ACOUSTIC CORRELATES OF TASK LOAD AND STRESS


Department of Psychology
University of Geneva, Switzerland
Klaus.Scherer@pse.unige.ch

ABSTRACT

It is argued that reliable acoustic profiles of speech under stress can only be found if different types of stress are clearly distinguished and experimentally induced. We report first results of a study with 100 speakers from three language groups, using a computer-based induction procedure that allows distinguishing cognitive load due to task engagement from psychological stress. Findings show significant effects of load, and partly of stress, for speech rate, energy contour, F0, and spectral parameters. It is further suggested that the mean results for the complete sample of speakers do not reflect the amplitude of stress effects on the voice. Future research should isolate and focus on speakers for whom the psychological stress induction has been successful.

1. INTRODUCTION

Since antiquity the voice has been considered a reliable readout of the speaker's affective arousal, particularly suited for the communication of emotion. Beginning in the early 20th century, researchers have attempted to identify the acoustic profiles of the major emotions (see [1] for a review), a quest that is starting to yield reliable results [2]. When the notion of "stress" became a fashionable concept in the fifties and sixties, researchers became interested in identifying the acoustic correlates of stress (see [3] for a review).

Deviations of fundamental frequency (F0) parameters (mean and variability) from baseline has been found to be the most reliable indicator of stress, as shown by a large majority of studies in the field [3, 4]. Similar effects have been found in studies of the magnitude and duration of spoken utterances. Mean envelope amplitude is generally greater, and mean utterance duration lower, for highly stressed subjects [3, 5]. In addition, formant structure seems to change under stress [5].

While some of these results seem relatively stable, allowing replication across studies, there is little evidence for a general "acoustic stress profile." As Protopapas and Lieberman [4, p. 2267] state: "...in all of the aforementioned studies it was evident that the acoustic correlates of emotion in the human voice are subject to large individual differences (i.e., among speakers). Streeter et al. (1983) concluded that there are no "reliable and valid indicators of psychological stress" (p. 1359)." Murray et al. [6, p. 12] summarize the insights gained during a workshop on "Speech under stress" as follows: In conclusion: Stress and its effect on speech seems to be a very complicated area of study which is very poorly understood at present."

This state of affairs is hardly surprising since this area of study is beset by two major problems: 1) The existence of many different kinds of stress and the absence of clear conceptual and operational distinctions between them, and 2) the difficulty of experimentally inducing stress in a reliable fashion in the laboratory. As to 1), the existence of the general concept of "stress" and its rather indiscriminate usage in everyday life and in scientific research seems to suggest that the enormous panoply of states loosely referred to by this term have something in common. However, this common factor has yet to be identified. For example, Ruiz et al. [7] compare a laboratory stress situation (a female subject performing a cognitively challenging task) with a real stressor (two pilots discussing technical problems just before an airplane crash). While both situations imply a higher-than-normal cognitive load, the danger component that looms large in the real situation is absent from the lab task. Most likely, then, the crash situation implies an emotional stress, whereas the lab situation should at most produce mild cognitive stress. As one might expect, the authors find important quantitative and qualitative differences between the situations and the speakers involved.

Unfortunately, in this study, as in many other studies in this area, single cases, involving only one or very few speakers, are examined. This procedure seems defensible in cases in which there are unique records of extreme stress or emotion in real-life settings (such as cockpit voice recordings). Acoustic analyses of such special cases may provide important clues as to the nature of affective voice changes and may inform the formulation of hypotheses. However, the analysis of single speakers is less appropriate when hypotheses are to be tested or when parametric estimates (e.g., of degree of F0 change) are intended. The single-case approach precludes the use of inferential statistics and thus the examination of the significance and effect size of observed differences as well as the calculation of confidence intervals. In consequence, there is little justification for using only one (or a very small number) of speakers in laboratory studies.

As part of a large-scale study on the psychological, physiological, and behavioral effects of different kinds of stress on different types of individuals, Scherer and his collaborators [see 5] systematically compared the effects of inductions of cognitive stress (difficult arithmetic tasks) and emotional stress (slides showing major injuries) in 60 subjects. Tolkk and Scherer [5] showed strong, statistically significant differences in the way different individuals reacted to the two types of...
stressors in this study (see following). Thus, rather than glossing over differences between different types of stress-inducing situations, it seems desirable to base future research on better-defined concepts and operationalization. Scherer [8] has suggested defining different types of “emotional stress” in terms of the underlying type of emotion (anger-stress, fear-stress, etc.). The stress quality is seen as due to the emotional reaction being more intense and more prolonged than in the case of normal emotions. Similarly, cognitive stress should be distinguished (e.g., by perceived difficulty, time pressure, coping potential, etc.) from increased processing load during cognitively demanding tasks.

We suggest that we will be unable to find reliable acoustic profiles for stress (should they exist) unless we 1) clearly differentiate between different types of psychological stress on the one hand and task load on the other, and 2) study only those speakers who are indeed in a stressed state (that is, for whom stress induction actually worked).

Directly addressing the first issue, this paper reports research from the EMOVOX project at the University of Geneva (funded by the Swiss National Research Fund). The project, which has grown out of the European VERIVOX research consortium [9], aims at a better understanding of the pattern of changes that different physiological and psychological speaker states produce in the acoustic speech signal. It is expected that increases in our knowledge concerning the fundamental mechanisms whereby emotions and stress modify voice and speech will be of major importance for speech technology, allowing improvement of performance and acceptance of speaker verification, speech recognition, and speech synthesis applications [10,11].

To produce experimental manipulations of affective speaker state in a controlled manner, we developed a computer-aided speech recording tool, in which the speaker is confronted with “computer tasks” meant to replicate natural working situations or situations from everyday life. Spontaneous speech and read speech are recorded while the speaker completes the different tasks. After each task, speakers are prompted to self-rate their emotional state, permitting an evaluation of the relative efficacy of different tasks in inducing the targeted emotional reactions. In this contribution, we focus on one of these tasks, which was intended to allow a distinction between task load and psychological stress as outlined above, in their effect on a number of acoustic parameters generally studied in this domain.

2. METHOD

2.1 Computer-aided tool for emotion induction

The program consists of six tasks, presented in random order to each speaker. In this paper we focus on one of these tasks, logic deductions plus auditory stimuli differentiation, a split attention task, simulating a context in which a person has to work while being disturbed by other stimulation. Participants are asked to perform a logical reasoning test under two different conditions: 1) focusing on the task without any disturbance, and 2) working on the task while attending to an auditory monitoring task at the same time. Concretely, they have to make logical deductions on the basis of given premises displayed on screen, and, in the dual task situation (simulating two telephones with different ringing sounds that one needs to attend to while doing other work) respond to a particular sound while ignoring another sound (both occurring in an unpredictable fashion). We presume that the single logical reasoning task produces simple cognitive load whereas the dual task, in addition to a greater cognitive load and different attentional demands, adds psychological stress due to encountering capacity limits under time pressure. In addition, the phone ringing sounds used had a highly arousing quality. Participants were randomly exposed to two repetitions of each condition, using different logical deduction items for each subtask.

2.2 Speech recording

Pop-up windows at the beginning and end of each task as well as at different times during the task prompt the user to utter standard phrases and series of numbers while performing the task (for example: “This is task number 345629”, with the number changing for each subtask). Thus, while part of the utterance is standardized, another part varies across subtasks. All speech recording is entirely controlled by the computer program, which starts and stops sampling via the computer’s audio card. A high-quality condenser microphone built into a headset is used (keeping the distance from the mouth relatively constant).

2.3 Verbal feeling report

After the task, participants are asked to self-rate their emotional state by choosing one or more emotion words (among a list, including “stressed”) and an intensity level describing their present state. This self-assessment of the emotional reaction is used as a control to evaluate the relative efficacy of different tasks in inducing the targeted emotional reactions.

2.4 Speakers

Using the computer-induction tool, 100 male speakers were recorded (25 native German, 16 native English, and 59 native French speakers). Each speaker produced about 100 sentences of read speech and several passages of spontaneous speech.

2.5 Acoustic analysis

An extensive set of automatic acoustic analysis routines was developed by one of the authors (GK), covering both established acoustic parameters that have been implicated in the expression of stress and emotion in the voice, as well as a number of more rarely used parameters.

2.6 Statistical analysis

The statistical analysis focuses on four types of speech samples: the standard utterance spoken a) after a pause, b) before the beginning of the task, following the instructions, c) during the single task, and d) during the dual task (averaging the acoustic
parameters across two repetitions in the case of c and d). By comparing the speech samples before and during the task, we can study the effect of cognitive load, due to task engagement. By comparing the speech samples after a pause and during the simple single task condition with those before task execution and during the dual task condition, we can examine the effect of psychological stress. We assume that such stress is induced by worrying about the nature of the difficult task that has been announced in the instructions during the before condition and by capacity limitations, worrying about one's coping potential, and unpleasant stimulation while working on the dual task. The statistical analyses consisted of Repeated measures ANOVAs with the factors Load and Stress (low vs. high, in both cases). Since no major differences between the three language groups were found, this variable will not be discussed.

### 3. RESULTS

In what follows, the theoretical predictions for the effect of activation (see [1]), which can be expected for both increases in stress and load, will be examined in the light of the preliminary results from this study. Figure 1 shows the changes in speech rate (which is faster if the average duration of each syllable is shorter). While there is no effect for stress, there is a highly significant ($p < .001$) increase in speech rate under high load.

Activation is also expected to lead to sharper attack and decay gradients in the amplitude contour. Figure 2 shows the average gradient of the stylized amplitude envelope in signal segments characterized by energy decay. Both high load and high stress combine to increase the steepness of the decay gradient, as shown by a significant interaction effect ($p = .011$). The gradient of energy attack is also increased by load ($p = .005$) but not by stress (n.s.). In consequence, it cannot be excluded that the speech rate effects reported in the literature, are mostly due to load rather than stress. The same could be the case for increases in the steepness of the amplitude attack and decay gradients.

As shown before, the literature suggests a strong effect of stress on F0. As expected, we found (see Figure 3) a significant effect of stress on F0 ($p = .01$), with load having a smaller, marginally significant effect ($p = .09$). There were no effects on F0 variability as measured by F0 standard deviation. While the effect of stress is reliable and significant, its magnitude seems small, less than 2 Hz on the average. However, note that the mean values reported hide the fact that some individuals show an important increase in F0 while others show no change or even a decrease.

Indeed, most researchers using stress and/or emotion induction concur that stress induction techniques will only work for a subset of the participants in any experiment. This is due to the fact that stress and emotion are elicited by subjective appraisal of the eliciting event or situation with respect to an individual's current goals or values and the perceived coping potential [12]. Since the very same event (or induction) can be appraised in a very different fashion by different individuals, it is not surprising that the resulting affect states are rarely very homogeneous.

Thus, Scherer [3, pp. 180-181] showed that stress induced by the requirement to deceive an aggressive interviewer, increased mean F0 only in those subjects scoring high on the personality trait Achievement via Independence. It seems possible that these subjects had a higher standard for achieving successful deception as required by the instruction and thus were more stressed. Tolkmitt and Scherer [5] found that, whereas F0 floor (lowest 5% of F0 values) was higher under both cognitive and emotional stress for high-anxiety subjects and anxiety-denying subjects, low-anxiety subjects showed the opposite change in F0 floor. Most importantly, the authors showed that de-peripheralization of vowels (as measured by the distance between observed and normative formant frequencies) occurred only for female subjects characterized by a strong tendency to deny anxiety. They showed a strong tendency to de-peripheralize vowels under emotional stress but to peripheralize them under cognitive stress. These differences can be explained by different types of self-ideals and coping strategies in subjects with different types of personality.

Strong effects of load and stress are theoretically expected for the spectral domain. As shown in Figure 4, high load leads to a

![Figure 1](image1.png) Differences in speech rate between load and stress levels

![Figure 2](image2.png) Differences in average gradient of energy decay between load and stress levels

![Figure 3](image3.png) Differences in mean F0 between load and stress levels

![Figure 4](image4.png) Differences in mean F0 between load and stress levels
significant decrease (p < .001) in the relative energy below 500 Hz (and a corresponding increase in the proportion of energy in the range from 500 to 1600 Hz; the two parameters correlating with r = -.86, p < .001. The effect for stress, while in the same direction, is not significant. Again, this weak effect could be due to stress induction having worked only for a subgroup of speakers.

![Graph](image)

**Figure 4:** Differences between load and stress levels in the proportion of energy below 500 Hz in voiced segments

We used the self-report data to examine this hypothesis. A paired t-test showed that overall subjective stress levels were higher in the low stress condition than in the high stress condition (p = .016). However, the only correlation of stress change with acoustic change was found for syllable duration (r = -.28, p = .014). In other words, speakers whose speech rate increased following the stress manipulation also tended to report a higher level of stress. However, this may reflect the tendency of subjects to report being stressed when under cognitive load, since, as shown before, there is no effect of experimentally manipulated psychological stress on speech rate. These results confirm the suspicion that self-report is not a good standard for measuring the success of affect or stress induction. It may be necessary to develop alternative ways to identify the individuals who have appropriately responded to a stress induction. One possibility is to look at the pattern of responses. For example, if one only considers speakers who show an F0 change in the predicted direction (N = 55), there is a significant stress effect for the decrease in the proportion of energy below 500 Hz and the corresponding increase in the range from 500-1600 Hz. Since there is no correlation between F0 mean and spectral energy distribution parameters, this probably reflects the effect of focusing on speakers who actually experienced stress.

4. CONCLUSIONS

As shown in the introduction, research on the effects of stress on voice and speech has been beset by the problems of an insufficient specification of the type of stress studied and by a neglect of strong individual differences in the susceptibility to experience stress. As a result, there has been little evidence for the existence of a replicable acoustic profile for stressed speech. The approach suggested in this contribution may contribute to finding a solution to this problem. The results suggest that cognitive load due to task engagement will reliably increase speech rate, energy attack and decay gradients, mean F0, and the proportion of energy in the higher frequency range. In contrast, only F0 and the change in spectral energy distribution seem to respond to the induction of psychological stress. In fact, it seems reasonable to expect that cognitive and attentional demands primarily affect fluency and rate whereas sympathetic activation mostly affects F0 parameters [13].

By defining the speaker states to be studied more carefully and by using well-defined induction procedures that are comparable between studies, it may be possible to replicate consistent patterns of vocal change across different studies. It is imperative for future studies in this area to adopt the established principles of experimental design and to use a number of speakers that is sufficient to test the significance of results, the size of the effects, and potential sources of individual differences. In addition, it seems essential to control for the success of the induction procedure since it is difficult to find vocal patterns of stress in non-stressed speakers. In this sense, the average differences found in this study, even though often statistically significant, and thus reliable, cannot be used as parameter estimates for the effect of load or stress on speech.

5. REFERENCES