



Fundamental Frequency as an Alternative Method for Assessing Mental Fatigue

Enrique Muñoz-de-Escalona^(✉), José Juan Cañas, and Jessica F. Morales-Guaman

Mind, Brain and Behaviour Research Centre, University of Granada, Granada, Spain
{enriquemef, delagado}@ugr.es, jessicamorales@correo.ugr.es

Abstract. The development of methods for measuring mental workload and fatigue is still a central issue in the research agenda of Human Factors and Ergonomics researchers. There exist several different methodologies which offers indexes which are sensitive to mental workload and fatigue changes. One of these methodologies, the physiological measures offers many advantages, but comes with the main disadvantages of being very intrusive and require expensive and specialized equipment. To overcome these disadvantages, we propose the use of voice parameters, particularly the fundamental frequency, as an alternative valid method for assessing mental fatigue in complex and dynamic situations. In this study we tested our proposal. We manipulated the time during which participants performed the MATB-II task battery with high and constant task demands in order to induce mental fatigue over time; while subjective reports of fatigue, performance, pupil size and voice speech were collected as our dependent variables. Results showed that all of these measures, including the fundamental frequency obtained from participants' voice, were sensitive to the emergence of mental fatigue. High correlations among all measurements also confirmed the appropriate use of fundamental frequency as an alternative sensitive measure for assessing mental fatigue.

Keywords: Mental fatigue · Mental workload · Fatigue measures · Fundamental frequency

1 Introduction

Research over the past 50 years has clearly shown that performance on complex tasks depends, among many other factors, on the mental resources that are demanded by tasks, but also on the operators' available mental resources to cope with those demands. The relation between demanded and available resources is what we call mental workload [1–4]. This relationship can be expressed as a fraction where the numerator is the amount of demanded resources and the denominator is the amount of available resources. When this fraction is greater than one, we say that we are in an overload condition, in which demanded resources are higher than available resources. If, on the contrary, the result of the fraction is less than one we have an underload condition, in which available resources are higher than task demands. The conditions of extreme overload and extreme underload negatively affect the performance of the task and the operator's welfare [1–4].

For this reason, human work is designed to avoid these situations. It should be noted that performance would not be impaired with intermediate levels of mental workload [5]. An overload condition is mainly caused by an excessive demand on resources when, for example, the task is very complex. However, overloading can also occur when there is a decrease in available resources as a result of the emergence of mental fatigue (during long periods performing the tasks) or when the tasks do not change the objectives for a long time [6, 7]. For this reason, estimating the amount of available resources at any given time has been and continues to be a very important research topic in the area of Human Factors and Ergonomics. Therefore, in the agenda of researchers who study the variables that affect performance in a complex task, there is one main objective which is to develop methods that would allow us to measure the mental resources that a person has at a given time. However, all the methods that have been proposed have advantages but also have disadvantages and more research is still needed to find methods to measure available resources that would avoid them.

In this paper we will describe an experimental study carried out to explore the possibility of estimating the available resources from an acoustic parameter of the voice, its fundamental frequency (F0). Previous research has shown that the fundamental frequency of operators' voice increases as mental workload gets higher [8]. We hypothesize that this increase in fundamental frequency that occurs with time on task may be due to a decrease in available resources because of the gradual emergence of mental fatigue. In the experiment we manipulated the time spent performing the task while keeping high constant task demands. We measured the participants' performance, their subjective estimation of fatigue, their pupil diameter and the fundamental frequency of their voice. We expected to find that as participants spent more time performing the task, their execution would worsen, their subjective perceptions of fatigue would increase, their pupil would shrink reflecting less physiological activation, and especially the fundamental frequency would increase. In Sect. 2 we will review the different methodologies for measuring mental fatigue to focus, afterwards, in the acoustic parameters, particularly in the fundamental frequency. Section 3 describes the design and methodology used in our study. In Sect. 4 we describe the obtained results and in Sect. 5 the discussion about our findings and future possible research in this line. Finally, in Sect. 6 we conclude the study and provide implications about using fundamental frequency as a mental fatigue index.

2 Related Work

2.1 Methods for Measuring Available Resources

Over the past few decades, researchers have designed several methods to measure the mental resources a person has while performing a task. All these methods can be classified in various ways according to various criteria, but a first criterion that is often used is the temporary moment at which the amount of resources is measured. According to this criterion we can distinguish between "offline methods" and "online methods".

Offline methods basically consist of estimating the amount of mental resources once the task has been completed. These methods have the advantage that they normally do not interfere with the performance of the task because they are collected at the end of the task but have the disadvantage that they do not allow an estimation during the

performance of the task. Sometimes, what we are interested in, is how many mental resources a person has left after the task is completed. For example, we would be in that situation when we want to know how tired a person is at the end of his or her working day and not how the mental resources have been managed during the performance of his tasks at work. In these situations, offline methods would be appropriate for our purposes. However, if our objectives were to estimate mental resources during task performance, we would need to use the so-called online methods.

Regardless of whether the evaluation is offline or online, the methods can be classified into three big categories. The first category of methods is the so-called “concurrent task performance-based methods” which consist of asking the person to perform a task that is called the secondary task while he or she performs his or her primary task, which is the one he or she has to perform at work. Basically, these methods measure the performance of the secondary task and apply the following logic: since both the primary task and the secondary task have to share the same resources, the performance of the secondary task is a reflection of how many resources are “left over” after performing the primary task [4]. If we observe that the secondary task is performed very well, we can interpret that the primary task requires few resources. Conversely, if the secondary task is poorly performed, it will mean that the primary task is consuming most of the resources. These performance-based methods are good online methods for diagnosing the causes of overload or underload. However, these methods have two fundamental drawbacks. First, in real situations (not in the laboratory or in a simulator) sometimes we cannot afford to have a secondary task drawing resources that may be needed to perform well the primary task (which is the main task the person has to perform). Secondly, it is often difficult to ask the person whose mental resources we are estimating to understand that he or she has to “prioritize” the primary task over the secondary task.

The second category of methods is made up of subjective methods that consist of asking the operators themselves to give a subjective estimation of their current available mental resources, normally using a numerical interval scale where the available mental resources can be expressed from less to more during the performance of the task. The main disadvantage of these methods is that they can interfere with the completion of the task when performed online. Although most of these methods only require operators to verbally give their estimation of the amount of mental resources, doing so may distract their attention from the task and impair their performance. Therefore, they are usually used offline at the end of the task or working day.

Finally, the third category of methods, which are known as psychophysiological methods, are based on our knowledge about the relationship between mental resources and certain psychophysiological parameters. There are currently several psychophysiological methods, all of these with advantages and disadvantages. Most of these psychophysiological recording methods that are currently being proposed (records of ocular parameters such as blink rate, pupil diameter, records of the electrical activity of the cerebral cortex, etc.) require high technical expertise and the use of some particular recording equipment, which is expensive and needs the operator to be equipped with it so that the physiological parameters can be properly recorded. For example, records of ocular parameters require the operator to be equipped with special eye-trackers, and the records of cortical electrical activity have to be done by equipping the operator with a

set of electrodes on the head that can be annoying and interfere with the task. Therefore, there is currently a great need to explore other psychophysiological methods for the estimation of mental resources, especially online, which meet two main requirements: they should not require expensive recording equipment and they do not need to ask the operator to be equipped with it during the performance of his task. The methods that we should explore must record psychophysiological parameters in a natural way from some aspect of the operators' open behavior while she or he is performing the task. One of these methods may be the estimation of mental resource states from acoustic parameters of the voice. In many situations where it is necessary to estimate the amount of mental resources that a person has, verbal communications are (or can be) recorded automatically, often for security reasons. One of these situations is the one we have at Air Traffic Management (ATM) [9].

2.2 Effects of Mental Workload on the Acoustic Parameters of the Voice

In the communication between two people it is very important that both are able to perceive the emotional state of the other. For this perception of the emotional state of the person we are communicating with, we can use several signals, the most important of which is the gesture of the face [10]. We humans manifest our internal psychological states very well through gestures of the face. However, another characteristic of human communication that also shows the internal psychological state is the voice. In the field of ATM, the relationship between mental states and communications has been known for a long time. For example, Bailey, Willems and Peterson [11] found that the number of communications between route controllers and pilots increased in high mental load conditions. However, the current interest in this area is primarily to explore the relationship between the acoustic characteristics of verbal communications and the mental states. In this line, many of the first studies that were carried out on the acoustic parameters of the voice were aimed at exploring the relationship between these parameters and stress. In an internal EUROCONTROL report, Hagmüller, Kubin and Rank [12] reviewed the empirical evidence on the influence of stress caused by mental workload on the acoustic parameters of the voice. The authors argue that since voice is a common communication tool of Air Traffic Controllers (ATCOs), the analysis of its acoustic characteristics can be very useful to identify its psychological states during the performance of the control task. The authors considered mental workload as a “psychological” stress-causing factor that is different from other physical (e.g., vibrations produced by machines), physiological (e.g., illness, lack of sleep), or perceptual (e.g., noise) factors. One of the most important conclusions of the review conducted by these researchers was that the Fundamental Frequency is sensitive to changes in stress caused by mental workload.

The fundamental frequency is the lowest frequency of a periodic waveform. In the case of the human voice, when a person produces a sound, the vocal cords vibrate at a certain speed by rapidly opening and closing with small bursts of air. In this way the sound produced is composed of a spectrum of frequencies that we can break down to obtain the lowest frequency of the spectrum. That lower frequency of the produced sound is what we call the fundamental frequency. We should not confuse the fundamental frequency with the tone of a sound even though they are related. Tone is a unit of measurement for “perception”. We could say that fundamental frequency is a “physical” measure of

sound that can be obtained by placing a wide-range microphone directly in the throat, above the vocal cords, but below the resonant structures of the vocal tract. While pitch is a “psychological” measure that reflects how frequencies are perceived in the human nervous system and is measured by the Mel Scale. The name Mel comes from the word melody to indicate that it is based on human perception of tone. The relationship between fundamental frequency and tone would be similar to the relationship between physical sound intensity and the decibel scale which is a perceptual psychological measure of intensity. In any case, since the relationship between frequency and tone is a complex relationship that leads us to the field of Psychophysics, researchers have preferred to measure fundamental frequency directly in their research on the relationship between mental states and the physical parameters of the voice.

Effects on fundamental frequency have been consistently found in studies where participants were asked to identify whether or not a person was stressed by hearing what this person said. In these studies, researchers have found that the characteristic that listeners use to make a meaningful classification is Fundamental Frequency [13, 14]. These studies soon distinguished between the different causes of stress. Although, in a general way we can say that stress is a psychological and physiological response of the organism to a danger, we can distinguish between the different types of dangers, called stressors, to which the organism responds. In this line, Scherer, Grandjean, Johnstone, Klasmeyer, and Bänziger pointed out that one thing is the stress produced by mental overload and another is the stress produced by other situations of danger [15]. Consider two examples: the stress of an ATCo may be caused by a high workload, but it is also possible that the stress we observe in an ATCo is due to a situation of danger from an imminent conflict. In the latter case the observed stress has a causal component with an emotional component due to the perceived risk of collision. In the experiment carried out by these authors, the participants had to perform a logical reasoning task in one of two conditions (1) focusing on this task without any distraction; (2) performing the task while simultaneously attending to an auditory monitoring task. In the logical reasoning task, they had to make deductions based on premises that appeared on a screen. In the auditory monitoring task, they had to respond to one sound while ignoring another sound, both appearing randomly. The authors hypothesized that in the condition of the logical reasoning task there would only be a mental workload effect, while in the dual task, in addition to increasing this mental workload by having to perform two tasks simultaneously, there would be an effect of emotional stress due to time pressure and the same stressful noise defect. In this study, the participants were asked to say a few sentences appearing on the screen at intervals while they were performing the tasks. The duration of the intervals was random. These sentences were of the type “This is task number 345629”. In the sentence, only the number changed from one time to the next. In this way, there was one part that was always fixed and another part that was changing, the number.

The results of the experiment showed a clear statistical independence of the effects of stress and workload on F0. The two effects did not interact, allowing the researchers to conclude that they can be considered independent effects. However, although the effect of stress was statistically significant, the effect of workload was only marginally so. The authors argued that this lack of significance was due to the large individual differences

they found in the sample of participants. Based on the results of questionnaires that the participants had to fill out, the authors suggested that the participants with higher levels of anxiety caused by the desire to perform the task well were those who showed the most significant results on the effect of F0. Huttunen, Keränen, Väyrynen, Pääkkönen, and Leino have conducted a study with military pilots in a flight simulator piloting task [16]. The three mental load conditions in which the pilots performed their task were (1) mental workload due to the complexity of Situation Awareness; (2) mental workload due to the complexity and critical value of information; and (3) mental workload due to the difficulty of the decision making processes. Taking as a baseline the recordings collected in a period of time before the simulation, the authors calculated the differences in F0 as a function of mental workload in these three conditions. The results showed that F0 increased as a function of mental workload in all three conditions.

In a word recall task under controlled laboratory conditions, Boyer, Paubel, Ruiz, El Yagoubi and Daurat found that the F0 increased with the number of words participants had to remember [17]. The authors interpreted that the number of words can be considered a factor that increases cognitive demand and, therefore, its effect on the F0 can be interpreted as evidence of the relationship between mental workload and this acoustic parameter of the voice. Therefore, we can say that there is experimental evidence of the effect of mental workload on an acoustic parameter of the voice, the fundamental frequency of people who are performing complex tasks. The fundamental frequency increases when the mental workload increases. However, it would be interesting to know if this effect of mental workload can be due to an increase on mental demands or a decrease in available resources. We believe that a decrease in the available resources due, for example, to the time spent performing the task may be also a factor affecting the observed increase in the Fundamental Frequency due to the emergence of mental fatigue.

2.3 Effects of Fatigue on Fundamental Frequency

Fatigue research has found results that show how acoustic parameters of the voice, such as fundamental frequency, can be sensitive to fatigue. For example, research in recent years has shown that fatigue can affect different phases of speech production. According to the review by Krajewski, Batliner, and Golz these phases would be as follows [18]:

1. Cognitive speech planning has been found to result in a slowdown in cognitive processing, impaired speech planning, impaired neuromuscular motor coordination processes, impaired fine motor control and slow movement of the articulator, and sluggish articulation and slower speech.
2. There are effects on breathing that manifest themselves in decreased muscle tension, flat and slow breathing, reduced subglottal pressure, and lower fundamental frequency, intensity, articulatory accuracy and joint rate.
3. The effects on phonation that have been found are as follows: decreased muscle tension, increased vocal cord elasticity and decreased vocal cord tension, decreased body temperature, change in vocal cord viscoelasticity, change in spectral energy distribution, breathable and lax voice, non-lifting larynx, decreased resonant (formant) frequency positions and increased formant bandwidth.

4. The effects on the articulation/resonance are decreased muscle tension, unrestricted pharynx and softening of the vocal tract walls, loss of speech signal energy, wider formant bandwidth; postural changes, lowering of the torso and head, change in the shape of the vocal tract, change in the position of the formant, increased salivation, loss of energy, decreased body temperature, reduced heat conduction, change in friction between the walls of the vocal tract and the air, changes in laminar flows, jet streams and turbulence, change in spectral energy distribution, increased formant bandwidth, increased formant frequencies especially in lower formants.
5. As for the radiation the effects are a decrease of the orofacial movement, facial expression and lip extension (visualization of open and relaxed mouth), lengthening of the vocal tract, lower positions of the first and second formant, reduction of the articulation effort, lower degree of openness, relaxed articulation, decrease of the first formant, oropharyngeal relaxation, decrease of the veil, coupling of the nasal cavity, increase of the nasality, band width of the extended Formant 1, lower amplitude of Formant 1.

Based on these known effects of mental fatigue on speech production, research has been aimed at identifying which acoustic parameters best reflect the effects of mental fatigue. In a study by Cho, Yin, Park, and Park, the authors found that when participants were divided into two groups according to a subjective fatigue scale, some acoustic parameters such as fundamental frequency shaking, brightness, HNR (the ratio of harmony to noise), SNR (the signal-to-noise ratio), and shaking amplitude were predictors of mental fatigue [19]. These results suggest that it would be possible to use some of these parameters to assess mental fatigue. Whitmore and Fisher conducted a study in which a group of American bomber pilots, divided into groups of four, participated in a pilot task for periods of 36 h in a flight simulator [20]. The piloting periods were interspersed with rest periods of 36 h. Approximately every 3 h the participants had to perform cognitive tasks, subjective fatigue assessments and repeat the sentence:

“Futility Magellan, this is xxx yyy. The time is zz:zz Zulu”.

where:

XX was the participant’s rank

YY was the Participant’s Name

ZZ:ZZ was the time when you say the phrases

The results showed that the fundamental frequency and duration of words were good indices of subjective fatigue and performance in the cognitive tasks. Other researchers have conducted research using methodologies based on voice data recorded during conversations in a natural context. This is what Krajewski, Batliner, and Golz did, who conducted a validation experiment to examine whether automatically trained voice database models can be used to recognize subject drowsiness [21]. Their methodological approach can be summarized in four steps:

- They collected individual speech data as well as associated sleepiness scores for each subject.
- Then, they extracted relevant acoustic characteristics from the speech data.

- With that data, they constructed statistical models of sleepiness scores based on the acoustic characteristics.
- Finally, they tested learned models on new speech data.

With this methodological procedure the researchers did not need to conduct a study in which, in a controlled manner, a group of people were required to verbally express a certain text every certain period of time, as has been done in other empirical studies designed to study this same topic. In contrast, with the procedure used by these authors, the voice data were collected in the natural context in which the people performed their tasks. Pattern recognition algorithms were applied to the data collected with this procedure. These algorithms were trained to recognize basic acoustic characteristics according to the acoustic-preceptual concepts of (1) Prosody (tone, intensity, rhythm, pause pattern and speed of speech); (2) Articulation (speech difficulty, reduction and elision); and (3) Speech quality (breathable, tense, high-pitched, hoarse or modal voice). The algorithms were also trained according to signal processing categories (time domain and frequency domain) and state space characteristics. With this procedure, the classification algorithms, working with an unusually large set of data, were able to determine whether a subject's sleep was beyond a critical threshold. Subsequently, the authors conducted a validation study with new data in which they achieved an accuracy rate of over 86% in unseen data, but from known speakers, with an SVM (artificial intelligence algorithm) classifier.

As this review showed, there is enough empirical evidence about the effect of mental fatigue on some acoustic parameters of the human voice. Empirical evidence on the effect of fatigue on the acoustic parameters of the voice supports the hypothesis that the effect of mental workload on these parameters may be reflecting a decrease in available resources. According to the definition of mental workload as a relationship that can be expressed as a division between demanded and available resources, a decrease in available resources due to factors such as fatigue may be the cause of the effect of the observed mental workload.

The aim of the experiment described below has been to test this hypothesis. In the experiment a group of participants performed a set of tasks in which their performance was measured at the same time as their pupil diameter, the subjective estimation of their fatigue and the fundamental frequency of their voice. The pupil diameter measurement was measured because it has proven to be a good index of the level of activation [22]. Task demands remained constant throughout the task. The variable that was manipulated was the time spent performing the task. We expected to find that the longer time performing the task, the greater the fatigue, causing a decrease in performance, a smaller pupil diameter, a greater subjective feeling of fatigue and, most importantly according to our hypothesis, an increase in the fundamental frequency of participant's voice.

3 Design and Methodology

3.1 Materials and Instruments

MATB-II Software. The participants of the experiment had to perform the Multiple Attribute Task Battery (MATB-II) [23]. This task battery is designed to assess the performance and workload of operators by means of different tasks similar to those performed

by flight crews. The software used to perform the tasks has a user-friendly interface that allows non-pilot participants to use it. MATB-II comes with default event files that can be easily modified to suit the needs or objectives of an experiment. The program records the events presented to the participants, as well as their responses. MATB-II contains the following four tasks: the System Monitoring Task (SYSMON), the Tracking Task (TRACK), the Communications Task (COMM), and the Resource Management Task (RESMAN) (see Fig. 1).

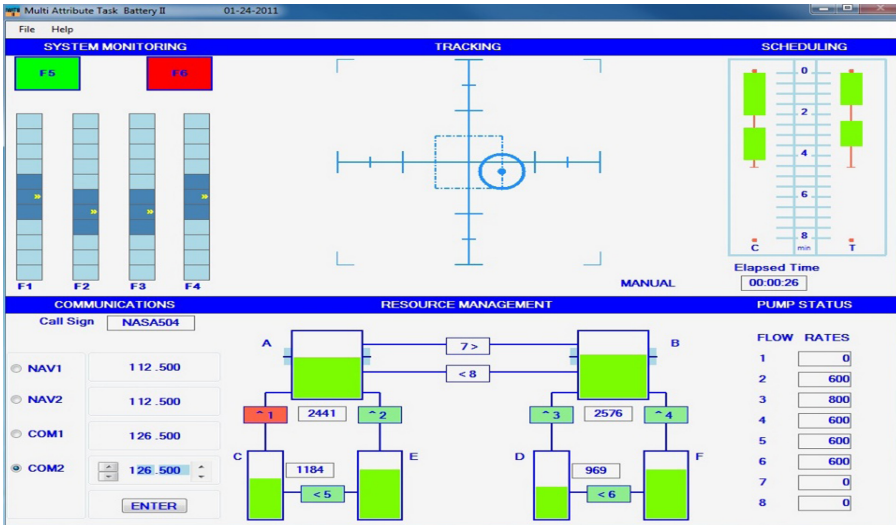


Fig. 1. MATB-II task display. Taken from <https://matb.larc.nasa.gov/>

- The task of SYSMON is divided into two sub-tasks: the lights task and the scales task. For the sub-task of lights, participants are required to respond as quickly as possible to a green light turning off and a red light turning on, and to turn the lights back on and off, respectively. For the scale sub-task, participants are asked to detect when the lights on four moving scales deviate from their normal position and respond accordingly by clicking on the deviated scale.
- In the TRACK task, there are two modes. Participants can work only in a manual mode. During this manual mode, participants have to keep a circular target in the centre of an inner box using a joystick with their left hand (the dominant hand was necessary for the use of the mouse). During the automatic mode of the task, the circular target will remain in the inner box by itself.
- In the COM task, an audio message with a specific callsign is displayed and the participant is asked to respond by selecting the appropriate radio channel and by setting the correct frequency, but only if the callsign matches his or her own (callsign: “NASA504”). The participant is not required to respond to messages in other callsigns.
- In the RESMAN task, participants have to maintain the fuel level in tanks A and B, within ± 500 units of the initial condition of 2500 units each. To maintain this

objective, participants must transfer the fuel from the supply tanks to A and B or transfer the fuel between the two tanks.

Praat Software. “Praat” is a scientific license free tool for analyzing spectrograms of audio records. It was developed at the University of Amsterdam by Paul Boersma and David Weenink in 1992 and it is constantly being updated with improvements implemented by authors, some of them suggested by users [24]. Once the audio file is loaded you can obtain multiple audio parameters such as fundamental frequency, intensity, volume, formants, etc. In this study we used Praat software for obtaining the fundamental frequency intervals average.

Tobii T120 Eyetracker. Pupil diameter measurements were obtained using an infrared eye tracking system with a sampling frequency of 120 Hz, the model Tobii T120 marketed by Tobii Video System. This equipment is completely non-intrusive, does not have a visible eye movement tracking system and provides high accuracy and an excellent compensating head movement mechanism, ensuring high quality data collection.

Instantaneous Self-assessment Scale. We employed an easy and intuitive instantaneous subjective fatigue scale called Instantaneous Self-assessment, which provides momentary subjective ratings of perceived mental fatigue during task performance. Participants evaluated the mental fatigue they experienced at any given time on a scale ranging from 1 (no mental fatigue) to 9 (maximum mental fatigue). Participants were taught to use the scale just before starting the experimental stage.

3.2 Participants

17 students from the University of Granada participated in the study. Participants’ ages ranged from 18 to 32, with an average of 23.6 and a standard deviation of 2.25. A total of 13 women and 4 men participated. It should be noted that there is a greater number of female participants due to the fact that psychology students at the University of Granada are mostly women. Recruitment was achieved through the dispersion of posters and flyers around the university, as well as an advertisement for the study on the university’s online platform for experiments (<http://experimentos.psiexpugr.es/>). The requirements for participation included (1) not being familiar with the MATB-II program, (2) Spanish as a native language, and (3) visual acuity or correction of visual impairment with contact lenses, as glasses impair the utilized eye-tracking device from collecting data. Participants’ participation was rewarded with extra credit.

3.3 Procedure

The participants went through an experimental session consisting of two phases:

1. **Training stage:** training took place for no longer than 30 min. The objective of this stage was for participants to familiarize themselves with the program so that they could carry out the tasks securely during the data collection stage. The procedure

was conducted as follows: upon entering the lab and after filling out the informed consent form, the participant was instructed to read the MATB-II instruction manual and inform the researcher once they had finished. The researcher then sat down with the participant to allow for questions and resolve any doubts on how to use the program. Afterward, on a computer monitor, participants were presented with each MATB-II task separately and were first given a demonstration as to how to execute the task and after what they were given time (3 min or more if needed) to perform the task themselves. The participants were always free to consult the manual and ask the researcher questions during the training stage in case of doubts or uncertainties. Once the participants had completed all four tasks and resolved all doubts, they were ready for the data collection stage, which followed immediately afterwards. During the training stage, participants could work in one room equipped for training with the MATB-II software, and no special attention to room conditions was needed.

2. **Data collection stage:** the data collection stage lasted a period of 60 min that was divided into thirty intervals of 2 min. The first interval was left as a training interval in order to focus the attention of participants into the task. During this first 2 min interval only the tracking task was activated. Then, during the second interval and until the end of the experiment, the three tasks of the MATB-II software were activated (TRACK, SYSMON and RESMAN), thus mental workload level was high and constant throughout the experimental session in order to facilitate the emergence of mental fatigue. The participants were instructed to verbalize the ISA (Instantaneous Self-Assessment) scale every 2 min when a scheduled alarm sounded. They were also instructed to verbalize every action they were performing in the SYSMON (e.g. “I press F3 button”) and the RESMON (e.g. “I activate fuel pump n°1”) tasks in order to collect our F0 variable through audio recordings. Prior to the start of the task-battery, the eye-tracker system was calibrated, and the participants were told to keep head and body movements to a minimum. During the data collection stage, standardizing room conditions was essential. Thus, the testing rooms were temperature controlled to 21 °C, and lighting conditions (the main extraneous variable in pupil diameter measurement) were kept constant with artificial lighting; there was no natural light in the rooms. Moreover, participants always sat in the same place, a comfortable chair spaced 60 cm from the eye-tracker system.

This study was carried out in accordance with the recommendations of the local ethical guidelines of the committee of the University of Granada institution: “Comité de Ética de Investigación Humana”. The protocol was approved by the “Comité de Ética de Investigación Humana” under the code: 779/CEIH/2019. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

3.4 Variables

Independent Variable

The only independent variable was the time spent performing the task. The period of one hour performing the task was divided into 2-min intervals resulting in 30 intervals. The first period of 2 min was taken as a baseline to calculate the pupil diameter so it

will not be analyzed and we are left with only 29 intervals that constitute the 29 levels of our independent variable “intervals” during the time performing the task.

Dependent Variables

Performance. MATB-II provides us with many indicators of participants’ performance: e.g. root mean square deviation (RMSD) for the TRACK task, number of correct and incorrect responses for the SYSMON and COMM tasks, and the arithmetic mean of tanks “A-2500” and “B-2500” in absolute values for the RESMAN task. However, for the purposes of this experiment we will only consider the RMSD performance indicator. The RMSD performance indicator reflects the distance of the circle to the target point, in such a way that, performance impairment is reflected by a higher score on this variable.

Pupil Size. Fatigue can be estimated by several physiological indexes such as EEG, HVR, and several ocular metrics. We decided to use pupil diameter as our physiological fatigue indicator, as it effectively reflects mental fatigue [19] and minimize intrusiveness. While our eye-tracking system allows continuous sampling rate recording at 120 Hz, we set a total of 30 intervals lasting 2 min each. Since expressing pupil size in absolute values has the disadvantage of being affected by slow random fluctuations in pupil size (source of noise), we followed a procedure for standardizing the values of pupil size for each participant.. To do this, for every participant, we took his/her pupil size value during the first interval of 2 min, and then subtracted it from the obtained value in each of the rest of 29 intervals, thereby giving a differential standardized value allowing us to reduce noise in our data. Analyses were carried out for the average of both the left and right pupils. A negative value meant that the pupil was contracting while a positive value meant that it was dilating.

Subjective Fatigue. We used an online subjective fatigue scale created for this purpose, the “Instantaneous Self-Assessment Scale”. Ratings were obtained at 2 min intervals throughout the 60 min of the experimental stage, obtaining a total of 30 subjective mental fatigue ratings. The rating from the first 2 min interval was discarded for the analysis.

Fundamental Frequency (F0). We recorded the verbalizations of participants through the microphone integrated in our eye-tracker system during the data collection stage. The 60 min obtained audio file was divided into 30 intervals which were analyzed with the “Praat” software in order to obtain the average F0 of each interval.

Synchronization of Measures. Performance, pupil size, subjective, and fundamental frequency measures were obtained continuously throughout the experimental session. Synchronization between measures was simple, as the eyetracker (pupil and F0) and MATB-II performance log files began to record data simultaneously at the start of the experimental session. The scheduled alarm (every 2 min) was also synchronized by the experimenter, as it was simultaneously activated with the MATB-II software. This would also allow the ISA scale to be synchronized with the performance, F0 and pupil size measures.

3.5 Experimental Design

The experimental design was One-Way with Intervals as the only within-subject variable with 29 experimental levels.

4 Results

We performed four one-way within-subject ANOVA to analyze collected results, one for each different measure. Firstly, the ANOVA for our performance variable revealed a significant effect of intervals $F(28,448) = 2.27$, $MSe = 52.71$, $p < .01$. As we can see in the graph, performance worsened considerably during the first intervals (2 to 4), and then, it stabilized. However, we can appreciate that it continued worsening through intervals until the end of the experiment. Trend analysis confirmed this statements, as linear trend found to be significant, $F(1,16) = 5.02$, $MSe = 330.21$, $p < .05$ (see Fig. 2).

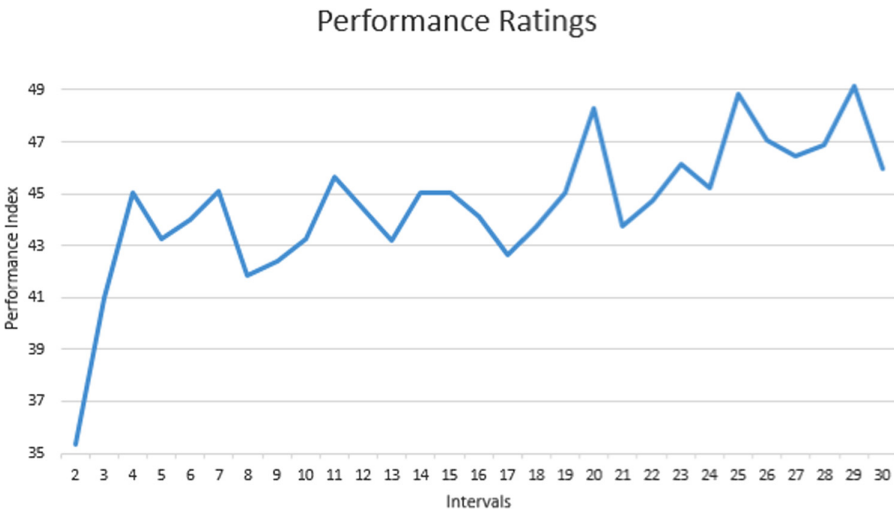


Fig. 2. Participants' performance during task development.

Regarding subjective mental fatigue scores, we found a linear increase through intervals. The ANOVA for this variable revealed a significant main effect of intervals $F(28,448) = 66.14$, $MSe = .79$, $p < .01$. Trend analysis supports the linear increase of mental fatigue through intervals, $F(1,16) = 189.87$, $MSe = 7.34$, $p < .01$ (see Fig. 3).

With respect to our first psychophysiological variable, pupil size, the ANOVA showed a significant effect of intervals $F(28,448) = 20.63$, $MSe = .01$, $p < .01$. The graph revealed a sudden dilation in pupil size from interval 2 to 3 and then, we can appreciate a linear decrease through intervals which tends to stabilize during the last 6 intervals. Trend analysis also revealed a linear trend for this variable, $F(1,16) = 33.0$, $MSe = .1$, $p < .01$ (see Fig. 4).

The ANOVA analysis for our second psychophysiological variable, F_0 , revealed again a significant main effect of intervals $F(28,448) = 2.04$, $MSe = 91.33$, $p < .01$.

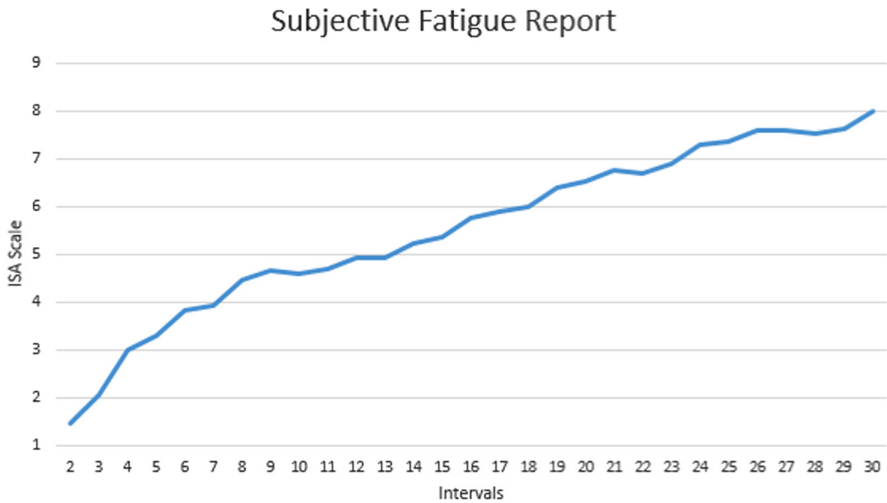


Fig. 3. Participants' subjective mental fatigue ratings during task development.

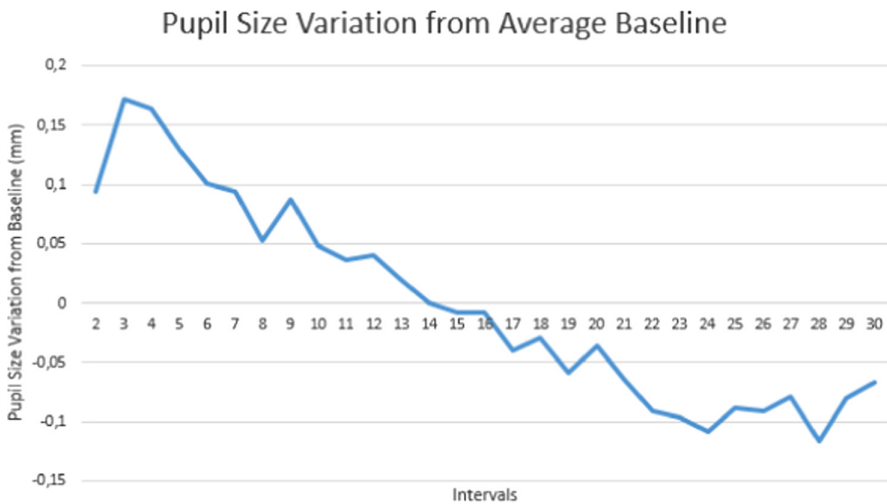


Fig. 4. Participants' pupil size variation ratings from the average baseline during task development.

We can see in the graph that F0 increased linearly through intervals, as trend analysis revealed, $F(1,16) = 6.35$, $MSe = 540.69$, $p < .05$ (see Fig. 5).

Finally, the correlation chart in Table 1 revealed significant very high correlations among measurements: we found a positive correlation between subjective fatigue and performance $.73$, $p < .01$; a negative correlation between subjective fatigue and pupil $-.94$, $p < .01$; a positive correlation between subjective fatigue and F0 $.83$, $p < .01$; a negative correlation between performance and pupil size $-.65$, $p < .01$; a positive

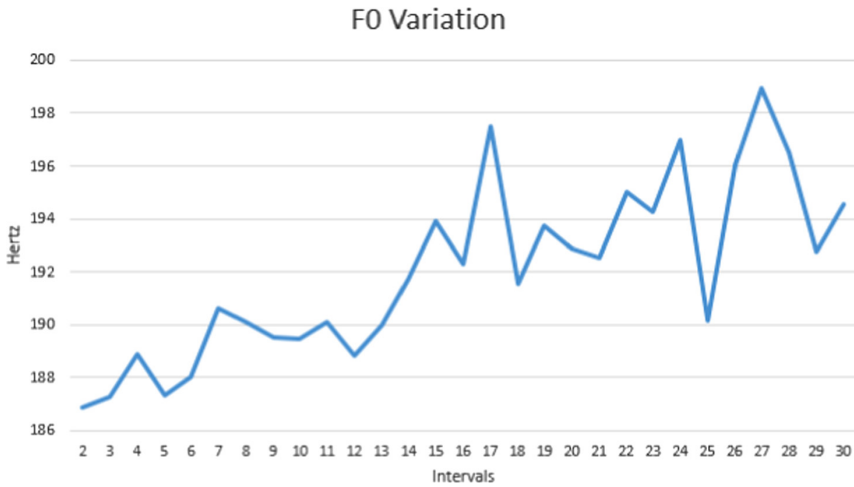


Fig. 5. Participants' F0 during task development.

correlation between performance and F0 $.56$, $p < .01$ and a negative correlation between pupil and F0 $-.86$, $p < .01$.

Table 1. Correlation chart between measures.

		Performance	Sub. fatigue	Pupil	F0
Performance	Spearman correlation	1	$.725^{**}$	$-.652^{**}$	$.564^{**}$
	Sig. (bilateral)		,000	,000	,001
	N	29	29	29	29
Sub. Fatigue	Spearman correlation	$.725^{**}$	1	$-.938^{**}$	$.834^{**}$
	Sig. (bilateral)	,000		,000	,000
	N	29	29	29	29
Pupil	Spearman correlation	$-.652^{**}$	$-.938^{**}$	1	$-.856^{**}$
	Sig. (bilateral)	,000	,000		,000
	N	29	29	29	29
F0	Spearman correlation	$.564^{**}$	$.834^{**}$	$-.856^{**}$	1
	Sig. (bilateral)	,001	,000	,000	
	N	29	29	29	29

*. The correlation is significant at the level .05 (bilateral).

**.. The correlation is significant at the level .01 (bilateral).

5 Discussion

The results show a clear effect of time performing the tasks on every dependent variables: subjective fatigue, task performance, pupil diameter and the fundamental frequency of participants' voice. Particularly, the more time participants spent performing the tasks the worse their performance got, the greater their reports of subjective mental fatigue were, the more contracted their pupil got (indicating that their activation level was decreasing) and the higher the fundamental frequency in their voice got as well. Therefore, considering the above mentioned results we can say that our hypothesis was fulfilled, and the results significantly showed that the fundamental frequency can also be considered an appropriate alternative index to other measures of mental fatigue such as EEG [25, 26], HRV [27, 28], blink rate [29]. The emergence of mental fatigue would be caused by a decrease in available resources (as our pupil size variable reflected) resulting from resources' depletion caused by time on task. These results were in line with literature research and proved that the fundamental frequency can be used as an appropriate index of mental fatigue. It is especially important to note that all measures of mental fatigue correlated significantly in the appropriate direction indicating that participants showed the effects of mental fatigue throughout the task on the one hand, and that every mental fatigue measure in this study was found to be sensitive to mental fatigue as well. However, we must consider that these results should be viewed under certain methodological limitations. First, our sample was small and only composed of young university students, so we believe that these findings should also be tested under working situations in real contexts in order to verify that the fundamental frequency can be used out of the laboratory with satisfactory results, also we must take into account that the number of levels of our independent variable is higher than the number of participants and it might affect to the statistical power of the ANOVA, so our findings must be taken cautiously. Finally, it would also have been interesting to consider different time periods, some of them longer and shorter than 60 min and also under different task demand levels. Further research should consider this methodological modifications in order to analyse the sensitivity of the fundamental frequency to the emergence of mental fatigue to different degrees.

6 Conclusions

The aim of this research was to provide empirical evidence in favour of the hypothesis on the effect that mental fatigue has on an important parameter of the human voice such as the fundamental frequency. The results obtained showed that indeed, the fundamental frequency increases with fatigue. These results allow us to propose the fundamental frequency as an appropriate psychophysiological index of fatigue that may be a viable alternative to other psychophysiological indices that have been proposed. The advantage of fundamental frequency over the other psychophysiological indices is that it does not require expensive and intrusive equipment to record it. The human voice can be recorded in a natural and easy way in many tasks where operators have to communicate to perform their tasks. A direct implication would be that mental fatigue could be predicted at certain works (such as ATM) directly from human voice records, in an automated, economic and

reliable way. Also, historical voice records could be used to analyse the mental fatigue that operators were experiencing at a given time or period to, for example, analyse work-related accidents or productivity at work.

Acknowledgement. This research has been supported by contract between CRIDA (<https://crida.es/webcrida/>) and the University of Granada to the first two authors and by the master thesis research grant from CRIDA to the third author.

References

1. Longo, L., Leva, M.C. (eds.): H-WORKLOAD 2017. CCIS, vol. 726. Springer, Cham (2017). <https://doi.org/10.1007/978-3-319-61061-0>
2. Longo, L., Leva, M.C. (eds.): H-WORKLOAD 2018. CCIS, vol. 1012. Springer, Cham (2019). <https://doi.org/10.1007/978-3-030-14273-5>
3. Longo, L., Leva, M.C. (eds.): H-WORKLOAD 2019. CCIS, vol. 1107. Springer, Cham (2019). <https://doi.org/10.1007/978-3-030-32423-0>
4. Wickens, C.D.: Mental workload: assessment, prediction and consequences. In: Longo, L., Leva, M. Chiara (eds.) H-WORKLOAD 2017. CCIS, vol. 726, pp. 18–29. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61061-0_2
5. Hancock, P.A.: Whither workload? Mapping a path for its future development. In: Longo, L., Leva, M.C. (eds.) H-WORKLOAD 2017. CCIS, vol. 726, pp. 3–17. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61061-0_1
6. Desmond, P.A., Neubauer, M.C., Matthews, G., Hancock, P.A. (eds.): The Handbook of Operator Fatigue. Ashgate Publishing, Ltd, Farnham (2012)
7. Hockey, B., Hockey, R.: The Psychology of Fatigue: Work Effort and Control. Cambridge University Press, New York (2013)
8. Hagsmüller, M., Kubin, G., Rank, E.: Evaluation of the human voice for indications of workload induced stress in the aviation environment. Eurocontrol (2005)
9. Luig, J., Sontacchi, A.: Workload monitoring through speech analysis: towards a system for air traffic control. In: 27th Congress of The International Council of Aeronautical Sciences, pp. 1–10 (September 2010)
10. X Belin, P., Campanella, S., Ethofer, T. (eds.). Integrating face and voice in person perception. Springer, New York (2012)
11. Bailey, L.L., Willems, B.F., Peterson, L.M.: The effects of workload and decision support automation on enroute r-side and d-side communication exchanges. Federal Aviation Administration Oklahoma City Ok Civil Aeromedical Inst (2001)
12. Hagsmüller, M., Kubin, G., Rank, E.: Evaluation of the human voice for indications of workload induced stress in the aviation environment. Eurocontrol (2005)
13. Protopapas, A., Lieberman, P.: Fundamental frequency of phonation and perceived emotional stress. *J. Acoust. Soc. Am.* **101**(4), 2267–2277 (1997)
14. Zhou, G., Hansen, J., Kaiser, J.: Nonlinear feature based classification of speech under stress. *IEEE Trans. Speech Audio Process.* **9**(3), 201–216 (2001)
15. Scherer, K.R., Grandjean, D., Johnstone, T., Klasmeyer, G., Banziger, T.: Acoustic correlates of task load and stress. In: Proceedings of the International Conference on Spoken Language Processing, pp 2017–2020, Denver, Colorado, USA (2002)
16. Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., Leino, T.: Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Appl. Ergon.* **42**(2), 348–357 (2011)

17. Boyer, S., Paubel, P.V., Ruiz, R., El Yagoubi, R., Daurat, A.: Human voice as a measure of mental load level. *J. Speech Lang. Hear. Res.* **61**(11), 2722–2734 (2018)
18. Krajewski, J., Batliner, A., Golz, M.: Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. *Behav. Res. Methods* **41**(3), 795–804 (2009). <https://doi.org/10.3758/BRM.41.3.795>
19. Cho, S., Yin, C.S., Park, Y., Park, Y.: Differences in self-rated, perceived, and acoustic voice qualities between high- and low-fatigue groups. *J. Voice* **25**(5), 544–552 (2011). <https://doi.org/10.1016/j.jvoice.2010.07.006>
20. Whitmore, J., Fisher, S.: Speech during sustained operations. *Speech Commun.* **20**(1–2), 55–70 (1996)
21. Krajewski, J., Batliner, A., Golz, M.: Acoustic sleepiness detection: framework and validation of a speech-adapted pattern recognition approach. *Behav. Res. Methods* **41**(3), 795–804 (2009). <https://doi.org/10.3758/BRM.41.3.795>
22. Muñoz-de-Escalona, E., Cañas, J.J., Noriega, P.: Inconsistencies between mental fatigue measures under compensatory control theories. *Psicológica J.* **41**, 103–106 (2020). <https://doi.org/10.2478/psicolj-2020-0006>
23. Santiago-Espada, Y., Myer, R. R., Latorella, K.A., & Comstock Jr, J.R.: The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user’s guide (2011)
24. Boersma, P., Van Heuven, V.: Speak and unSpeak with PRAAT. *Glott Int.* **5**(9/10), 341–347 (2001)
25. Shen, K.Q., Li, X.P., Ong, C.J., Shao, S.Y., Wilder-Smith, E.P.: EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clin. Neurophysiol.* **119**(7), 1524–1533 (2008)
26. Borghini, G., et al.: Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6442–6445. IEEE (August 2012)
27. Hong, L.Y.J.Y., Zheng, Y.X.L. W.: Study of Mental Fatigue Based on Heart Rate Variability. *Chinese J. Biomed. Eng.* **29**, 1 (2010)
28. Melo, H.M., Nascimento, L.M., Takase, E.: Mental fatigue and heart rate variability (HRV): the time-on-task effect. *Psychol. Neurosci.* **10**(4), 428 (2017)
29. Stern, J.A., Boyer, D., Schroeder, D.: Blink rate: a possible measure of fatigue. *Hum. Factors* **36**(2), 285–297 (1994)