ESCA-NATO Workshop on
Speech under Stress
Lisbon, Portugal
September 14-15, 1995

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

E. Absil[1], B. Grammatica[2], B. Harmegnies[1], C. Legros[2], D. Poch[3], R. Ruiz[2]

[1]Département de communication parlée, Université de Mons-Hainaut, 18, place du parc, B-7000 Mons, Belgium.
[2]Laboratoire d'acoustique, Université de Toulouse-le-Mirail, 5, allées Antonio Machado, 31058 Toulouse Cedex, France.
[3]Laboratori de Fonètica, Universitat Autònoma de Barcelona, Campus Bellaterra, 08193 Bellaterra, Spain.

ABSTRACT

The paper presents pitch-based analyzes 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 CVR of a crashed airplane. Analyzes relying on pitch are presented, and an index of microprosodic variation, μ, is introduced. The discriminating powers of the parameters used are studied in both situations by means of inferential statistics and discriminant analyzes. The results confirm the validity of laboratory experiments on stress, but emphazise quantitative as well as qualitative differences between the situations and speakers involved.

1. INTRODUCTION

Research dealing with effects of stress on the speech signal has taken into account a large variety of stressing factors. Experimental devices using different strategies supposed to artificially cause stress have been set up [1,2,3]. Various real events (specially aircraft incidents) have also constituted opportunities for collecting significant samples of stressed voices [4,5].

Nevertheless, most of the papers focus on a single stressing situation. It is therefore quite difficult both to check for the validity of laboratory induced stresses, and to correlate findings from a specific experiment with findings about the same parameter drawn from other experiments.

In this paper, we will therefore study both laboratory- and real event drawn recordings, in order to apply the same treatments to both kinds of corpora. Our analyzes will be based upon pitch, which as revealed to be highly sensitive to stress in previous experiments.

2. EXPERIMENTAL CONDITIONS

Two corpora will be analyzed. The first one has been obtained from a laboratory situation: it derives from the application of a French version of the so called "Colour-Word Test" (1935) by Stroop; 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 colours names as quickly as he can. During the first two phases, the subject has firstly to read colours names, then to name the colours of little squares. For the third part of the test, he is given 50 rows of 4 colours names each. This time, the words are nevertheless no more written in black on white: the ink used for each word is of a different colour than the one referred by the word itself. The subject has to tell the word's ink colour. The third phase is considered as providing a stressing 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 a reading material.

The second corpus is drawn from a real case: it is extracted from the Cockpit Voice Recorder (CVR) of a crashed aircraft. The pilot's and the copilot's utterances have been analyzed. The timing of the extract firstly shows the occurence 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

For the laboratory stress situation, formants analyzes were performed on a subsample of 60 vowels in each phase (out of 200 by phase), i.e., a total of 180 vowels, allowing context invariance for the selected vowels. For the real situation stress, the analyzes were 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 were taken into account: the vowel is selected under the requirement that the consonant-vowel combination appears in all three parts of the CVR corpus. The sample size is 366, for the copilot, and 307, for the pilot.

For the laboratory corpus, pitch values were determined by means of the F0 routine of a KAY CSL 4300 analyzer. For the CVR recording, both cepstrum-based measurements by ILS software, and evaluation on the basis of narrow band spectrograms (on a KAY 5500 DSP analyzer) were used.

The results will first be analyzed in terms of the observed frequency values. Inferential treatment will then be applied on frequencies transformed to Mel values.

We will finally study the discriminating ability of new parameter, μ, which assesses the amount of microprosodic variation inside a voiced sound, following equation 1:

$$\mu = \frac{f_c}{\frac{f_i + f_f}{2}} \qquad (1)$$

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 center is higher than those at the boundaries. This pattern seems to be frequent under certain forms of stress, on the basis of clinical observations.

4. RESULTS

4.1. Fundamental frequencies

4.1.1. Laboratory stress

During the first phase, F0 is rather low and stable: the average frequency is 127 Hz. During the first two thirds of the second phase, F0 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 increase does not affect a lot the average F0, which remains low (130.5 Hz), for the second phase taken as a whole. In the third phase, nevertheless, a sensible increase is to be found: the average F0 is 164.8 Hz. The greater pitch value observed is 250 Hz (during the third phase), i.e., a 195% increase, relative to the the fundamental frequencies from phases 1 and 2.

After fundamental frequencies conversion into Mels, a 2-way analyzis of variance has been performed. The dependent variable is the fundamental frequency expressed in Mels. Phases 1 to 3 form one independent variable, as well as the four different vowels form another one, since vowel-related effects could be observed. The design is therefore of the form *phase* X *vowel*. A strong between-phase effect is revealed ($F=21.03$, $p < .001$), although no other significant effect can be observed.

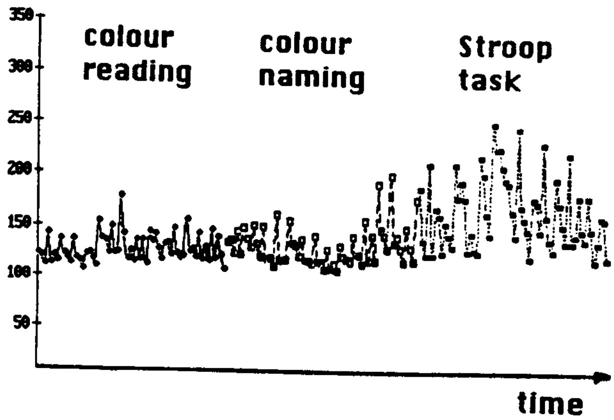

Figure 1: pitch frequencies (in Hz) of each selected vowel during the first (filled circles), second (open boxes), and third (filled boxes) phases of the Stroop test.

The same treatment has been reproduced, but after dropping data from the third phase. In this case, no more effect is to be found ($F=.299$, $p=.590$). The inferential analyzis thus confirms that the observed frequency increase which characterizes phase 3 is significant over phases 1 and 2, and that these phases cannot be differenciated one from another.

4.1.2. Real case stress

F0 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 is increased during the first flight incident, and 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 reactions. For the pilot, F0 increase is stronger during the discussion relating to the first incident (150 Hz, vs 117 Hz under stress 0 condition and 144 Hz under stress 2 condition). On the contrary, the copilot does not show important F0 increase during this discussion (153 Hz, vs 142 Hz under stress 0 condition), but he is characterized by a dramatic increase just before the crash (204 Hz under stress 2 condition).

For both speakers, there is a phase of the flight that provokes very high frequencies: phase 1 for the pilot, and phase 2 for the copilot. During these phases, their pitches culminate at very high values: 290 Hz, for the pilot, and 340 Hz for the copilot. These frequencies represent 248% and 239% increases, relative to the quiet, initial period of the flight.

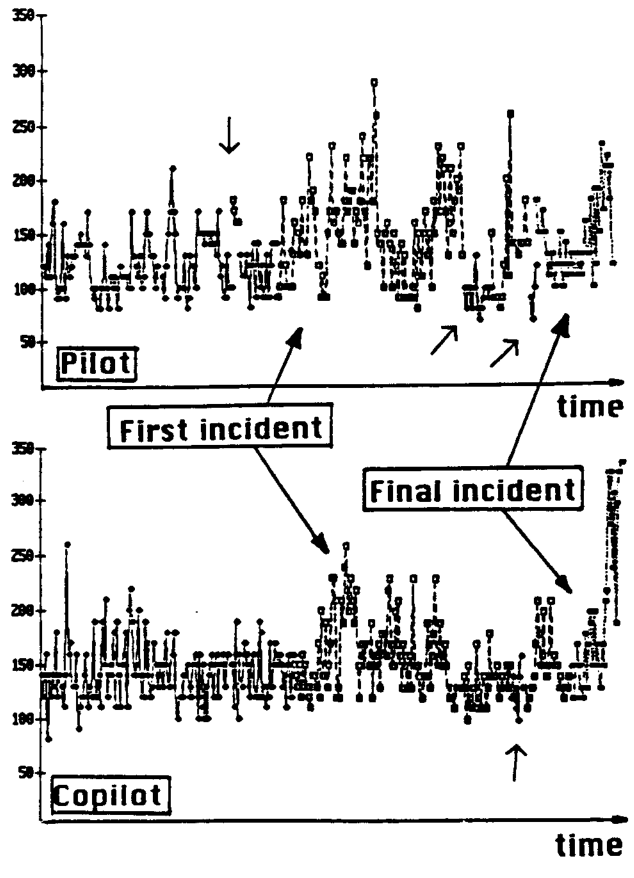

Figure 2: pitch frequencies (in Hertz) of each selected vowel under stress 0 condition (filled circles), stress 1 condition (open boxes), and stress 2 condition (filled boxes) phases of the CVR recordings, both for the pilot and the copilot; small arrows emphasize transient states of variable stress.

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 ($p<.0001$). For the copilot, a significant stress X vowel interaction ($p=.008$) is also revealed, suggesting that the magnitude of the observed effect depends on the vowel uttered.

4.2. The μ index

4.2.1. 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 .07). This indicates that within the observed vowels, the pitch tends to be rather flat. The values of the index increase sligthly in phase 2 (average of 1.11, standard deviation of .09) and strongly in phase 3 (average of 1.32, standard deviation of .20). It therefore appears that variations of the μ index are in good agreement with expected values of a stress-dependent index.

As for the F0 values, analyzes of variance have been performed on μ, which was regarded as the dependent variable of a stress X vowel factorial design. Their results reveal significant between-phase differences on the whole ($p<.001$), but no difference between phases 1 and 2.

In order to evaluate the discriminant power of the index, we moreover performed discriminant analyzes 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 the discriminant functions. This treatment was firstly reproduced for each possible pair of phases. Each treatment delivered a percentage of correct recognition (phase 1 vs. phase 2: 60.00%, phase 1 vs. phase 3: 90.00%, pase 2 vs. phase 3: 80.00%). The whole procedure was thereafter reproduced, but with F0 as the discriminant variable. The correct recognition percentage are the following: phase 1 vs. phase 2: 53.33%, phase 1 vs. phase 3: 83.33%, pase 2 vs. phase 3: 76.67.

As can be seen, differenciation between phases 1 and 3 is achieved with 90% recognition rate. In all cases, the recognition rates are better with the μ procedure than with the F0 one. In case of phases 1-3 discrimination, the power of the pitch-based technique is 92% of the one of the μ procedure.

4.2.2. 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 .9996). This indicates that the within vowel pitch variations are few, if any. No between-stress states differenciation can be made on the basis of the index for the pilot. Nevertheless, the analyzis of variance reveals significant differences between stress states for the copilot (p=.007). This astonishing result must be considered keeping in mind that even if the observed between average differences are not important, the μ standard deviations are very little.

As for the laboratory drawn data, we simulated recognition tasks based both upon the μ index and the fundamental frequency. In the case of the CVR recordings, the F0-based procedure however appears to be more discriminant than the μ-based one.

5. CONCLUSION

The results presented in this paper confirm the sensitivity of the pitch-based measures of stress. All our analyzes proved able to discriminate different degrees of stress on the basis of F0, whatever the subject and the situation involved.

This is an interesting observation, because confronting such situation in a single experiment is rather original. It appears very clearly that pitch-related differences can be observed in real- as well as in laboratory situations. This fact rather pleads for the generalizability of laboratory experiments, for which criticism about their ecological validity is frequently objected. Between situation differences are nevertheless obvious: 1. the pitch increase in the case of the plane crash is greater than the one in laboratory situation (245% against ca 195%); 2. the microprosody variations are more important in the laboratory- than in the real case situation.

These differences are probably linked to differential characteristics of the situation, that can be described in terms of linguistics-, pragmatics- or psychology based approaches(*in the Stroop situation there is one speaker only, in the flight incident situation, there are two speakers; in the Stroop situation, speech is not used to communicate, in the flight incident situation speech is used to communicate; in the Stroop situation, the speaker does not fear for his own life, in the flight incident situation he does; in the Stroop situation, the arousal of the stressor is controlled, in the flight incident situation it is not; in the Stroop situation, speech is the only tool to solve the problem, in the flight incident situation, there are many means of trying to solve the problem; in the Stroop situation the differences between phases are clear-cut, in the flight incident situation they are not; in the Stroop situation, the stress produced has a cognitive source only, in the flight incident situation the stress produced has, among others, an emotional origin; the speaker in the Stroop situation does not use the same variety of French than those in the flight incident situation, etc.*). Further research should carry on systematic, intensive investigation of the consequences of such differences.

The study of the CVR recordings also emphasizes the importance of subject-dependent variations. Obviously, the pilot and the copilot do not show the same reactions to the events they are faced 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 receive 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 an imminent danger is as obvious for him as for the pilot.

Finally, this research shows the interest of microprosodic examination of speech. The μ index reveals very sensitive, principally in medium-stress situations (such as in 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 F0 variation can be achieved. From this point of view, μ therefore appears as a promising tool for the detection of arousing stress.

6. REFERENCES

[1] M.BRENNER, T.DOHERTY, T.SHIPP, *Speech Measures Indicating Workload Demand*, Aviat. Space Environ. Med. 1994, 65:21-6.
[2] G.R.GRIFFIN, C.E.WILLIAMS, *The effects of different levels of task complexity on three vocal measures*, Aviat. Space Environ. Med. 1987, 58,1167-1170.
[3] B.HARMEGNIES, A.LANDERCY, *A multi-variate approach for the analyzis of speech under cognitive stress*, Proceedings of ESCA, Cannes-Mandelieu, 1992, 231-234.
[4] J.SULC, V.REMEK, *Possibilities and limits of using speech signals in aviation and space psychophysiology*, Acta Neurobiological exp., 1985, 46, 347-352.
[5] B.GRAMATICA, R.RUIZ, C.LEGROS, *Modification de la fréquence fondamentale de la voix des pilotes: incidents réels et simulés*, Journal de Physique IV, 1992, C1-335-338.

7. ACKNOWLEDGMENTS

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