EAA

Volume

ACTA ACUSTICA
UNITED WITH
ACUSTICA

The Journal of the European Acoustics Association (EAA) · International Journal on Acoustics

Reprint

S. Hirzel Verlag · P. O. Box 10 10 61 · D-70009 Stuttgart

EAA European Acoustics Association

ACTA ACUSTICA UNITED WITH ACUSTICA

The Journal of the European Acoustics Association (EAA) · International Journal on Acoustics

Abstracted/Indexed in:
Cambridge Scientific Abstracts / Electronics and Communication Abstracts;
Current Contents / Physical, Chemical and Earth Sciences;
Current Contents / Engineering, Computing & Technology;
Ei Compendex Plus (Engineering Information);
FRANCIS (CNRS); INSPEC;
Research Alert; SciSearch; Science Citation Index.
Recognized by the European Physical Society

Acta Acustica united with Acustica,
the journal of the European Acoustics Association (EAA), is an international journal on acoustics. Manuscripts of articles and letters should be submitted to the Editor-in-Chief. They will be forwarded to the Associate Editor responsible for the subject area of the paper. A detailed list of Associate Editors and Instructions for Authors are given on the last pages of each issue. Manuscripts are refereed by two experts and reviewed by the Associate Editor and Editor-in-Chief before acceptance.
No honorarium is paid. The first author of an article will receive 25 offprints free-of-charge. Further offprints may be ordered at extra costs. A price list is available from the publisher.
All articles published in this journal represent the opinions of the authors and do not necessarily reflect the opinions of the editors. The publisher cannot be held responsible for unsolicited manuscripts. Submission of a manuscript implies that the work has not been published before and that it is not under consideration elsewhere.

Internet
http://www.Acta-Acustica-united-with-Acustica.com
http://www.ingentaconnect.com/content/dav/aaua
http://eaa-fenestra.org/acta

Subscription Information
Acta Acustica united with Acustica is published bimonthly. Annual institutional subscription rate (Vol. 96, 2010, No 1–6): € 930.00 plus shipping & handling costs (Europe: € 27.00; rest of the world: € 48.00). Airmail costs upon request. Price for a single copy: € 212.00 plus shipping & handling costs. Institutional subscriptions include online access to the full text via http://www.ingenta.connect.com provided a licence agreement is concluded with the publisher. No online availability for additional supplement issues (congress abstracts issues) which are published in irregular intervals. Every print issue includes a CD-ROM with the contents of the current issue. The CD is not sold separately.
Members of a national acoustical member-society of the European Acoustics Association (EAA) receive the journal from their society.
Subscriptions are made on a continuation basis unless ordered for a limited period. Cancellations may only be made for a complete annual volume, they must be received by the publisher by 15 November of the preceding year at the latest.

Publisher
S. Hirzel Verlag GmbH & Co.
Birkenwaldstrasse 44
D-70191 Stuttgart, Germany
Tel.: +49 711 2582-0, Fax: +49 711 2582-290
E-mail: ActaAcustica@Hirzel.de

Production Manager
Dr. Johann Wempen, Daucher Strasse 98
D-85053 Ingolstadt, Germany
Tel.: +49 841 975 852
E-mail: wempen@t-online.de

Responsible for Text Material
Prof. Dr. Dick Botteldooren, Acoustics Group, Department of Information Technology, Ghent University, St. Pietersnieuwstraat 41, B-9000 Gent, Belgium.

Responsible for Advertisements:
S. Hirzel Verlag, Birkenwaldstrasse 44
D-70191 Stuttgart, Germany
Tel.: +49 711 25 82-245, Fax: +49 711 25 82-252
E-mail: anzeigen@hirzel.de.
Present effective rate card: October 1, 2007.

Printed by W. Kohlhammer Druckerei GmbH + Co. KG, Augsburger Strasse 722, 70329 Stuttgart

ISSN 1610-1928 (Print), ISSN 1861-9959 (Online)
Acta Acustica united with Acustica
Volume 96 Number 1
(Acta Acustica Vol. 18 No.1, Acustica Vol. 96 No.1)

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 96 (2010) 567–579

DOI 10.3813/AAA.918307

Advanced Voice Analysis of Pilots to Detect Fatigue and Sleep Inertia

Robert Ruiz[1)], Philippe Plantin de Hugues[2)], Claude Legros[3)]

[1)] Laboratoire de Recherche en Audiovisuel (LA.R.A), Université de Toulouse, 5 allées Antonio Machado, 31058 Toulouse cedex 9, France. robert.ruiz@univ-tlse2.fr

[2)] Bureau d'Enquêtes et d'Analyses (B.E.A) pour la sécurité de l'aviation civile, Aéroport du Bourget, 93352 Le Bourget cedex, France

[3)] Laboratoire d'Acoustique de l'Université Toulouse- le Mirail (L.A.U.T.M), Université de Toulouse, 5 allées Antonio Machado, 31058 Toulouse cedex 9, France

Summary

During long flights, pilots can nap in the cockpit. On daily short-haul rotations, pilots may be tired. Their speech was analyzed in two experiments corresponding to these flight situations: firstly after being woken up in a laboratory sleep inertia experiment, secondly at different airplane stopovers. Measurements of vowel acoustic parameters covered the time and spectral domains as well as the phase space. In addition to classical features and to the maximal Lyapunov exponent drawn from the chaos theory, an additional parameter was introduced: the Digital Amplitude Length (*DAL*). The results show that a large number of acoustic characteristics are modified in the sleep inertia experiment. Only a few of them are modified in the case of fatigue and drowsiness induced in the consecutive daily short flights. The airplane's Cockpit Voice Recorder (CVR) was used for the second experiment and a comparative vowel analysis was performed against a recording on the ground in laboratory conditions.

PACS no. 43.70.Dn, 43.72.Ar

1. Introduction

Voice analysis in aeronautical situations contributes to the study of human factors in flight incidents. The latter do not always imply dramatic consequences for passengers. With the development of technology, the number of airplane accidents has decreased. However, for the remaining accidents, human factors are involved more and more often. Knowledge of the pilot's condition is of a great interest in evaluating the effects of fatigue, stress and workload. By using the medium of sound, the evaluation is not invasive in a context where all the communications use a microphone and are recorded.

Some of the studies deal with the subject by performing simulated flight condition experiments and by analyzing real situations [1, 2, 3, 4, 5, 6, 7]. They were a part of research on voice distorsions due to factors such as emotion, alcohol, stress, illness (pathological voices), workload.

Here, the cause of possible voice distorsions is related to a situation encountered by pilots during both short and long flights: fatigue or drowsiness due to consecutive short flights in a day and the ability to react rapidly after a period of sleep on long flights (sleep inertia). On daily short-haul rotations, pilots can also be in a sleepy state late in the day or even during early morning flights.

Sleep inertia refers to the feeling of grogginess and slowness that occurs immediately upon awakening. Difficulties can be severe enough to impair thinking, decisions and performance. The vulnerability of pilots is important when they sleep during long flights. It is not a desire for sleep or drowsiness. This one appears when a feeling of physical and mental tiredness exists. In the case of pilots this fatigue could result from an important workload during consecutive short daily flights. In the paper, even if fatigue and drowsiness are indistinctly used, the second one is a consequence of the first one.

Pilots voices were recorded in a laboratory sleep inertia experiment and during successive flights of a typical work day.

The same acoustic features have been measured in the two vowel groups removed from the recordings. They covered the time and spectral domains but also the signal phase space. Some of them, such as fundamental frequency, jitter and shimmer are known to be sensitive to various distorsion factors. Others have not been widely tested, like the maximal Lyapunov exponent. A further parameter is proposed: the Digital Amplitude Length of the signal with its derived parameters.

The aim of this paper is to evaluate the ability of a large and representative panel of vocal features to be modified by sleep inertia, drowsiness and fatigue.

Received 15 May 2009,
accepted 10 December 2009.

2. Experiments

Two experiments were carried out: one on sleep inertia linked to long flight conditions and another on fatigue and drowsiness associated with short daily flights.

2.1. Sleep inertia. Long flight conditions

During long flights, pilots sometimes decide to sleep alternately. They can also succomb to short periods of drowsiness. If a technical problem occurs, like an alarm in the cockpit, it is important they react as rapidly as possible not only to evaluate the issue but also to solve it. The question that arises is whether the ability to concentrate and work efficiency is always at an optimum level.

The purpose of the experiment was to create such a situation in the laboratory in order to record speech before and after a sudden awakening.

Some physiological variations are known (like heart rhythm) and voice modifications are also possible. Because of the non-invasive acoustic recording for pilots such research is of great use.

The experiment was conducted with a pilot in a French hospital (sleep disorder department). Three recordings were made:

Recording 1: upon arrival in the lab (11 AM)

Recording 2: after lunch (2 PM)

Recording 3: The pilot was taken to a room where his sleep was monitored. A few minutes after falling asleep he was suddenly awakened by a very powerful light. He then had to perform a number of tasks on a computer similar to those usually performed in flight (3 PM).

In the laboratory the pilot read the same five sentences in each of the three recordings. The acoustic conditions were those of a quiet room with limited background noise and no noticeable presence of standing waves.

2.2. Fatigue, drowsiness. Short flight conditions

A pilot's work period is sometimes made up of several flights from an airport to another within the French territory. The same crew carries out this flight plan for three days running. Flight personnel explain that major fatigue occurs during the first day because of the sudden change in daily rythm after a period of rest. The following days they adapt themselves.

Fatigue or drowsiness are induced because

- pilots have to wake up early in the morning (the first flight takes off early and the airport is some distance from the city centre),
- each flight (about one hour) is too short to have any rest time and air space is very crowded,
- each stopover is short (less than one hour) and the pilots stay in the cockpit to prepare the next flight.

The recordings were made on the first day. Both pilots in the cockpit crew read the same eight sentences at each stopover and also on arrival in the morning. The signal to noise ratio was significant and large enough to ensure good measurement of the acoustic features.

Table I. Distribution of vowels by recording and by phoneme for sleep inertia-long flight conditions experiment.

vowel	recording 1	recording 2	recording 3	total
[a]	12	10	14	36
[ə]	4	5	4	13
[i]	7	4	5	16
[o]	10	8	11	29
total	33	27	34	94

Among all the daily short-haul rotations, the following recording periods were chosen: recording 1 on departure (06h00), recording 2 at one stopover (08h00), recording 3 on arrival (15h00).

To gather more information from the experiment, the airplane's Cockpit Voice Recorder (CVR) was used. The purpose was to test the efficiency of signal processing for non controlled utterances in the presence of background noise during flight. A few sentences were provided from pilot 1's conversations before final arrival.

3. Acoustic material

Except for the CVR analysis, all the sentences were from aeronautical terminology. A sample sentence was: "*Bravo, Victor, Charlie montez au niveau deux cinq zero*" ("*Bravo, Victor, Charlie climb level two five zero*"). They all contain an airplane registration ("*Bravo, Victor, Charlie*" for example). This was a real one in all sentences in the fatigue experiment. They were different in each sentence for the sleep inertia experiment. Pilots were familiar with this choice of vocabulary and the meaning of the sentences.

The sound recording was made with a headset proximity microphone to maintain constant distance between the lips and the transducer for all the head movements (AKG C555L). The recorder was a Sony Pro D.A.T with a sampling frequency of 44.1 kHz and a 16 bit resolution.

Vowels were the phonemes used in the study because consonants have an important intra-speaker variability that can mask abnormal values of acoustic features. Oral vowels were processed because of their greater occurrence in speech than others.

They were segmented with an audio editing software so as to only keep quasi-stationary part of the signal. The attack, the decay and the transition with adjacent phonemes were suppressed. The monophthongs were analysed during their stable period.

The following tables give the distribution of the vowels by periods in the relevant experiment.

The phoneme corpus counted 108 vowels for the "sleep inertia - long flight conditions" experiment: 37 in recording 1, 30 in recording 2 and 41 in recording 3. For spectral analysis, 14 [e] and [ə] were excluded from the set because there were too few in each recording.

The phoneme corpus counted 190 vowels for "fatigue, drowsiness - short flight conditions experiment": 94 for pilot 1 and 96 for pilot 2. Measurements in the time domain

Table II. Distribution of vowels by recording, by pilot and by phoneme for fatigue, drowsiness-short flight conditions experiment and for spectral measurements.

vowels	recording 1		Recording 2		recording 3		total
	pilot 1	pilot 2	pilot 1	pilot 2	pilot 1	pilot 2	
[a]	5	3	5	6	6	5	30
[e]	5	1	5	5	6	6	28
[ɛ]	5	3	3	5	4	5	25
[i]	5	4	9	7	4	10	39
[o]	6	6	5	6	7	7	37
total	26	17	27	29	27	33	159

Table III. Distribution of vowels by pilot and by phoneme for fatigue, drowsiness-short flight conditions experiment and for spectral measurements.

vowels	for all recordings		total
	pilot 1	pilot 2	
[a]	16	14	30
[e]	16	12	28
[ɛ]	12	13	25
[i]	18	21	39
[o]	18	19	37
total	80	79	159

and in the phase space were made on these sets. For spectral analysis, 31 vowels, like [ə], were excluded (14 for pilot 1 and 17 for pilot 2) because they were too few in each recording.

4. Acoustic measurements

The acoustic features were taken from the time domain (amplitude vs time), the spectral ones (sound level vs frequency) and from the phase space.

In the time domain and for each vowel signal, the following characteristics were measured: the mean fundamental frequency, its standard deviation, the coefficient of variation, the mean jitter, the jitter factor, the shimmer, the shimmer factor and the Digital Amplitude Length.

In the spectral domain and for each vowel signal, the center of gravity, the spectral balance frequency, the energy balance frequency, the second, third and fourth spectral moments and the frequency of the first four formants were measured. Additionally, a spectral distance calculation was made between vowel sets.

In the phase space, the phase portraits were plotted and the maximal Lyapunov exponents were computed for each vowel.

Measurements were made by specific laboratory programs developed in the Matlab environment.

4.1. Time domain

Mean Fundamental Frequency ($\langle F_0\rangle$): A peak detection of glottal pulses gives each period duration. The averaging of the corresponding frequencies leads to the vowel mean fundamental frequency $\langle F_0\rangle$ and its standard deviation σ_{F_0}. A cepstral analysis of the signal was also made to check the $\langle F_0\rangle/F_{0\text{cepstrum}}$ ratio was close to one. Such a method made it possible to simultaneously measure peak amplitude for shimmer measurement.

Coefficient of Variation (CV): This is the ratio between standard deviation and mean fundamental frequency.

$$CV(\%) = 100\frac{\sigma_{F_0}}{\langle F_0\rangle}. \tag{1}$$

This dimensionless number gives a better idea of dispersion than standard deviation. It provides data on the medium term instability of F_0.

Mean Jitter ($\langle J\rangle$): Short term instability of F_0 is measured by mean jitter. If there are N periods in the vowel signal and if $F_0(i)$ is the frequency of the i period, mean jitter $\langle J\rangle$ is defined as

$$\langle J\rangle(\text{Hz}) = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|F_0(i) - F_0(i+1)\right|. \tag{2}$$

Jitter Factor (JF): The jitter factor compares mean jitter to mean fundamental frequency. It increases with the instability of $\langle F_0\rangle$.

$$JF(\%) = 100\frac{\langle J\rangle}{\langle F_0\rangle}. \tag{3}$$

Shimmer (S): This explores the small variations in the glottal pulses peak amplitude.

$$S(\text{dB}) = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\log_{10}\frac{A(i)}{A(i+1)}\right|. \tag{4}$$

$A(i)$ is the amplitude of the peak i, N is the number of periods.

Shimmer Factor (SF): Its definition is similar to the one for jitter factor. $\langle A\rangle$ is the mean amplitude over the N periods. This shows the amplitude instability of the vowel signal.

$$SF(\%) = 100\frac{S}{20\log_{10}\langle A\rangle}. \tag{5}$$

Digital Amplitude Length of the signal (DAL): The amplitude-time diagram of an alternating signal is unique

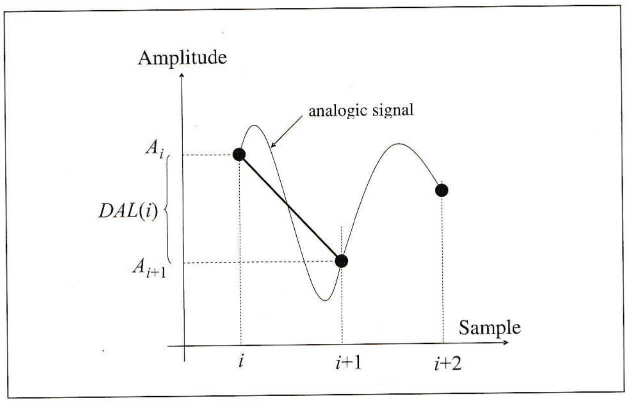

Figure 1. Digital Amplitude Length $DAL(i)$ between two samples.

and generally not predictable except for deterministic signals. Two signals of the same duration have different amplitude-time representations.

An indicator of the shape differences should be the total length measured by applying time stretching in order to transform this alternative signal in a continuous one. However, for digital signals such a method does not lead to an exact value because of the lack of amplitude information between two samples. When the sampling frequency f_s increases the bias is reduced.

Another way to approach the Analog Amplitude Length of the signal (AAL) is to compute the sum of amplitude absolute differences between successive samples from the beginning to the end of the signal. This sum is called DAL (in an amplitude unit),

$$DAL = \sum_{i=1}^{N-1} DAL(i) = \sum_{i=1}^{N-1} \left| A_i - A_{i+1} \right|. \tag{6}$$

N is the total number of signal samples (or for a part of it).

Two errors occur in estimating AAL through DAL (Figure 1): firstly the absence of amplitude between successive samples and secondly the fact that real signal trajectory between successive samples is ignored and replaced by its minimal value $|A_i - A_{i+1}|$.

An approached value of the analog real trajectory is the linear interpolation through its metric calculation by the hypotenuse length (Figure 1). But it is impossible because of the different units on the two axes.

Instead of computing DAL on the entire length of the signal, an alternative choice is to do it for each period T_0 of the vowel signal. In this case N is the number of samples of the period T_0 and DAL is noted DAL_{T_0} (in an amplitude unit). Two resulting quantities can then be defined,

$$\frac{DAL_{T_0}}{T_0} \quad \text{and} \quad \frac{DAL_{T_0}}{\left| A_{\max} - A_{\min} \right|_{T_0}}$$

for each period T_0 in seconds. $|A_{\max} - A_{\min}|_{T_0}$ is the absolute maximal amplitude difference during the period T_0. Another way should be to define the second one in dB with an amplitude reference $|A_{\max} - A_{\min}|_{T_0}$ equal to this difference for the period or the one for the entire signal,

$$20 \log \frac{DAL_{T_0}}{\left| A_{\max} - A_{\min} \right|_{T_0}}$$

or

$$20 \log \frac{DAL_{T_0}}{\left| A_{\max} - A_{\min} \right|_{\text{signal}}}.$$

In this first exploratory study about these types of parameters, the values in dB were not computed.

$\langle DAL_{T_0} \rangle$ is the mean value of DAL_{T_0} over all the periods of the vowel. Therefore, a jitter and a jitter factor can be defined for DAL_{T_0}. It is the same for DAL_{T_0}/T_0 and $DAL_{T_0}/|A_{\max} - A_{\min}|_{T_0}$. If M is the number of periods in the vowel signal,

$$\text{jitter}(DAL_{T_0}) \quad \text{(in amplitude unit)}$$
$$= \frac{1}{M-1} \sum_{i=1}^{M} \left| DAL_{T_0}(i) - DAL_{T_0}(i+1) \right|, \tag{7}$$

$$\text{factorjitter}(DAL_{T_0}) \quad \text{(in \%)}$$
$$= 100 \frac{\text{jitter}(DAL_{T_0})}{\langle DAL_{T_0} \rangle}, \tag{8}$$

$$\text{jitter}\left(\frac{DAL_{T_0}}{T_0}\right) \quad \text{(in amplitude unit per second)}$$
$$= \frac{1}{M-1} \sum_{i=1}^{M} \left| \frac{DAL_{T_0}}{T_0}(i) - \frac{DAL_{T_0}}{T_0}(i+1) \right|, \tag{9}$$

$$\text{factorjitter}\left(\frac{DAL_{T_0}}{T_0}\right) \quad \text{(in \%)}$$
$$= 100 \frac{\text{jitter}\left(\frac{DAL_{T_0}}{T_0}\right)}{\text{mean}\left(\frac{DAL_{T_0}}{T_0}\right)}, \tag{10}$$

$$\text{jitter}\left(\frac{DAL_{T_0}}{\left| A_{\max} - A_{\min} \right|_{T_0}}\right) \quad \text{(dimensionless)}$$
$$= \frac{1}{M-1} \sum_{i=1}^{M} \left| \frac{DAL_{T_0}}{\left| A_{\max} - A_{\min} \right|_{T_0}}(i) - \frac{DAL_{T_0}}{\left| A_{\max} - A_{\min} \right|_{T_0}}(i+1) \right|, \tag{11}$$

$$\text{factorjitter}\left(\frac{DAL_{T_0}}{\left| A_{\max} - A_{\min} \right|_{T_0}}\right) \quad \text{(in \%)}$$
$$= 100 \frac{\text{jitter}\left(\frac{DAL_{T_0}}{\left| A_{\max}-A_{\min} \right|_{T_0}}\right)}{\text{mean}\left(\frac{DAL_{T_0}}{\left| A_{\max}-A_{\min} \right|_{T_0}}\right)}. \tag{12}$$

DAL and AAL can be computed on deterministic signals. For a sinusoid $y = a \sin(x/b)$, the theoretical expression of the one-cycle length AAL is

$$AAL = 4\sqrt{a^2 + b^2}\, E\left(\frac{a}{\sqrt{a^2 + b^2}}\right), \tag{13}$$

where E is an elliptic integral of the second kind.

Here, the purpose is to understand how DAL varied with the shape of the amplitude-time display. For vowel signals, comparisons with an AAL theoretical expression is not possible because this one does not exist. Therefore, DAL is studied on the basis of a digital one-cycle sinusoid with a frequency f, a maximal amplitude a and N samples spaced by $T_s = 1/f_s$ seconds (f_s in Hz).

By choosing $N > 4$ ($N = 4$ is a one-cycle triangular signal), DAL is a constant whatever is the number of samples in the cycle and whatever is the frequency f of the sinusoid. DAL only depends on the maximal amplitude a: $DAL = 4$ for $a = 1$, $DAL = 0.8$ for $a = 0.2$. Therefore, a general rule appears: $DAL = 4a\,\forall f, \forall N > 4$; for n cycles: $DAL = 4an\,\forall f, \forall N > 4$.

Introducing harmonics of the frequency f leads to an increasing value of DAL. From numeric simulations of the one-cycle analog signal,

$$\sum_{k=1}^{K} a\sin(2\pi k f t), \quad (k \in \mathbf{N}^0),$$

$DAL \simeq 4aK, \forall f, \forall N > N_{\min}$. For example, with $K = 3$ and $a = 1$, $DAL = 12.3168$ on one period and $DAL = 2.4634$ with $a = 0.2$ and $K = 3$.

For n cycles: $DAL \simeq 4aKn, \forall f, \forall N > N_{\min}$.

$N_{\min}$ is different from 4 because the shape is more complicated. It needs a greater minimal number of samples to begin to be well displayed. $N_{\min}$ is about 12 for $K = 3$. The general rule is: $N_{\min} = 4K$ to obtain an approximate DAL value close to the one obtained with greater values of N.

One can notice that a small difference exists between computed values and those approached by the equations. These last ones are those of a signal with N values close to $N_{\min}$. For example, the signal for $K = 3$ with $a = 0.2$ and $N_{\min} = 12$ samples (that is equivalent to $f_s = 12f$), leads to a computed $DAL = 4aK$. The display of the signal shows linear segments between maximal and minimal local values. As N (or f_s) increases, they disappear and DAL tends to 2.4634 for the preceding example. So differences between DAL computed values on the signal itself and the ones issued from equations are linked to the accuracy of the signal simulation. For a unique one-cycle sinusoid such differences do not exist since $N > 4$ (or $f_s > 4f$).

Finally, simulations with deterministic signals demonstrate that DAL increases with the complexity of the signal shape.

DAL and its derivative parameters measurements are then used to test if the shape of the signals change with speech distorsions. For example, small peaks or irregularities on the time domain display can be more numerous so that DAL increases when vowels are uttered in a state of fatigue.

The presence of additive noise in speech recordings can falsify the results. Further research should try to evaluate the sensitivity of these features to background noise. But here, the realistic hypothesis is that signal-to-noise ratio are very high, the noise characteristics do not fluctuate as much during the experiments and the comparisons between sets of vowels are done with a similar and very low noise level.

4.2. Spectral domain

The mean spectral envelope of the vowel is obtained from an Auto-Regressive model of each frame by the linear prediction technique (autocorrelation method) with a p order of 48, a pre-emphasis coefficient of 0.98, a Hamming window of 512 samples. A 50% overlapping was applied between consecutive frames.

The prediction coefficients $a(i)$ ($i = 0$ to p) are averaged: each $\langle a(i) \rangle$ is the arithmetic mean of the same order coefficient for all the frames. By applying an FFT on this set (1024 FFT bins), mean spectral envelope of the vowel is obtained [5]. The following features result from these 1024 sound level values. With a sampling frequency of 44.1 kHz, bins take place every 43 Hz on the spectrum. The calculations are made on the 0–5000 Hz band for vowel analysis.

The measurement of the three first parameters (spectral center of gravity, spectral balance frequency, energy balance frequency) corresponds to the search for a unique equilibrium frequency that should be significantly changed in disordered speech. The spectral moments give information about the shape of the spectrum. Formant frequencies are keys to the distinguishability of vowel sounds. Spectral distances between phoneme sets can be modified when a group is composed of rest utterances and the other of monophtongs uttered in a different speaker condition.

Spectral Center of Gravity (SCG):

$$SGC(\mathrm{Hz}) = \frac{\sum_{i=1}^{N} f_i \cdot 10^{L_i/10}}{\sum_{i=1}^{N} 10^{L_i/10}}. \tag{14}$$

f_i is the i FFT bin of the spectrum, L_i is its sound level, $N = 1024$. To keep energy units, sound level of each bin is divided by ten (instead of 20 for the amplitude). By analogy with mechanics, SCG indicates the frequency where the energies gather to give an equilibrium.

Spectral Balance Frequency (SBF):

A Cumulative Spectral Probability Diagram ($CSPD$) is computed for two frequency bands of the spectrum [5, 7]. The Spectral Balance Frequency (SBF) is the one for which the area between the two $CSPD$ is minimal. It is not exactly an area but the sum of the absolute differences for all the level classes between the two $CSPD$. SBF is a bin of the envelope spectrum. The level class width is 0.5 dB. The number of classes is double than the absolute difference between maximal and minimal sound level of the spectral envelope.

Energy Balance Frequency (EBF): This is the frequency for which the energies in the lower and upper band are closest:

$$\left(\sum_{i=1}^{N} 10^{L_i/10}\right)_{0-EBF} \simeq \left(\sum_{i=1}^{N} 10^{L_i/10}\right)_{EBF-5000\,\mathrm{Hz}}.$$

Spectral Moments (M_n): These are related to the spectral center of gravity. M_2 is the variance of frequencies in the spectrum, M_3 is the non-normalized spectral skewness and M_4 is the non-normalized spectral kurtosis. To normalize M_3 and M_4 it is necessary to divide them by 1.5 power of the second moment for M3 and by the square of the second moment and substract 3 for M_4.

$$M_n = \frac{\sum_{i=1}^{N} (f_i - SGC)^n \cdot 10^{L_i/10}}{\sum_{i=1}^{N} 10^{L_i/10}}. \tag{15}$$

Normalized skewness is a measure of asymmetry of the spectrum around its spectral center of gravity. The skewness is a measure for how much the shape of the spectrum below the SCG is different from the shape above SCG. It is zero if the shape appears to have normal distribution. Negative values indicate data that are skewed to the left and positive values indicate data that are skewed to the right. Skewed to the left means that the left tail is long relative to the right tail. Similarly, skewed to the right means that the right tail is long relative to the left tail.

Normalized kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution (normalized kurtosis = 0). Data sets with positive kurtosis tend to have a distinct peak near the SCG, decline rather rapidly, and have heavy tails. Data sets with negative kurtosis tend to have a flat top near the SCG rather than a sharp peak.

Formant Frequencies (F_i): The frequency of the first four formants are estimated by a second derivative computing method. Differences between adjacent bins are calculated to aproximate the second derivative. When the sign of successive differences changes, formant frequency is detected. Comparisons with peak picking method show very close results. This is due to the smooth envelope obtained with an order of the linear prediction (48) not much higher than the sampling frequency expressed in kHz (44.1).

Spectral Distance (SD): This is the euclidean distance between two sets of Mel Frequency Cepstral Coefficients ($MFCC$) [8].

The vowel signal is divided in successive frames of 512 samples (with 50% overlapping). For each one of them, the process is the following: pre-emphasis (coefficient = 0.98); Hamming weighting; FFT (512 bins); Mel filtering (32 filters between 133-5000 Hz); Neperian logarithm of the filtered signal; Discrete Cosinus Transform (DCT).

It makes it possible to obtain a set of 32 $MFCC$ for every frame. Averaging is done for each $MFCC$ coefficient over all the frames to obtain the mean $MFCC$ set of the vowel.

Spectral distance is then computed between two mean $MFCC$ vectors of vowels uttered by the same speaker in two different recordings of each experiment.

4.3. Phase Space and Maximal Lyapunov Exponent λ

The phase portraits are drawn for an embedding dimension of three which is an appropriate choice for vowels [9]. The reconstruction delay τ is chosen equal to the first zero of the autocorrelation function of the monophtong signal. This choice is similar to Rosenstein's one [10] and insures the independance of the points in the trajectory.

For $n = 1$ to $(N - 2\tau)$, the reconstructed trajectory $X_n = (x(n), x(n+\tau), x(n+2\tau))$ gives the coordinates of the points in the phase portrait that form the attractor. N is the total number of samples. X_n describes the state of the system at discrete time n.

For $n = 1$ to $(N - 2\tau)$, the set of coordinates is processed to obtain the maximal Lyapunov exponent λ of the time series by Kantz's method [11]. The algorithm tests exponential divergence of nearby trajectories from experimental time series without an explicit equation defining the dynamics of the system.

A neighborhood U_{X_n} of each X_n is searched in the trajectory by defining a sphere of diameter ε. The X'_n points inside the sphere have an euclidean distance with X_n less than ε and they belong to U_{X_n}. To evaluate the divergence of the trajectory, the following function $S(\Delta n)$ is computed (from [11])

$$S(\Delta n) = \frac{1}{N}\sum_{i=1}^{N} Ln\left(\frac{1}{|U_{X_n}|}\sum_{X'_n \in U_{X_n}} \left|X_{n+\Delta n} - X'_{n+\Delta n}\right|\right). \tag{16}$$

$|U_{X_n}|$ is the number of neighbours of X_n inside the sphere, Δn is the time span.

The linear part of the plot of $S(\Delta n)$ as a function of Δn gives the maximal λ. Indeed, slope value varies with the number of time spans Δn involved in the computation of the linear regression on the corresponding values of $S(\Delta n)$.

For automatic processing on a set of vowels, the implementation of Kantz's method is based on a fixed maximal value of Δn before analysis instead of computing $S(\Delta n)$ for various values of Δn to choose the maximal slope. This is obtained with $\Delta n = 1$ to 5 for each vowel signal.

Validation of the algorithm for Logistic and Henon time series [12] is done by comparing the theoretical λ with their estimation with time series lengths equivalent to those of the vowels. For the Henon map, theoretical λ = 0.418 and estimated one is 0.421 with 4000 samples. For the Logistic map, theoretical λ = 0.693 and estimated one is 0.654 with 3000 samples ($\Delta n = 1$ to 5 and $\varepsilon = 2\%$).

The analysis conditions (ε and Δn_{max}) are the same for all monophtongs in order to better compare them.

The size of the neighbourhood is variable from a vowel to another but the rule is the same. ε is a percentage of the maximum peak-to-peak amplitude of the signal [13] chosen here equal to 2%. With lower values of ε, the small number of neighbours ignore some parts of the attractor and greater values of do not improve the results.

4.4. Acoustic measurements for the cockpit voice recorder (CVR)

For the experiment on fatigue, some sentences extracted from the airplane CVR were used to test acoustic measurements. Two main differences existed with recordings inside the cockpit at the stopover: background noise and sampling frequency.

This was 8400 Hz for the CVR instead of 44.1 kHz. This led to reduce the frame length to 128 samples. Such a length provided many frames per vowel to give a better spectral estimation. The measurement of time domain parameters was not affected by the sampling frequency change. In the phase space, attractors were less well described by a reduced number of samples when the sampling frequency was smaller. Moreover, the phoneme duration may decrease in free speech conditions compared to the one when a speaker has to read a text. Lyapunov exponent estimation is all the more sensitive since time series is short.

The background noise level was higher for the cockpit voice recording during the flight than for recordings with laboratory equipment and with the airplane stopped on the ground. The ambiant noise had a wide band and deteriorated vowel-to-noise ratio for medium and high frequencies because of decreasing sound level for vowel spectrum and relative constant level for noise between 0 and 5000 Hz. Without any modification of the analysis conditions, vowel spectra showed high sound levels at high frequencies. By applying a null pre-emphasis, spectral envelopes became more usual. In another experiment with V.H.F communications from a CVR, a first order low-pass filtering was applied before LPC analysis to obtain correct spectra together with a zero pre-emphasis [14]. Here, this additional pre-processing was not necessary because noise level was lower.

The maximal Lyapunov exponent estimation is very sensitive to noise level. In order to suppress noise points in the neighbourhood, a minimal sphere of diameter is defined in the reconstructed trajectory. The neighbours are then the points belonging to the space between the two spheres of diameter δ and ε ($\delta < \varepsilon$) [13].

The lower ratios between "speech + noise" and "noise" amplitudes during an utterance were about 6 (the maximal ones were greater than 40). This meant that amplitude noise varied from 16% to 2.5% of the "speech+noise" amplitude. By choosing a high δ value, true neighbours could be ignored for important signal-to-noise ratios and if δ was too small noised neighbours could be taken into account for Lyapunov exponent estimation. Considering that utterances were predominantly composed of high vowel-to-noise ratios, was chosen equal to 2%. For laboratory recording conditions in the two main experiments, ε was 2% with no noise sphere. To maintain such an interval, $\varepsilon = 4\%$ for the CVR analysis.

The total number of time spans Δn was decreased to 3 because the smaller sampling frequency led to an inferior number of samples for approximatly the same vowel duration.

5. Results

For all the parameters excepting the spectral distance, the analysis method was to study their chronologic variations on a two-dimensional graphical representation and to proceed to statistical t-tests to compare their mean values for the different recordings of the experiments (see 2.1 and 2.2).

All the detailed results are in study reports (in French) and can be obtained from the authors. Only the more remarkable variations are presented here.

Before analysis, the assumptions of the study were as follows:

Hypothesis 1 (H1): Sleep inertia experiment: the vowel features of recording 3 would be significantly modified compared with those of recordings 1 and 2. They were uttered just after a sudden awakening.

Fatigue, drowsiness experiment (H1.1): Features of recording 3 were different from recording 1 and 2 because fatigue was supposed to increase throughout the day. A regular variation of parameters would be expected. (H1.2): The drowsinness could appear at any moment. Therefore, in every recording voice modifications were possible.

Hypothesis 2 (H2): The chronological variations of the parameters suggested the possibility that a relevant modification existed by visual inspection of the figures. In that case, the observation had to concern a set of consecutive vowels.

Hypothesis 3 (H3): Statistical comparison of mean values between recordings of the same experiment (t-test) might show the existence of a significant change.

Hypothesis 4 (H4): The second assumption did not imply the third. But the inverse was true. Even for small vowel samples, the adapted statistical test could not be significant whereas visual observation showed the emergence of variations.

All the results were analyzed with regard to these assumptions. In Table IV, the crosses indicate if one of the first three hypothesis is verified for each acoustic parameter in the two experiments.

Table IV gives a global picture of the sensitivity parameters that correspond to sleep inertia, fatigue or drowsiness.

A greater number of parameters vary in the first experiment than in the second one. The cause may be due to the "violence" of the situation in which the speaker has been plunged into for sleep inertia compared to the fatigue and drowsiness manifestations.

Only some *DAL* derivative features and the maximal Lyapunov exponent (λ) are modified in the two experiments. This produces a promising result for which had never been tested in such situations and for the derivative *DAL* parameters.

More detailed results about significant variations are presented in the following paragraphs.

5.1. Time domain results

Figure 2 shows a significant and observable jump in mean fundamental frequency ($\langle F_0 \rangle$) for the vowels belonging to recording 2 in the second experiment and for the pilot 1. This is why hypothesis 1, 2 and 3 are true in Table IV.

A more noticeable jump is observed for jitter in the first experiment (Figure 3). It appears for the first vowels uttered after the awakening. The same significant increase is present for the coefficient of variation, the jitter factor, the