ELSEVIER

Speech Communication 20 (1996) 111–129

SPEECH
COMMUNICATION

Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions

Robert Ruiz [a,*], Emmanuelle Absil [b], Bernard Harmegnies [b], Claude Legros [a],
Dolors Poch [c]

[a] *Laboratoire d'Acoustique, Université de Toulouse-le Mirail, 5, allées Antonio Machado, 31058 Toulouse Cedex 1, France*
[b] *Laboratoire de Phonétique, Université de Mons-Hainaut, 18, place du parc, B-7000 Mons, Belgium*
[c] *Laboratori de Fonètica, Universitat Autònoma de Barcelona, Campus Bellaterra, 08193 Bellaterra, Spain*

Received 15 April 1996; revised 15 June 1996

Abstract

Stress induced by various types of situation leads to vocal signal modifications. Previous studies have indicated that stressed speech is associated with a higher fundamental frequency and noticeable changes in vowel spectrum. This paper presents pitch- and spectral-based analyses of stressed speech corpora drawn from both artificial and real situations. The laboratory corpus is obtained by means of the Stroop test, the real-case corpus is extracted from the Cockpit Voice Recording of a crashed aeroplane. Analyses relative to pitch are presented and an index of microprosodic variation, μ, is introduced. Spectrum-related indicators of stress are issued from a cumulative histogram of sound level and from statistical analyses of formant frequencies. Distances to the $F1$–$F2$–$F3$ centre are also investigated. All these variations, throughout the two different situations, show the direct link between some new vocal parameters and stress appearances. The results confirm the validity of laboratory experiments on stress, but emphasize quantitative as well as qualitative differences between the situations and the speakers involved.

Zusammenfassung

Der durch verschiedene Situationstypen hervorgerufene Streß führt zu Veränderungen des Stimmsignals. Vorhergehende Untersuchungen haben gezeigt, daß die gestresste Stimme durch eine höhere Grundfrequenz und Schwankungen im Vokalspektrum gekennzeichnet ist. Der folgende Text stellt die gemeinsamen Analysen dieser beiden Parameter anhand des in realen und künstlichen Situationen gestressten Stimmkörpers dar. Der Stimmkörper des Labors ist der des Strooptestes, und der, der realen Situation ist der Auszug einer Cockpitgesprächsaufnahme eines verunglückten Flugzeuges. Die Grundfrequenz ist makroskopisch untersucht und ein Index μ für die mikroprosodischen Veränderungen eingeführt worden. Die Ludikatonen des Streßspektrums sind von einem Histogramm abgeleitet, das den Geräuschpegel und statistische Analysen der Formantfrequenzen verbindet. Die Abstände $F1$–$F2$–$F3$ im Zentrum sind ebenfalls untersucht worden. All diese Abweichungen zeigen in Bezug auf die zwei Situationen, daß ein direkter Zusammenhang zwischen den neuen Parametern des Stimmsignals und dem Erscheinen des Streß besteht. Die Ergebnisse bestätigen die Gültigkeit der Laborexperimente, aber zeigen auch quantitative und qualitative Unterschiede zwischen den Situationen und den betreffenden Sprechern auf.

---

[*] Corresponding author.

PII S0167-6393(96)00048-9

Résumé

Le stress provoqué par divers types de situations conduit à des modifications du signal vocal. Des études précédentes ont indiqué que la parole stressée est caractérisée par une fréquence fondamentale plus élevée et des altérations du spectre des voyelles. Cet article présente les analyses conjointes de ces deux paramètres à partir de corpus de parole stressée obtenus à la fois dans une situation réelle et dans une situation artificielle. Le corpus de laboratoire est celui du test de Stroop et le corpus de la situation réelle est extraît d'un enregistreur des conversations d'un avion accidenté. La fréquence fondamentale est étudiée macroscopiquement et un index μ de la variation microprosodique est introduit. Les indicateurs spectraux du stress résultent d'un histogramme cumulé du niveau sonore et d'analyses statistiques des fréquences des formants. Les distances par rapport au centre $F1$–$F2$–$F3$ sont aussi étudiées. Toutes ces variations, à travers les deux situations, montrent un lien direct entre certains nouveaux paramètres du signal vocal et les apparitions du stress. Les résultats confirment la validité des expérimentations de laboratoire, mais mettent également en évidence des différences quantitatives aussi bien que qualitatives entre les situations et les locuteurs concernés.

Keywords: Voice analysis; Pitch; Spectrum; Stress; Emotion

1. Introduction

Research dealing with the effects of stress on the speech signal has taken into account a large variety of factors causing stress. Broadly speaking, they can be divided into two main categories.

On the one hand, researchers have tried to set up various experimental devices aimed at reproducing in their laboratories conditions supposed to provoke stress in the subjects under study. Quite a large variety of stressors has been used for this purpose.

Among them, the first one to be studied was probably audio-feedback perturbation. First observations, reported by E. Lombard in 1911, led the author to introduce the so-called "signe de l'élévation de la voix" (Lombard, 1911; Lane and Tranel, 1971). Many other researchers, such as Black (1951), Webster and Klumpp (1962), or more recently Junqua (1992, 1993), have later taken part in the study of this phenomenon.

Emotionally induced stress was also described and studied quite early on. For example, Fairbanks and Hoaglin (1941) worked on emotion simulation by professional actors. Thompson and Mason (1995) studied the effect of anxiety in visual and audio speech databases in an unfamiliar situation of recording environment. Scherer (1981) had subjects purposely lie to the experimenter or to another person. Doherty and Hollien (1978) used aversive stimulations such as ammonia inhalation and observed vocal modifications. Others made use of varied electric shocks (Rubenstein, 1966).

Cognitive stressors can also be used to manipulate the variable in tasks, including, e.g., arithmetic computations, word associations, and so on (Harmegnies and Landercy, 1992). Griffin and Williams (1987), for instance, studied the influence of three complexity task levels on the subjects' vocal production. When such cognitive stressors are used in order to provoke stress, the experimenter may, in addition, ask the subject to execute the task as quickly as possible (e.g., Allen and Crowell, 1989); this means adding some stressful condition to task difficulty.

On the other hand, researchers have tried to benefit from the opportunities introduced by the natural, non-provoked emergence of obviously stressful events. In such cases, stress is supposed to exist on its own, as a result of the environmental aggression on the subject. Here, the study does not take place in a laboratory, and the recording conditions are no longer artificial. Whether the research is feasible or not depends upon the availability of a recording device at the moment when the stressing event appears.

Various real events have also constituted opportunities for collecting samples of stressed voices. The most popular ones are related to aeronautics, since flight incidents are obviously dangerous – and therefore supposed to be stressful for those who are aware of their occurrence –, and also because recording devices, either remote or embarked (Cockpit Voice Recorders), are generally functioning when the incident occurs. Pilot radio communication tape recordings or cosmonaut voice recordings have therefore been the object of extensive study (Simonov and Frolov, 1973, 1977; Sulc and Remek, 1985; Ruiz et al., 1990; Protopapas and Lieberman, 1995; Benson, 1995).

Among the natural factors expected to provoke stress, mental illness has also received attention (e.g., Newman and Matter, 1938; Helfrich et al., 1984). In some cases, variations of the patients' states (either improvement or degradation) constituted the independent variable of the study (Hargreaves and Starkweather, 1964; Ostwald, 1965; Scherer, 1981).

From the epistemological viewpoint, the main difficulty in this field is precisely the existence of those two research streams, the first one relying on real events, and the other on artificial laboratory situations. Apart from the methodological variability, differences in the concept of stress obviously appear: as a matter of fact, it seems difficult to agree, at first, with the idea that the sudden inhalation of ammonia in the quiet atmosphere of a university laboratory should provoke the same kind of stress as the evidence of imminent death in an aircraft crash.

Moreover, most of the papers focus on a single stressful situation, and each reported study seems to belong either to one research stream or to the other. The analysis methodology is therefore often specifical to the study. Thus, it is quite difficult both to check for the validity of laboratory induced stresses, and to correlate findings from other experiments. Research dealing at the same time both with real and artificial stress, and using unified methodologies therefore appears highly commandable, in order to collect comparable results.

The aim of this paper is precisely to introduce a first contribution to this end. Treatments previously applied on real-event stress (Ruiz and Legros, 1994) are used for the processing of laboratory data, although procedures firstly used to treat speech samples deriving from laboratory stress (Harmegnies and Landercy, 1992), are applied to recordings collected in a real-life environment.

The laboratory stressful situation uses a psychometric stress test based on a cognitive task called "Color-Word Test" (Stroop, 1935). The real situation is constituted by an aircraft incident involving two pilots and preceded by long discussions about the failures.

The purpose of this work is to test various acoustic features (pitch based parameters, as well as spectral data) in the two stressful situations, and to evaluate their ability to vary under the influence of an increasing stress, firstly in a controlled environment and secondly in real conditions. Being both a methodological and an exploratory study, the work which is reported in this paper aims at testing the suitability of analysis procedure in various conditions, and at getting first insight about comparability of real-life studies and laboratory experiments.

2. Experimental conditions

2.1. Laboratory situation

The laboratory stress corpus was derived from the application of a French version of the so called "Colour-Word Test" by Stroop (1935); one young Belgian French-speaking subject undertook it. The test is divided into three different phases (referred to as phases 1, 2 and 3), within each of which the subject has to utter colour names as quickly as he can.

During the first two phases, the subject has firstly to read colour names, then to name the colours of small squares. For the third part of the test, he is given 50 rows of 4 colour names each. This time, however, the words are not written in black on white: the ink to write each word is of a different colour from the one referred to by the word itself. The subject has to name the colour of the ink used. The colours are green, yellow, red and blue.

The third phase is considered as providing a stressful situation, because it provokes a cognitive conflict between a reading (more natural for a cultivated subject) and a naming task; the subject has to refrain from reading, although he is presented with a reading material.

The choice of the Stroop task as a stress inducer arises here from its ability to create unambiguously contrasted control and stress situations.

2.2. Real situation

The second corpus is extracted from the Cockpit Voice Recorder (CVR) of a crashed aircraft. The pilot's and the copilot's utterances have been analysed.

The timing of the extract firstly shows the occurrence of a failure in flight commands. The incident arouses discussions between the pilot and the copilot, about its severity and its causes. At the end of the recording, the plane crashes. This corpus obviously contains a stressful phase (at the end of the flight) and other phases with lower stress (during discussions relating to technical problems); a rest period exists before the flight incident.

Therefore, three possible emotional states can be considered: one of important stress, at the end of the flight, one of weaker stress, during the discussions about the technical problem, and one of absence of stress, corresponding to verbal exchanges not involving any technical problem, and taking place during quiet moments. They will be referred to as levels 2, 1 and 0 of the independent variable stress.

3. Method

3.1. Speech data

The laboratory corpora consists of 180 oral vowels (60 in each phase), i.e., 45 utterances of /u/ from "rouge" (red), /ɛ/ from "vert" (green), /o/ from "jaune" (yellow), /ø/ from "bleu" (blue). This selection procedure allows context invariance for the comparisons.

The real situation analyses are performed on a restricted, balanced sample of words extracted from the three periods of the flight. Only vowels with the same anterior context in the three subsamples are taken into account: the vowel is selected under the requirement that the consonant–vowel combination appears in all of the three parts of the CVR corpus. Any vowel matching this condition has been taken into account, i.e., the corpus contains tokens of all the oral vowels of French. The sample size is 366, for the copilot, and 307, for the pilot.

3.2. Acoustic measures

Speech and especially monophthongs are of interest, for the moment, to detect acoustical stress manifestations.

Source-dependent information will be drawn from analyses of pitch. This will be studied both in terms of macroscopic changes under the effect of stress, and from the viewpoint of its microscopic variations. Characteristics of oral resonance will be analysed thanks to a spectral approach based upon the study of energy distribution in the frequency domain, and also by means of formants frequencies evaluation on the basis of spectrographic observation.

3.2.1. Pitch

Fundamental frequency F_0 is the common acoustic feature of most of the research, dealing with vocal manifestations of task-induced stress (Hecker et al., 1968; Williams and Stevens, 1972; Benson, 1995). All the previous results indicate pitch increasing under stress but also modifications of regularity of F_0 time contours and of jitter (Benson, 1995; Brenner et al., 1994; Gramatica et al., 1992; Griffin and Williams, 1987; Hecker et al., 1968; Protopapas and Lieberman, 1995; Williams and Stevens, 1972, 1981). It seems to be the unique voice parameter sensitive to different types of stress factors (Ruiz et al., 1990). In this paper, pitch is studied both in the real and laboratory stress situations and its variations are analysed in two ways: a macroscopic and a microprosodic one.

The former consists in the study of mean fundamental frequency variations in vowels. The latter, inside these voiced sounds, examine the discriminating ability of a new parameter, μ, following:

$$\mu = \frac{f_c}{(f_i + f_f)/2}, \tag{1}$$

where f_i, f_c and f_f are, respectively, the initial, central and final fundamental frequencies in a given voiced sound. The μ index equals 1 when the fundamental frequency remains invariant during the vowel utterance. It is greater than 1 when the frequency at the centre is higher than those at the boundaries. This pattern seems to be frequent under certain forms of stress, on the basis of clinical observations.

The pitch data have been determined on the basis of the pitch tracking algorithms implemented on the CSL 4300 and DSP 5500 Kay analysers; these are based upon computation of the inverse of the time between dominant amplitude peaks in 15 ms frames and frame-to-frame correlation control of the resulting pitch values. ILS software cepstrum-based measurements were also used in the case of the CVR recording. Each time between-procedure discrepancies or suspicious values appeared, the between-harmonic range, computed on the basis of narrow-band spectrograms was considered instead.

The results are first analysed in term of the observed frequency values. Inferential treatment are then applied on frequencies transformed into Mel values. Two-way analyses of variance are performed. The same treatment is also performed on μ.

3.2.2. Spectrum

Comparative observations of sonagrams at rest and under stress (or simulated emotions) have indicated changes in shape (e.g. attack, regularity of successive pulses, accuracy of articulation,...) (Hecker et al., 1968), and modification of the amount of high frequency energy (Hecker et al., 1968; Williams and Stevens, 1981). Such effects are difficult to quantify and sonagrams are very useful to this effect. Analyses of octave and third octave long term average spectra led to the same conclusion (Williams and Stevens, 1972, 1981). A specific indicator from third octave spectral analysis has been introduced to evaluate stress in different phases of a Russian space flight. It also indicates energy increase in high frequency with an increasing heart rate (Simonov and Frolov, 1973, 1977).

Studies of formant variations under stress have shown interesting results. In (Williams and Stevens, 1972), the authors noticed modifications of the formant structure. Shifts of the first two formant frequencies under stress induced by a simulated motion disorientation task have also been reported (Mosko et al., 1983).

A specific analysis of an audio tape of a pilot's speech during a serious aircraft failure (Benson, 1995) shows pitch, formant frequency and spectrum slope variations indicating higher pitch and first formant frequencies but also spectrum flattening.

Finally, it appears that two major kinds of modification exist: energy displacements and shifts of formant frequencies. In this paper, these two types of acoustic features are studied.

3.2.2.1. Frequency distribution of sound level. The first acoustic feature is evaluated by the estimation of Cumulative Spectral Probability Diagrams (CSPD) in two complementary spectral bands (Ruiz and Legros, 1994).

The CSPD is a cumulative histogram of sound level issued from the discrete envelope spectrum obtained by FFT from the set of LPC coefficients filled with zeros in order to obtain a frequency resolution of 9.75 Hz.

CSPDs are computed with a 1 dB class width and smoothed by polynomial regression. The signal is low-pass filtered at 5 KHz, sampled at 10 KHz. A 15th linear prediction model of each pre-emphasized and Hamming weighted frame is used.

Envelope spectra result from an LPC transform of each frame of the signal. A mean spectral envelope of the steady-state vowel is computed by averaging the sets of prediction coefficients. The mean spectral envelope is usually not very different from the envelope of each frame because of the quasi-stationarity of the signal and the choice of a 75% overlapping between consecutive frames.

The CSPD is computed on the mean discrete spectral envelope of the vowel. More than a simple comparison of energy, this measure indicates the way sound level is distributed. Two different spectra can have the same energy in two bands but not necessarily the same CSPD (Ruiz and Legros, 1995). Therefore, it is preferable to examine the CSPD's two bands than the corresponding energy ratio.

Because of the zero-mean property of the all-pole model, this new method leads to mean sound level independent spectrum variability analysis. Derived-CSPD features are then directly related to frequency distribution of sound level:

- Δ is the area between low and high frequency CSPD with a frequency limit equal to the second antiresonance of each spectrum;
- F_s is the spectral balance frequency, i.e., the frequency for which Δ is minimal.

CSPD, Δ, F_s and spectral averaging are computed by a specific routine and LPC analysis is performed by ILS software.

Vowel sound level normally decreases with frequency leading to a greater area Δ between high and low frequency CSPD than when this general tendency is modified. For example, a shift in energy from low frequency band to high band induces a low value of Δ. F_s is the frequency for which Δ is minimal. In the general case of sound level diminution with frequency, F_s has a very low or high value: near 0 Hz or near the half sampling frequency. If a lower spectral slope is only due to a lower dynamic range of the envelope spectrum, F_s will continue to take extreme values. But if a lower spectral slope is due to a non-uniform decrease of sound level with frequency, for example in the case of higher levels in a frequency band, the balance can appear for middle frequencies. Studies of F_s and Δ are two equivalent ways to quantify CSPD changes. One with a fixed frequency limit between high and low bands (Δ), and one with a variable frequency method (F_s). However similar the two approaches are, the stress sensitivity of these parameters is unknown, this is why they are studied together.

3.2.2.2. Formant frequencies. For each vowel, its first, second and third formants are evaluated at its centre, by means of a KAY DSP 5500 analyzer, on the basis of wide band spectrograms, FFT spectra and LPC formant tracking routine.

Formant values in Hz have been converted into Mels prior to statistical treatment. Two-way *stress* × *vowel* analyses of variance have been performed in order to test the existence of stress-related formant variabilities, notwithstanding the effect of the vowel.

Three analyses of this kind were performed: one for each formant. In order to locate the effects, one-way analyses were also performed vowel by vowel, in case a significant stress effect had been observed in the two-way analysis.

In order to avoid tedious enumerating of statistical results, only tests for which the null hypothesis probability is small are reported in this paper. We moreover essentially concentrate on sources of variance dealing with stress, either as a main effect, or as a variable contributing to an interaction effect.

Given the exploratory nature of the paper, p-values that are less than or close to the traditional 0.05 threshold are reported. Results usually regarded as *weakly significant* ($0.10 > p > 0.05$), as well as *significant* ($p < 0.05$) are therefore presented here. In order to emphasize strong evidence of significance, we moreover use the asterisk code – * if ($0.05 > p > 0.01$), ** if ($0.01 > p > 0.001$) and *** if ($p < 0.001$) – which is traditionally used to symbolize degree of significance (Pedhazur and Pedhazur Schmelkin, 1991).

It is a paramount importance to notice, however, that the use of inferential testing in this kind of research is not properly directed towards a seek of generalisability in the usual sense. Since the experimental approach developped here is a within-subject one, significance indicates simply that random processes alone could not account for the variations observed in the behaviour of the particular individuals taken into account. Now, the aim of exploratory research is precisely to reveal the existence of particular phenomena and to decide whether they are worth studying by extended work aimed at studying their laws. Such investigations are interesting, since they allow the microscopic study of the effects of stress. Nevertheless, they do not provide information about the global dynamics of the reported effects. A macroscopic viewpoint can be adopted by considering each vocalic sound uttered by the speaker as a point in a multidimensional space (in this case, the $F1$–$F2$–$F3$ space). Then, the question of whether these vowel-points form clusters with variable degrees of consistency may be addressed.

One way of doing this is to test whether the distances to the centre of the formant space depend upon stress. This involves choosing a reference centre for the formant space. Schwa, a theoretical vowel with formant values regularly spaced in the frequency domain (500 Hz, 1500 Hz, 2500 Hz, ...) is often regarded as such. In order to cope with inter-speaker variabilities, it is also possible to evaluate each subject's own $F1$–$F2$–$F3$ space centre, statistically evaluated on the basis of a descriptive analysis of the subject's utterances.

4. Results

4.1. Fundamental frequencies

4.1.1. Macroscopic analysis

4.1.1.1. Laboratory stress (Fig. 1). During the first phase, F_0 is rather low and stable: the average frequency is 127 Hz. During the first two thirds of the second phase, F_0 has the same flat profile. Nevertheless, a gradual increase can be observed in the last third of the second phase, where the fundamental frequency culminates at 199 Hz. This late

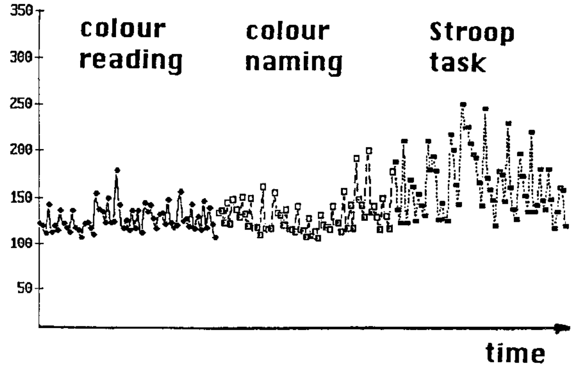

Fig. 1. Pitch frequencies (in Hz) of each selected vowel during the first (filled circles), second (open boxes) and third (filled boxes) phases of the Stroop test.

increase does not affect much the average F_0, which remains low (130.5 Hz), for the second phase taken as a whole. In the third phase, however, a dramatic increase is to be found: the average F_0 is 164.8 Hz, i.e. an increase of 29.8% between phase 1 and 3. The greatest pitch value observed is 250 Hz (during the third phase), i.e. a 96.8% increase, compared to the mean fundamental frequencies from phase 1.

A two-way analysis of variance was performed. Phases 1 to 3 form one independent variable, the four different vowels form another one, since vowel-related effects could be observed. The design is therefore of the form *phase* × *vowel*. The dependent variable is F_0 converted to Mels. A strong between-phase effect is revealed $(F = 21.03^*,\ p < 0.001)$, although no other significant effect can be observed.

The same treatment has been reproduced, but after dropping data from the third phase. In this case, no more effect is to be found $(F = 0.299,\ p = 0.590)$. The inferential analysis thus confirms that the observed frequency increase which characterizes phase 3, is significant over phases 1 and 2, and that these phases cannot be differentiated one from the other.

4.1.1.2. Real case stress (Fig. 2). F_0 values associated with levels 1 and 2 of the stress variable tend to be higher than those drawn from the rest periods. It can be easily observed that the fundamental frequency increases during the first flight incident, and becomes particularly high during the final phase of the flight.

A closer examination of the data nevertheless reveals that the two speakers do not show the same kind of reaction. For the pilot, F_0 increase is stronger under stress 1 condition (discussion relating to the first incident): 150 Hz, versus 117 Hz under stress 0 condition and 144 Hz under stress 2 condition, i.e. an increase of 23% between phases 0 and 2. On the contrary, the copilot does not show important F_0 increase during this discussion (153 Hz, versus 142 Hz under stress 0 condition), but he is characterized by a dramatic increase just before the crash: mean of 204 Hz under stress 2 condition, i.e., an increase of 43.6% compared to stress 0 condition.

For both speakers, there is a phase of the flight that provokes appreciably increased frequencies: phase 1 for the pilot and phase 2 for the copilot. During these phases, F_0 culminates at very high

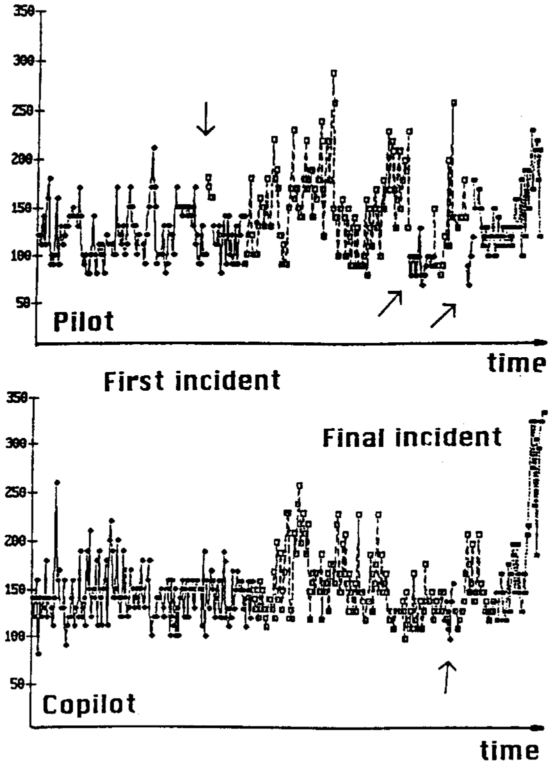

Fig. 2. Pitch frequencies (in Hz) of each selected vowel under stress condition 0 (filled circles), stress 1 conditions (open boxes) and stress 2 conditions (filled boxes) phases of the CVR recordings, both for the pilot and the copilot; small arrows emphasize transient states of variable stress.

values: 290 Hz for the pilot and 340 Hz for the copilot. These maximal frequencies respectively represent (compared to the mean F_0 from phase 0) 148% and 139% increases, relative to the quiet, initial period of the flight.

The same statistical treatment as the one performed on the Stroop data has been performed on these CVR values. Its results confirm the significance of the variations observed, for both speakers* ($p < 0.0001$). For the copilot, a significant *stress* × *vowel* interaction** ($p = 0.008$) is also revealed, suggesting that the magnitude of the observed effect depends on the vowel uttered.

4.1.2. Microprosodic analysis: the μ index

4.1.2.1. Laboratory stress. Under the Stroop test condition, μ index has values very close to 1 in phase 1 (average of 1.06, with standard deviation of 0.07). This indicates that within the observed vowels, the pitch tends to be rather flat. The values of the index increase slightly in phase 2 (average of 1.11, standard deviation of 0.09) and strongly in phase 3 (average of 1.32, standard deviation of 0.20). It therefore appears that variations of the μ index are in good agreement with expected values of a stress-dependent index (Fig. 3).

As for the F_0 values, analyses of variance have been performed on μ, which was regarded as the dependent variable of a *stress* × *vowel* factorial design. Their results reveal significant between-phase differences on the whole*** ($p < 0.001$), but no difference between phases 1 and 2.

In order to evaluate the discriminative power of the index, we moreover performed discriminant analyses with the phases as a priori categories. The μ index was considered as the discriminant variable. Once the discriminant functions had been obtained, we simulated a recognition task aiming at assigning each vowel to one of the three categories on the basis of discriminant functions. This treatment was firstly reproduced for each possible pair of phases. Each treatment delivered a percentage of correct recognition (phase 1 versus phase 2: 60.00%; phase 1 versus phase 3: 90.00%; phase 2 versus phase 3: 80.00%). The whole procedure was thereafter reproduced, but with F_0 as the discriminant variable. The correct recognition percentage are the following: phase 1 versus phase 2: 53.33%; phase 1 versus phase 3: 83.33%; phase 2 versus phase 3: 76.67%.

As can be seen, in all cases, the recognition rates are better with the μ procedure than with the F_0 one. In the case of phases 1–3 discrimination, the power of the pitch-based technique is 92% of the one of the μ procedure.

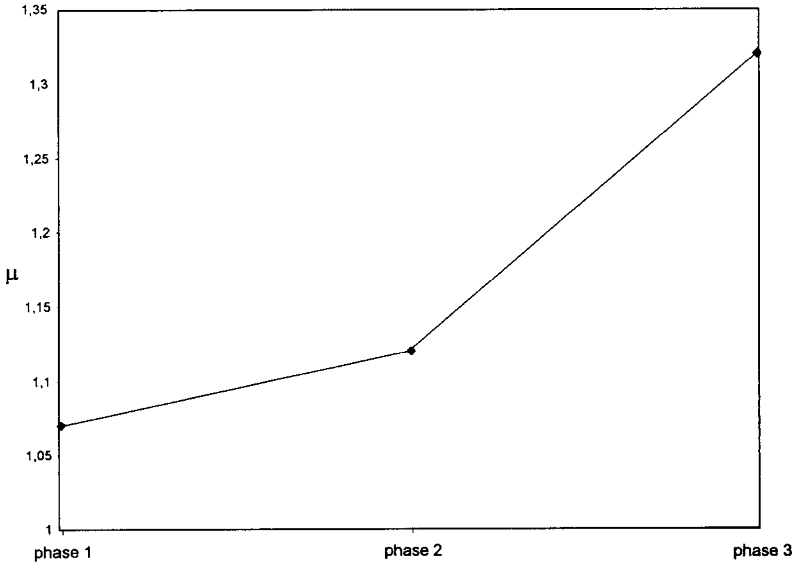

Fig. 3. μ mean values in the different phases of the Stroop test.

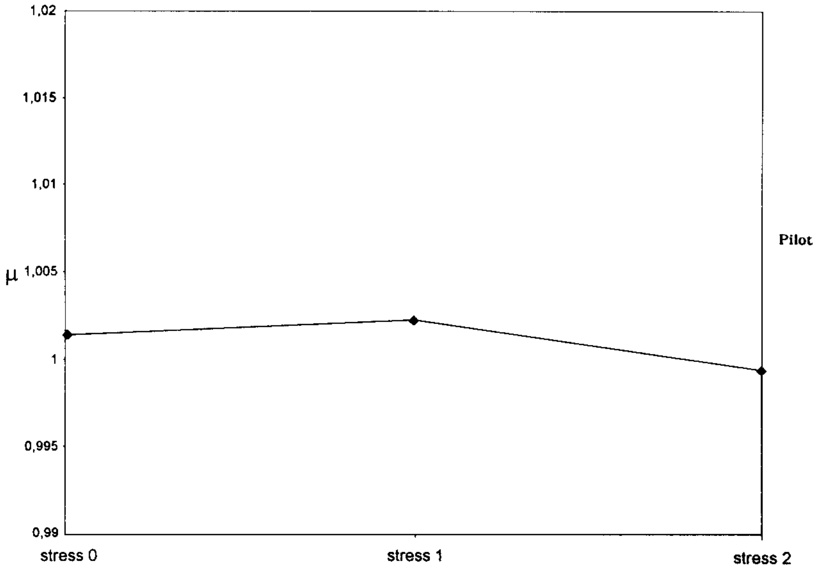

Fig. 4. μ mean values in the different stressful situations of the real case stress for the pilot.

4.1.2.2. Real case stress. Values of μ drawn from the CVR recordings are in all cases very close to 1 (the averages from stress 0 up to stress 2 are, for the copilot: 1.0008, 1.0020, 1.0163, and for the pilot: 1.0013, 1.0020 and 0.9996) (Figs. 4 and 5). This indicates that the within vowel pitch variations are few, if any. No differentiation between stress states can be made on the basis of the index for the pilot. Nevertheless, the analysis of variance reveals significant differences between stress states for the copilot

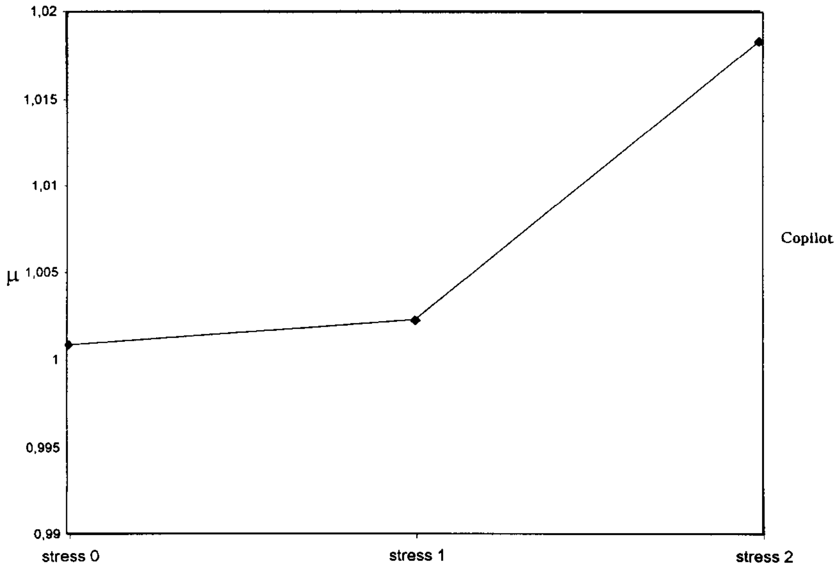

Fig. 5. μ mean values in the different stressful situations of the real case stress for the copilot.

($p = 0.007$). We must consider this result keeping in mind that even if the observed between-average-differences are not important, they may be great, relative to μ standard deviations which are very small.

As for the laboratory data, we simulated recognition tasks based upon both the μ index and the fundamental frequency. In the case of the CVR recordings, the F_0-based procedure however appeared to be more discriminant than the μ-based one.

4.2. Spectrum analysis

4.2.1. Frequency distribution of sound level

Steady-state French vowels are extracted and analysed throughout the recordings: /a/ for the CVR and /u/, /ϵ/, /o/, /ø/ for the Stroop test.

4.2.1.1. Δ area. Both for real and laboratory situations, variations of Δ do not indicate a unique direction of variation with time, i.e., with increasing stress degree. Low and high values of Δ exist, indicating motions of sound level distribution around the second antiresonance frequency.

Even if very low values are measured in the real situation, it is not possible to correlate them with stress arousal. The changes occur throughout the recordings (Figs. 6 and 7). In the case of the laboratory situation, lower Δ are measured at the beginning of the third phase of the stress test only for the colour "vert" (continuous line in Fig. 8) and "bleu" (dotted line in Fig. 8). For the two other colours no decreasing tendancy is observed (Fig. 9).

4.2.1.2. Spectral balance frequency F_s. The major finding concerning the spectral balance frequency F_s is its tendancy to take three types of value. For each type of vowel studied /a/, /ϵ/, /u/, /ø/, /o/, even if the spectrum form is different, three classes of F_s can be defined: two groups containing extreme F_s values (lower and higher) and a class with intermediary values.

Plotting F_s versus another characteristic, like for example the fundamental frequency F_0 of the same vowels, shows this property. In Figs. 10–12, these three groups are separated by vertical lines: 1000 and 3500 Hz for the CVR; 500 and 4500 Hz for the Stroop test.

It can be observed that higher values of F_0 do not belong to a unique group of F_s. This remark suggests that these two features act differently for the same stress factor. Vowels of the middle group of F_s are not only higher F_0 ones.

Third phase analysis of the Stroop test reveals that the F_s band between 500 Hz and 4500 Hz contains: 13/15 vowels for /u/, 10/15 for /ø/, 7/15 for /o/ and 6/15 for /ϵ/.

The same analysis for the CVR indicates that

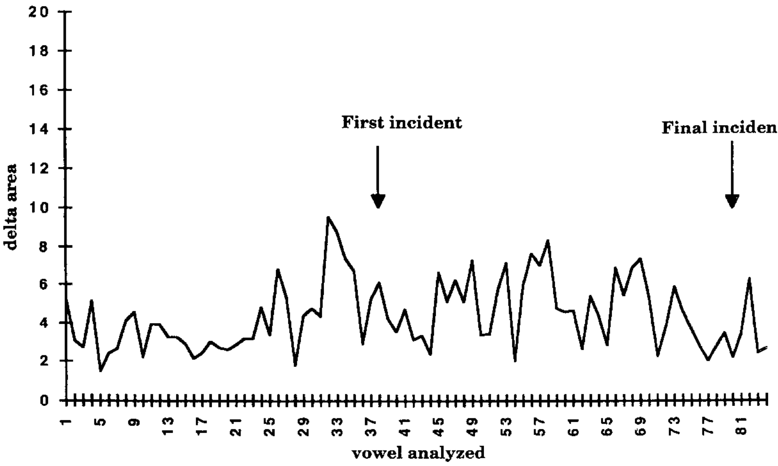

Fig. 6. Delta area (Δ) versus vowel /a/ uttered by the pilot of the real stress situation.

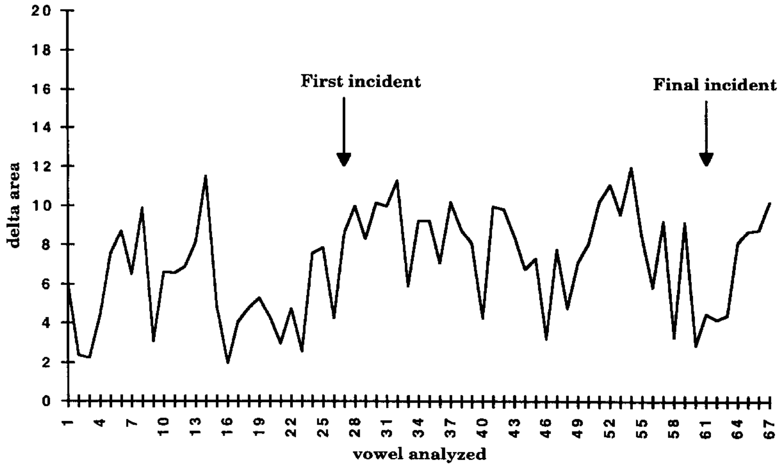

Fig. 7. Delta area (Δ) versus vowel /a/ uttered by the copilot of the real stress situation.

most of the /a/ (7/9) uttered by the pilot after the final incident belong to the middle band of F_s (1000 to 3500 Hz), but none of those uttered by the copilot. Concerning the first incident, 10 /a/ of 38 uttered by the pilot between the first and final incident take place in this band but none for the copilot. If belonging to the middle band can be considered as a vowel stress manifestation, such a result indicates that the pilot is much more stressed than his crew colleague, like F_0 results suggest.

This sensitivity of F_s to stress is confirmed by third phase analysis of the Stroop test. This phase is obviously a stressful situation and a lot of vowels have a middle F_s.

Therefore, spectral balance frequency seems to be related to the appearance of stress, probably in a different manner to fundamental frequency.

4.2.2. *Formant frequencies*

4.2.2.1. Laboratory stress. For the laboratory stress situation, formant analyses were performed on a

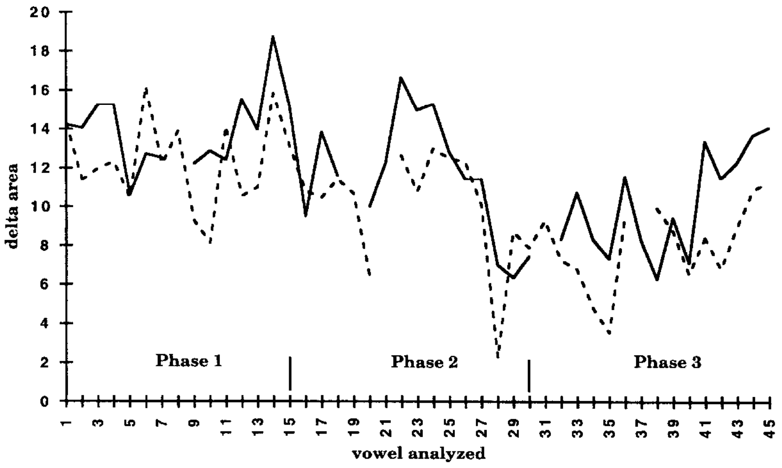

Fig. 8. Delta area (Δ) versus vowel (/ø/ dotted line and /ɛ/ continuous line) uttered by the speaker of the laboratory stress situation.

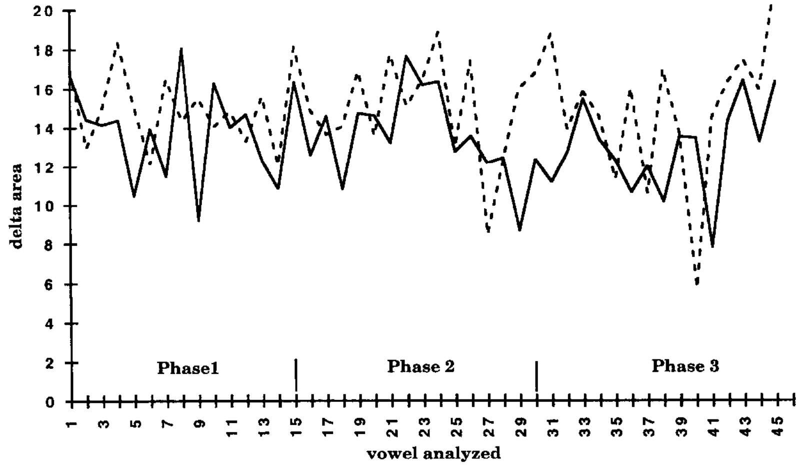

Fig. 9. Delta area (Δ) versus vowel (/o/ dotted line and /u/ continuous line) uttered by the speaker of the laboratory stress situation.

subsample of 60 vowels in each phase (out of 200 by phase), allowing context invariance for the selected vowels. After formant values convertion into Mels, the analyses of variance were applied. The results are presented in Table 1.

Considering all the phases of the experiment, significant differences appear in each formant. The first one is characterized by variations in /u/: the mean value increases between phases 1 and 2, then decreases between phases 2 and 3. First formant frequencies decrease for /ø/ and increase for /ɛ/ under the effect of stress. Variations in $F2$ are important in vowel /ɛ/ (Fig. 13). Finally, $F3$ variations are observed in some vowels: decrease for /u/, increase for /ɛ/ and increase after decrease for /ø/.

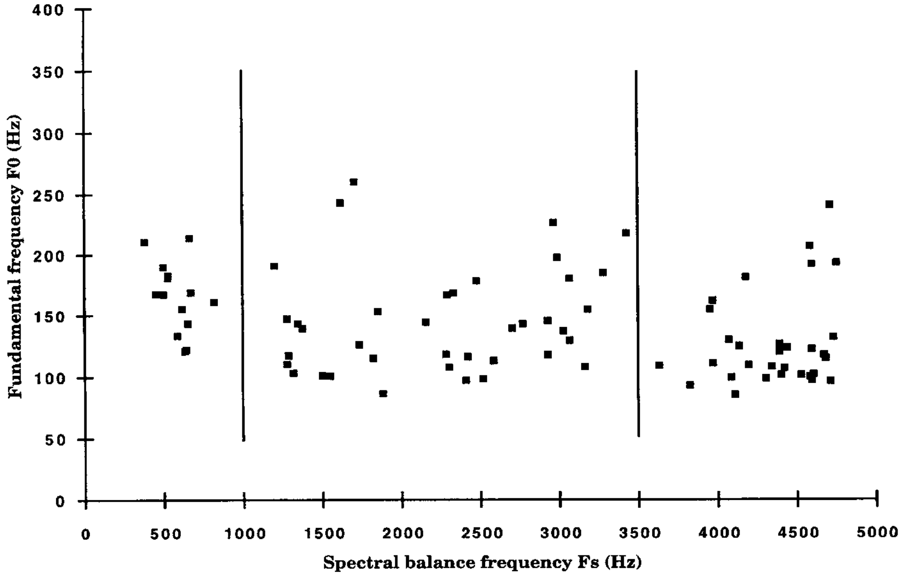

Fig. 10. Fundamental frequency F_0 (in Hz) versus spectral balance frequency F_s (in Hz) for the vowel /a/ uttered by the pilot of the real stress situation.

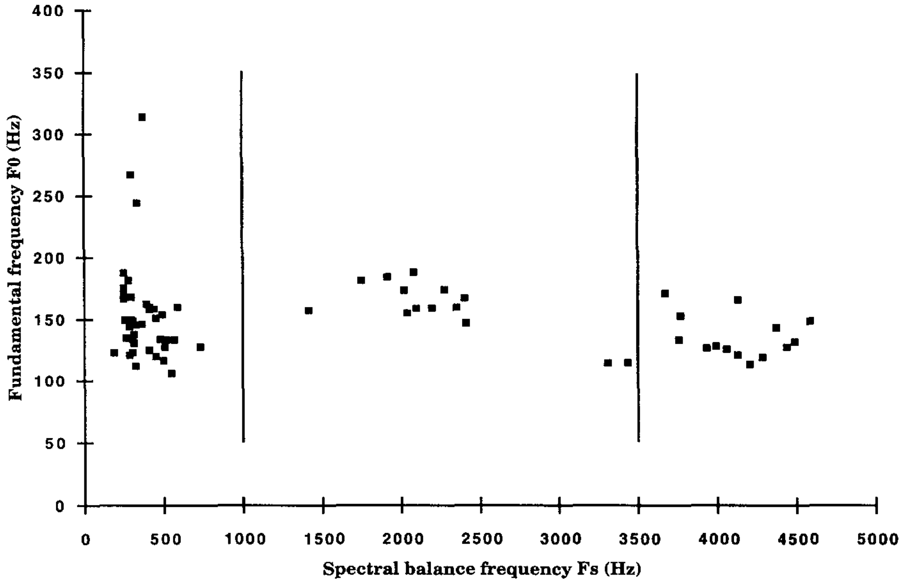

Fig. 11. Fundamental frequency F_0 (in Hz) versus spectral balance frequency F_s (in Hz) for the vowel /a/ uttered by the copilot of the real stress situation.

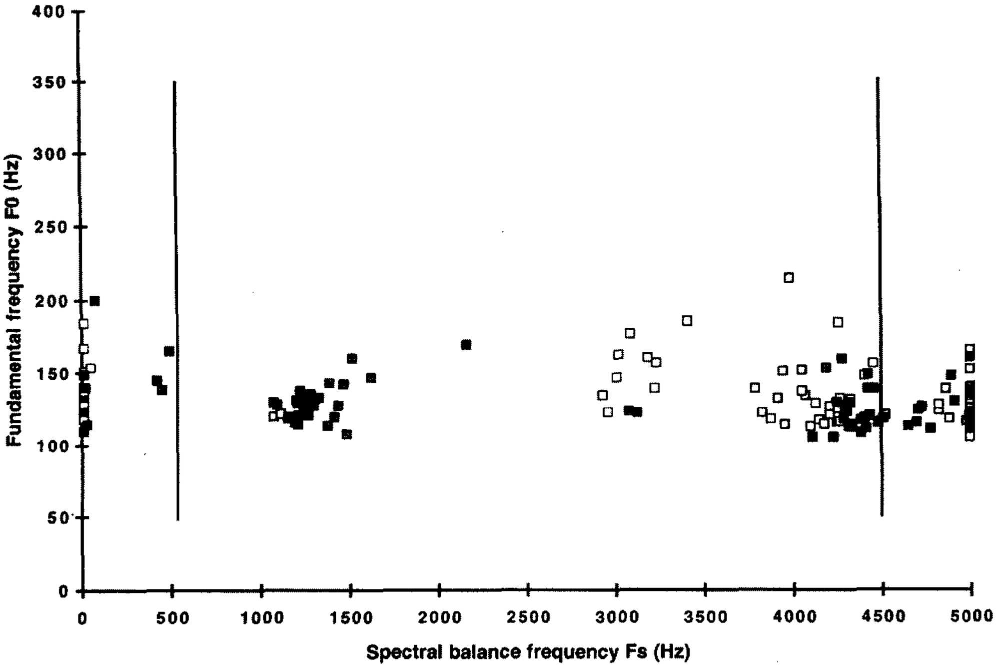

Fig. 12. Fundamental frequency F_0 (in Hz) versus spectral balance frequency F_s (in Hz) for the four vowels /ø/ (in blue), /o/ (in yellow), /ɛ/ (in green), /u/ (in red), uttered by the speaker of the laboratory stress situation.

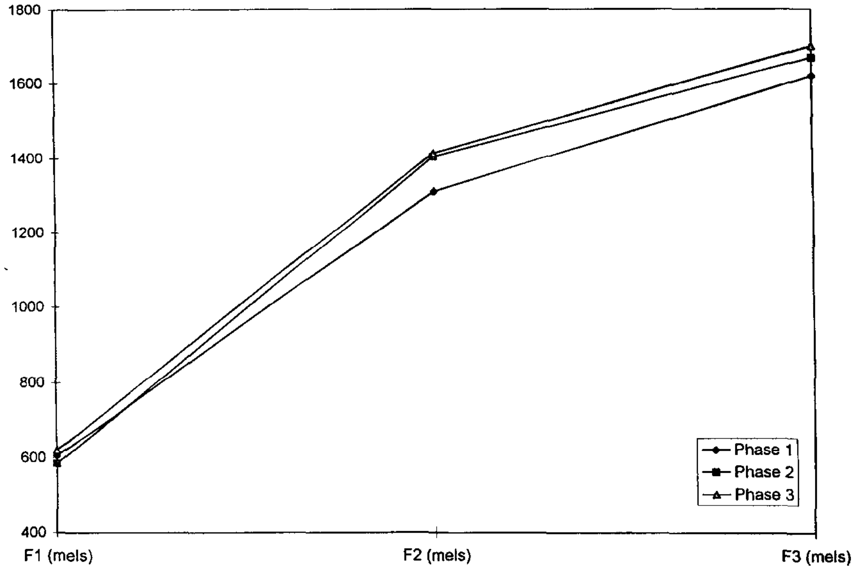

Fig. 13. Formant mean values (in Mels) for the vowel /ɛ/ in the different phases of the Stroop test.

Table 1
Probabilities of H_0 for the significant tests; the arrows indicate the direction of variation of the formant involved, from phase 1 to phase 2 (left arrow), and from phase 2 to phase 3 (right arrow)

$F1$	$F2$	$F3$
Phase ($p = 0.016$): /u/ ↗ ↘ /ø/ ↘ ↘ /ɛ/ ↗ ↗ Phase × vowel interaction: $p = 0.001$	Phase ($p < 0.001$): /ɛ/ ↗ ↗ Phase × vowel interaction: $p < 0.001$	Phase ($p = 0.083$): /u/ ↘ ↘ /ø/ ↘ ↗ /ɛ/ ↗ ↗ Phase × vowel interaction: $p < 0.001$

Table 2
Probabilities under H_0 (p) for the significant tests (arrows indicate the sense of the frequency variation under the effect of stress)

Main effects on	Pilot	Copilot
$F1$	NS	$p = 0.08$ /i/ ↗ /u/ ↘
$F2$	$p = 0.002$ /a/ ↗ /e/ ↗ /ɛ/ ↗	NS
$F3$	$p = 0.01$ /ɛ/ ↗ /ã/ ↗	$p = 0.005$ /u/ ↘
Interactions stress × vowel	Pilot	Copilot
$F1$	NS	$p = 0.08$
$F2$	$p = 0.03$	NS
$F3$	NS	NS

Table 3
Central individual frequencies: mean values (in Mels)

Speaker	$F1$	$F2$	$F3$
Pilot	605	1149	1626
Copilot	524	1209	1584

Table 4
Probabilities of H_0 for the analysis of variance

Independent variable	Pilot	Copilot
Stress	$p = 0.021$	$p = 0.556$
Interaction	$p = 0.539$	$p = 0.003$

Table 5
Synthesis of the results

Laboratory stress	Real stress
Mean F_0 per vowel	
Gradual significant increase: about 20% between rest and last stress phase and 96% for the greatest pitch value	Gradual significant increase: about 20% between no stress and final stress situation and 140% for the greatest pitch value
μ index	
Increase with test phases	Differentiation for the copilot
Δ area	
No direct relation between stress arousal and the Δ area variations	No direct relation between stress arousal and the Δ area variations
Spectral balance frequency F_s	
Belonging to the middle band of F_s seems to be related to a stress manifestation	
36/60 vowels of the more stressing phase are in the middle band	7/9 vowels of the final incident are in the middle band for the pilot and none for the copilot; 10/38 vowels of the first incident are in this band for the pilot and none for the copilot
Formant frequencies	
Significant variations of the three first ones (not for all the types of vowels studied) but variable sense of variation depending on vowel type	Significant variations exist but not for all the types of vowels, not for all the formant frequencies and with different senses of variation with stress arousal
Distances to the $F1$–$F2$–$F3$ space center	
No significant effect of the test phase	Significant effect for the pilot

Situation × *vowel* interaction is highly significant for each formant.

The treatment performed on data from situation 1 and 2 only shows very significant between-phase differences for the second formant. We observe in this case an increase of the mean values for the vowels /u/ and /ɛ/.

The treatment performed on data from situation 1 and 2 only shows very significant between-phase differences for the second formant. We observe in this case an increase of the mean values for the vowels /u/ and /ɛ/.

4.2.2.2. Real case stress. For the real situation stress, the analyses were performed on a restricted, balanced sample of words extracted from the three periods of the flight. The aim was to cope with the possible important effect of vowel phonetic environment. For that purpose, only vowels with the same anterior context in the three subsamples were taken into account, i.e., vowels are selected only if their consonant–vowel combinations appears in every part of the CVR corpus. The results are presented in Table 2.

For the pilot, one can note a significant effect of stress on the second formant. Detailed analyses reveal frequency increases in vowels /a/, /e/ and /ɛ/. A significant variation of $F3$ with a formant frequency increase for /ɛ/ and /ã/ is also to be noted. Finally, a significant interaction effect appears between stress and vowel variables in $F2$.

For the copilot, the analysis reveals a significant effect of variable stress on $F3$. In this case, third formant values of /u/ decrease. We also observe an effect on the first formant, consisting in frequency increase for /i/ and frequency decrease for /u/ values. A significant *stress* × *vowel* interaction is also observed, for the first formant.

A similar treatment, performed on data from stress levels 0 and 1, shows significant effects of stress on $F2$ for the pilot* ($p = 0.02$) and on $F3$ for the copilot* ($p = 0.046$).

4.2.3. Distances to the F1–F2–F3 space centre

4.2.3.1. Laboratory stress. The same treatments as those performed on CVR data have been applied on the Stroop test corpus. Nevertheless, since the recordings contained utterances of the French vowels /u/, /ø/, /ɛ/ and /o/ only, the theoretical Schwa vowel had to be considered the centre of the $F1$–$F2$–$F3$ space.

Data processing consists in 2-way analyses of variance, with phase and vowel as independent variables, and the Euclidean distances of the analysed sounds to the $F1$–$F2$–$F3$ space centre as the dependent variable.

No significant effect of the phase is revealed by the analysis of variance ($p = 0.41$). The phase–vowel interaction effect is however very significant*** ($p < 0.001$); this can be interpreted in terms of effects in particular vowels under the Stroop situation.

4.2.3.2. Real case stress. Values of the first three formants of vowels /œ/ and /ə/ have been used to approximate an $F1$–$F2$–$F3$ vowel space centre (Table 3).

The same treatments as those performed on laboratory data have been applied on the Stroop test corpus. The results are presented in Table 4.

For the pilot, they indicate an influence of the more stressful situation of the flight on the distance to centre parameter. No stress–vowel interaction is noticed here. A reversed configuration can be observed in the copilot's data (no significant effect of stress, but a significant *stress* × *vowel* interaction).

5. Conclusion

The results presented in this paper confirm that some speech characteristics suffer from the arousal of stress. The phenomenon has been observed in artificial as well as in real situations (Table 5).

This observation is rather interesting, because experiments focused both on real and artificial stressing conditions are few. Laboratory phenomena thus do not seem so far from the real ones, concerning the observed variations from phase to phase in each case, all the more that the stress induced in this real situation can be considered as extreme. It seems therefore that artificial stressing conditions are able to lead the subject to produce vocal changes similar to those obtained in natural situations.

This fact rather pleads for the generalizability of laboratory experiments, for which criticism about their validity is often objected.

Some acoustical parameters used have moreover demonstrated sensitivity to the psychological factors involved in our study.

All our analyses proved able to discriminate different degrees of stress on the basis of F_0, whatever the speaker (ordinary speaker versus trained professionals) and the situation involved.

The maximum pitch increase, in the case of the plane crash, is greater than the one in the laboratory situation (148% against 96.8%). But if average F_0 is considered, the results show less proportional increase between phase 1 and 3 in the Stroop situation, and between phase 0 and 2 for the two speakers in the real situation. One can note that such modifications contribute less to distinct stress degrees than one by one vowel study of the mean F_0.

This observation suggests that an investigation that would pay more attention to local than to global changes could be of some interest. This idea seems to be supported by findings about the μ index, at least in the case of the Stroop corpus, where significant between-phase differences are revealed by ANOVA and discriminant analyses. The increase of the index value is obvious in the case of the Stroop situation (24%). It is however quite less dramatic in the case of the CVR recordings (1.54% for the copilot, no change for the pilot); the ANOVA procedure nevertheless confirms significance** ($p = 0.007$) in the case of the copilot.

The μ index therefore seems sensitive, principally in light-stress situations (such as in the Stroop task). In case of very heavy stress (fear of an imminent death), the fundamental frequency itself remains more efficient. A ceiling effect probably affects μ when the fundamental frequency reaches values so high that no more F_0 variation can be achieved. From this point of view, μ therefore can be considered as a promising tool for the detection of arousal stress, although F_0 increase remains the best cue to major reactions to perceived aggression.

Spectral data also appear to convey stress-dependent information. In the case of Δ, nevertheless, no systematic variation exists either for the real situation or for the laboratory one.

A stressing condition (flight incident, cognitive conflict) does not necessarily imply an increase or a decrease of the Δ area. Variations are not systematic and so, not closely linked with the stressful condition. On the contrary, study of F_s gives promising results. Firstly, its variations allow one to set up two groups of values (middle and extreme) which seem to be correlated respectively with stress and no stress vocal manifestations, especially for the laboratory situation. Secondly, this property occurs for different types of steady-state vowels, leading to think that this spectral characteristic seems to be free of vowel effect. This assumption must, of course, be verified by further experiments and reconsidered on the basis of theoretical speculations about F_s versus spectral ones. Now, it remains to develop this derived CSPD characteristics to prove its vowel independence, in regard to stress detection.

The formants-based analyses reveal significant between-phases differences, for the copilot, as well as for the pilot in the CVR recordings. These differences are nevertheless located in specific vowels. The same observations can be made for the Stroop corpus. Stress-related effects also appear when applying ANOVA analysis to the distance-to-centre parameter. The vowel influence is also quite obvious in this case, since the *stress* × *vowel* interaction turns out to be significant in the Stroop data, as well as for the copilot in the CVR recordings; only the pilot shows a stress effect free of vowel influence.

Further research should therefore investigate more deeply the role of the segment taken into account, and try to point out specific points of the speech chain that are more likely to reveal anatomic, physiologic or behavioural changes linked to stress.

Whatever the point of view taken into account (excitation or resonance), the recordings thus appear to convey acoustical cues to stress. The parameters used nevertheless show various kinds of behaviours, in reaction to various kinds of situations. The absolute value of F_0, for instance, has been regarded as a good index to reveal heavy stress, although the μ parameter rather seems useful for the detection of arousing stress. This observation emphasizes the importance of further research aimed at exploring the physical determinants of the reported variations. They could enlighten the development of new, compound, indices, built up in a multivariate framework, and able to take into account the various aspects or stress induced variations of the speech signal.

Notwithstanding indices variability, the results also show important between-speaker variations. In the case of the plane crash, e.g., the pilot and the copilot obviously do not have similar reactions to the events they have to cope with. A possible interpretation is the following: the pilot is working with the commands in his hands, while the copilot has only a remote view on the phenomenon. The pilot therefore receives physical, proprioceptive stimuli that do not reach the copilot. Moreover, being above the pilot in rank, the copilot probably tries to appear as the moderator in the situation.

Although the pilot develops anxiety on the occasion of the first incident, the copilot is actually afraid only during the very last seconds, at a time when the information about imminent danger is as obvious for him as for the pilot.

This functional, situational explanation emphasizes the interest of investigations taking into account variables usually considered in pragmatics and/or sociolinguistics, such as the characteristics of the communication situation, the number of speakers involved, their mutual relationship, the tasks they are faced with, etc.

Moreover, if stress as different as the fear of an imminent death and the disturbances resulting from of a psychometric test turn out to cause similar emotional manifestations, it might be interesting to consider resistance to stress, rather than stress itself as a criterion for further research.

This remark addresses, in turn, the question of how to assess the individual, intrinsic resistance, as well as the one of how to define the magnitude of stress. In the very case of our experiments, it seems reasonable to consider that the pilot and the copilot, who were well-trained professionals, were likely to respond to stress in a different way than the non-trained ordinary subject undergoing our laboratory experiment. On the other hand, in the real case, study of phase by phase variations seems closely linked to the choice of the definition and limits of the different phases. Actually, this problem does not appear in the Stroop situation because the phase cutting out is completed beforehand.

All those remarks clearly emphasize the need of a closer examination of the concept of stress itself. The experiments, their methodologies, and finally their results will vary quite a lot according to whether stress is regarded as a quantum of aggression, a given reaction to aggression, a specific personal trait, a psychological state, or anything else.

Together with the development of speculative reasoning about the mathematics of new indices of stress, elaboration of a psychologically-based framework for the concept of stress therefore appears as a *sine qua non* contribution for the improvement of research in the field.

Acknowledgements

Thanks are due to the "Service Technique de la Navigation Aérienne", who provided the CVR corpus.

References

M.T. Allen and M.D. Crowell (1989), "Patterns of autonomic response during laboratory stressors", *Psychophysiology*, Vol. 26, pp. 603–614.

P. Benson (1995), "Analysis of the acoustic correlates of stress from an operational aviation emergency", *Proc. ESCA-NATO Tutorial and Research Workshop on Speech under Stress*, Lisbon, September 1995, pp. 61–64.

J.W. Black (1951), "The effect of noise induced temporary deafness upon vocal intensity", *Speech Monogr.*, Vol. 18(b), pp. 74–77.

M. Brenner, T. Doherty and T. Shipp (1994), "Speech measures indicating workload demand", *Aviat. Space Environ. Med.*, Vol. 65, pp. 21–26.

E.T. Doherty and H. Hollien (1978), "Multiple-factor speaker identification of normal and distorded speech", *J. Phonetics*, Vol. 6, pp. 1–8.

G. Fairbanks and L.W. Hoaglin (1941), "An experimental study of the duration characteristics of the voice during the expression of emotion", *Speech Monogr.*, Vol. 8, pp. 85–90.

B. Gramatica, R. Ruiz and C. Legros (1992), "Modification de la fréquence fondamentale de la voix des pilotes: Incidents réels et simulés", *J. Physique III*, Vol. 2, pp. C1-335-338.

G.R. Griffin and C.E. Williams (1987), "The effects of different levels of task complexity on three vocal measures", *Aviat. Space Environ. Med.*, Vol. 58, pp. 1167–1170.

W.A. Hargreaves and J.A. Starkweather (1964), "Voice quality changes in depression", *Language and Speech*, Vol. 7, pp. 84–88.

B. Harmegnies and A. Landercy (1992), "A multi-variate approach for the analysis of speech under cognitive stress", *Proc. ESCA*, Cannes-Mandelieu, pp. 231–234.

M.H.L. Hecker, K.N. Stevens, G. Von Bismark and C.E. Williams (1968), "Manifestation of task induced stress in the acoustic speech signal", *J. Acoust. Soc. Amer.*, Vol. 44, No. 4, pp. 993–1001.

H. Helfrich, R. Standke and K.R. Scherer (1984), "Vocal indicators of psychoactive drug effects", *Speech Communication*, Vol. 3, No. 3, pp. 245–252.

J.C. Junqua (1992), "The variability of speech produced in noise", *Proc. ESCA*, pp. 43–52.

J.C. Junqua (1993), "The Lombard reflex and its role on human listeners and automatic speech recognizers", *J. Acoust. Soc. Amer.*, Vol. 93, No. 1, pp. 510–524.

H. Lane and B. Tranel (1971), "The Lombard sign and the role of hearing in speech", *J. Speech and Hearing Research*, Vol. 14, pp. 679–709.

E. Lombard (1911), "Le signe de l'élévation de la voix", *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, Vol. 37, pp. 101–119.

J.D. Mosko, K.N. Stevens and G.R. Griffin (1983), Interactive voice technology: Variations in the vocal utterances of speakers performing a stress-inducing task, Naval Aerospace Medical Research Laboratory Report 1300, Pensacola, FL.

S.S. Newman and V.G. Matter (1938), "Analysis of spoken language of patients with affective disorders", *Amer. J. Psychiatry*, Vol. 94, pp. 912–942.

P.F. Ostwald (1965), "Acoustic methods in psychiatry", *Scientific American*, Vol. 212, pp. 82–91.

E. Pedhazur and L. Pedhazur Schmelkin (1991), *Measurement, Design and Analysis* (Lawrence Erlbaum Associates, Hillsdale, NJ).

A. Protopapas and P. Lieberman (1995), "Effects of vocal F_0 manipulations on perceived emotional stress", *Proc. ESCA-NATO Tutorial and Research Workshop on Speech under Stress*, Lisbon, September 1995, pp. 1–4.

L. Rubenstein (1966), "Electro-acoustical measurement of vocal responses to limited stress", *Behav. Res. & Therapy*, Vol. 4, pp. 135–138.

R. Ruiz and C. Legros (1994), "The cumulative spectral probability diagram: Theory and experiments", *Acta Acustica*, Vol. 2, No. 3, pp. 215–222.

R. Ruiz and C. Legros (1995), "Vowel spectral characteristics to detect vocal stress", *Proc. 15th Internat. Congress on Acoustics*, Trondheim, Norway, Vol. 3, pp. 141–144.

R. Ruiz, C. Legros and A. Guell (1990), "Voice analysis to predict the psychological or physical state of a speaker", *Aviat. Space Environ. Med.*, Vol. 61, pp. 266–271.

K.R. Scherer (1981), "Vocal indicators of stress", in: J. Darby, Ed., *Speech Evaluation in Psychiatry* (Grune and Stratton, New York), pp. 189–220.

P.V. Simonov and M.V. Frolov (1973), "Utilization of human voice for estimation of man's emotional stress and state of attention", *Aerospace Med.*, Vol. 44, No. 3, pp. 256–258.

P.V. Simonov and M.V. Frolov (1977), "Analysis of the human voice as a method of controling emotional state: Achievments and goals", *Aviat. Space Environ. Med.*, Vol. 46, pp. 1014–1016.

J.R. Stroop (1935), "Studies of interference in serial verbal reactions", *J. Exp. Psychol.*, Vol. 18 pp. 643–662.

J. Sulc and V. Remek (1985), "Possibilities and limits of using speech signals in aviation and space psychophysiology", *Acta Neurobiological Exp.*, Vol. 46, pp. 347–352.

J. Thompson and J.S. Mason (1995), "Effects of anxiety in visual and audio speech databases'', *Proc. ESCA-NATO Tutorial and Research Workshop on Speech under Stress*, Lisbon, September 1995, pp. 21–24.

J.C. Webster and R.G. Klumpp (1962), ''Effects of ambiant noise and nearby talkers on a face-to-face communication task'', *J. Acoust. Soc. Amer.*, Vol. 34, pp. 936–941.

C.E. Williams and K.N. Stevens (1972), ''Emotions and speech: Some acoustical correlates'', *J. Acoust. Soc. Amer.*, Vol. 52, pp. 1238–1250.

C.E. Williams and K.N. Stevens (1981), ''Vocal correlates of emotional states'', in: J. Darby, Ed., *Speech Evaluation in Psychiatry* (Grune and Stratton, New York).