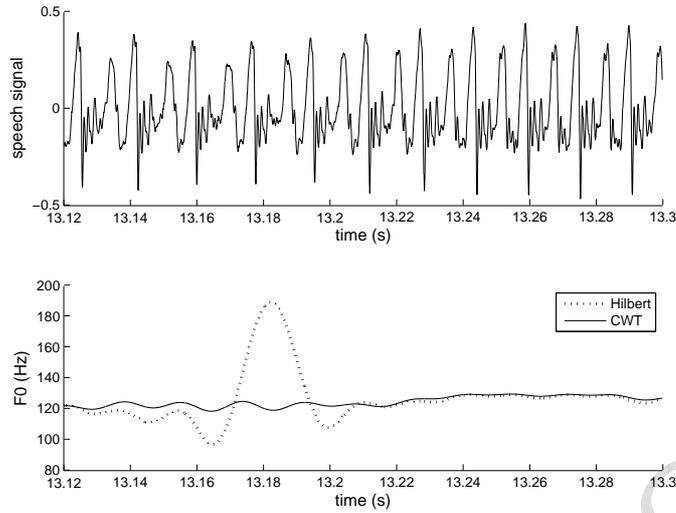


Fig. 10. Zoom of the speech signal, and F_0 traces obtained by the CWT-based (plain line) and the Hilbert-based methods (dashed line), for a 64-year old control speaker.

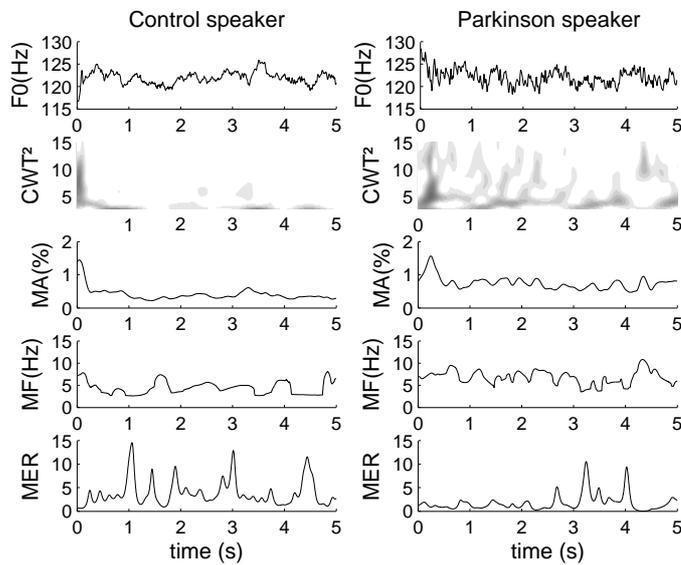


Fig. 11. Phonatory frequency, CWT^2 coefficients, modulation amplitude, frequency and energy ratio, for a control and a Parkinson speaker.

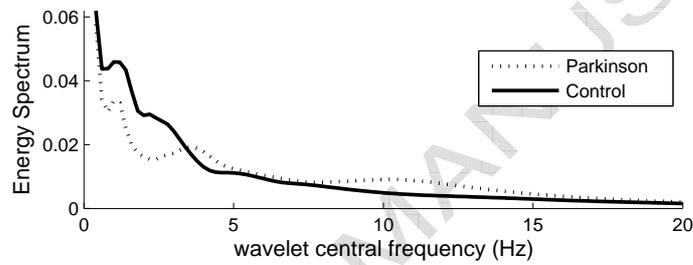


Fig. 12. Energy spectrum of the phonatory frequency for a control speaker (plain line) and a Parkinson speaker (dotted line).

Table 1

Quartiles of the average phonatory frequency F_0 , the average modulation amplitude MA, the average modulation frequency MF and the average modulation energy ratio MER.

		Male		Female	
		Control	Parkinson	Control	Parkinson
F_0 (Hz)	Minimum	88	74	141	107
	Percentile 25	100	120	155	142
	Median	117	144	191	156
	Percentile 75	129	167	205	203
	Maximum	192	239	244	225
MA(%)	Minimum	.29	.38	.37	.44
	Percentile 25	.55	.55	.47	.81
	Median	.78	.86	.65	1.06
	Percentile 75	.95	1.12	.73	1.95
	Maximum	2.31	2.62	1.23	2.68
MF(Hz)	Minimum	3.80	4.05	4.04	4.39
	Percentile 25	4.34	5.26	4.69	4.72
	Median	5.00	6.16	5.06	5.26
	Percentile 75	5.74	6.77	6.04	5.95
	Maximum	7.18	8.13	6.81	6.33
MER	Minimum	2.22	.72	1.89	2.81
	Percentile 25	4.11	1.56	3.71	3.23
	Median	6.16	2.76	6.61	4.79
	Percentile 75	8.91	4.67	7.32	6.33
	Maximum	31.81	18.05	12.87	8.26

Table 2

Results of the two-factor analysis of variance for the average phonatory frequency F_0 , the average modulation amplitude MA, the average modulation frequency MF, and the average modulation energy ratio MER.

Factor	Cue	F	p
Health x Gender	F_0	7.251	.009
	MA	4.900	.030
	MF	3.115	.082
	MER	1.082	.302
Health status	F_0	.152	.698
	MA	8.761	.004
	MF	4.482	.038
	MER	4.477	.038
Gender	F_0	22.914	.000
	MA	1.152	.287
	MF	.934	.337
	MER	.002	.965

Table 3

Correlation between the average modulation amplitude MA, average modulation frequency MF, average modulation energy ratio MER, and average phonatory frequency F_0 .

			F_0	MA	MF	MER
Male	F_0	Pearson Cor. Sig. (2-tailed)	1	.240 .075	.333 .012	-.269 .045
	MA	Pearson Cor. Sig. (2-tailed)		1	.022 .875	.177 .191
	MF	Pearson Cor. Sig. (2-tailed)			1	-.653 .000
	MER	Pearson Cor. Sig. (2-tailed)				1
Female	F_0	Pearson Cor. Sig. (2-tailed)	1	-.449 .081	-.173 .523	.347 .187
	MA	Pearson Cor. Sig. (2-tailed)		1	.335 .205	-.292 .273
	MF	Pearson Cor. Sig. (2-tailed)			1	-.786 .000
	MER	Pearson Cor. Sig. (2-tailed)				1

Table 4

Standardized function coefficients and correlation coefficients of the discriminant analysis, for male speakers.

	Standardized Function Coefficients	Correlation between variables and discriminant function
F_0	.561	.697
MA	.080	.171
MF	.729	.817

Table 5

Results obtained by Schoentgen (2002) (*Sch1* and *Sch2*) and Winholtz and Ramig (1992)(*WR*) for the average phonatory frequency F_0 , modulation amplitude MA and modulation frequency MF, for vowels [a] sustained by healthy speakers.

		Male			Female		
		<i>Sch1</i>	<i>Sch2</i>	<i>WR</i>	<i>Sch1</i>	<i>Sch2</i>	<i>WR</i>
F_0 (Hz)	Minimum	74		106	133		197
	Percentile 25	107		117	175		242
	Median	113		119	191		252
	Percentile 75	127		147	204		270
	Maximum	179		147	325		367
MA(%)	Minimum	.6	.3	.5	.7	.4	.5
	Percentile 25	1.2	.5	.9	1.2	.6	.8
	Median	1.8	.8	1.5	1.9	.9	1.1
	Percentile 75	2.4	1.1	1.8	2.9	1.3	1.3
	Maximum	4.0	2.2	2.8	5.2	1.6	1.4
MF(Hz)	Minimum	2.1	2.0	4.8	1.8	2.0	4.7
	Percentile 25	2.6	3.0	5.5	2.5	2.1	4.9
	Median	3.5	3.2	6.6	2.9	3.1	5.0
	Percentile 75	3.8	4.1	8.0	3.6	4.1	6.1
	Maximum	5.4	5.2	10.6	4.7	5.6	6.6