Speech Analysis for Fatigue and Sleepiness Detection of a Pilot

Carla Aparecida de Vasconcelos; Maurílio Nunes Vieira; Göran Kecklund; Hani Camille Yehia

BACKGROUND: Mental fatigue and sleepiness are well recognized determinants of human-error related accidents and incidents in aviation. In Brazil, according to the Center for Investigation and Prevention of Aeronautical Accidents (CENIPA), the rate of accidents in the aerial modal is 1 per 2 d. Human factors are present in 90% of these accidents.

CASE REPORT: This paper describes a retrospective study of the communication between a pilot and an air traffic control tower just before a fatal accident. The objective was the detection of fatigue and sleepiness of a pilot, who complained of these signs and symptoms before the flight, by means of voice and speech analysis. The in-depth accident analysis performed by CENIPA indicated that sleepiness and fatigue most likely contributed to the accident. Speech samples were analyzed for two conditions: 1) nonsleepy data recorded 35 h before the air crash (control condition), which were compared with 2) data from samples collected about 1 h before the accident and also during the disaster (sleepy condition). Audio recording analyses provided objective measures of the temporal organization of speech, such as hesitations, silent pauses, prolongation of final syllables, and syllable articulation rate.

DISCUSSION: The results showed that speech during the day of the accident had significantly low elocution and articulation rates compared to the preceding day, also indicating that the methodology adopted in this study is feasible for detection of fatigue and sleepiness through speech analysis.

KEYWORDS: mental fatigue, sleepiness, voice, speech acoustics.

de Vasconcelos CA, Vieira MN, Kecklund G, Yehia HC. ***Speech analysis for fatigue and sleepiness detection of a pilot*****. Aerosp Med Hum Perform. 2019; 90(4):415–418.**

Fatigue can, from a behavioral perspective, be defined as the state of performance impairment following a period of mental or physical effort.[9] Sleepiness refers to the transition between wakefulness and sleep and is defined as the tendency to fall asleep even when the individual is supposed to be active, for example, performing work. However, both fatigue and sleepiness are characterized by decreased ability to work, to operate machinery safely, attention failures, cognitive slowing, and memory problems.[2,7]

The etiological factors most commonly cited for fatigue are intense physical activity, prolonged or excessive exertion, insufficient sleep, prolonged wakefulness, and physical and emotional illnesses.[7,12] In addition, several work-related and environmental factors in an aircraft can cause or contribute to fatigue in pilots. These include high mental workload, shift scheduling characteristics, reduced cockpit space, diversified maneuvers, alternating acceleration forces, poor airflow, low luminosity, and continuous background noise and vibration. The consequences of factors contributing to fatigue and sleepiness are likely to be potentiated when the individual is, at the same time, carrying out safety-critical activities that demand sustained attention and concentration.[1,4] It is emphasized that fatigue and sleepiness are often used as synonyms, but they differ because sleepiness is mitigated by sleep, while fatigue mitigation demands rest.[5]

Speech evaluation has been suggested as a promising method for detection of fatigue and sleepiness in real life operations,

From the Graduate Program in Neurosciences and the Center for Research on Speech, Acoustics, Language and Music (CEFALA), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil; and the Stress Research Institute, Stockholm University, and Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden.

This manuscript was received for review in April 2018. It was accepted for publication in January 2019.

Address correspondence to: Carla Aparecida de Vasconcelos, Neurosciences, Universidade Federal de Minas Gerais, Avenida Antônio Carlos, 6627, Department of Electronic Engineering, Rm. 2505, Belo Horizonte, Minas Gerais 31270901, Brazil; carlavfono@yahoo.com.br.

DOI: https://doi.org/10.3357/AMHP.5134.2019

such as aviation. It is a challenge to obtain reliable and accurate assessment of fatigue and sleepiness for pilots, particularly because invasive methods involving wearable sensors are not viable options. Several studies have proposed that acoustic analysis of speech might be a practicable method which is both noninvasive and does not interfere with the pilots work task.[6,10,12]

CASE REPORT

This study aimed at verifying if there were variations in the speech and phonatory aspects of a pilot who complained of fatigue and sleepiness prior to an accident. The pilot was a Brazilian man, 45 yr old, with no history of preexisting diseases. The accident occurred at 10 o'clock in the morning. The flight operation on the day of the accident was uneventful and there were no reports of problems with the airplane or any other complaints by the pilot. Near the time of landing, the pilot committed a technical failure that caused the aircraft to tilt down, with no time to recover. Due to ethical considerations, detailed information concerning the flight will be avoided in the text.

The Center for Investigation and Prevention of Aeronautical Accidents (CENIPA) made available the following audio recordings containing dialogues of the pilot: 1) in days off (private); 2) with the flight plan center 35 h before the air crash; and 3) during the flight. There were approximately 16 min of recordings from the control day (before the accident) and 30 min from the crash day. On that day, there were dialogues from the taxiing of the aircraft and the preparation for takeoff until the last communication with the Air Traffic Controller (ATC).

Acoustic analyses were carried out with PRAAT® version 5.3.85 (University of Amsterdam, Netherlands), a free package widely used by speech scientists, while MS-Excel and Minitab version 18 (Minitab Inc., State College, PA, United States) were used for statistical analysis. The objective was to compare the same acoustic parameters acquired during the day before the accident (control condition) with those acquired during the day of the accident. Thus, the design of the study involved a retrospective case-control for an individual comparison of voice and speech. Similar data for the ATC personnel was not available.

Considering the noisy and band-limited (3.4 kHz) recordings, only parameters robustly extracted in such conditions were analyzed. This excluded usual voice and speech parameters such as fundamental frequency, intensity, and formant patterns. Linguistic classification was used to identify syllables, whereas the syllabic division was phonetic. In Brazilian Portuguese, the minimum structure of a syllable is the nucleus (a vowel), which may be preceded by an onset or followed by a coda, both formed by consonants.

The entire audio was manually segmented in PRAAT for the analysis of the temporal organization of speech and the following parameters[3] were extracted:

- Total Articulation Time (TAT), composed by "fluent syllables," that is, speech without disfluencies such as silent pauses, repetitions, prolongations, hesitation, etc.;
- Total Pause Time (TPT), composed by "silent pauses" plus "filled pauses," that is, nonsilent disfluencies;
- Elocution Time (ET), also known as Total Speech Duration, defined by ET = TAT + TPT;
- Number of Pauses and Average Pause Duration;
- Elocution Rate (ER) defined by ER = NS/ET, where NS = Number of Syllables;
- Articulation Rate (AR) defined by AR = NS/TAT;
- Percentage of Disfluency (PD) defined by PD = ND/NS×100, where ND = Number of Disfluencies; that is, breaks, irregularities or nonlexical vocables that affect the flow of words. Filled pauses named "prolongations" were not subtracted from the number of fluent syllables in the calculation of articulation and elocution rates, but they have been considered in the Percentage of Disfluency, decreasing it. This procedure was chosen because the subtraction of what would be normal syllabic production from what was already prolongation would be arbitrary and it would be difficult to correctly calculate the TAT. Examples of acoustic speech measures are shown in **Fig. 1**.

After the syllabic segmentation, the parameters of the temporal organization were computed manually (Tiers 3 and 4 in Fig. 1). Next, the detailed typology of typical pauses was determined (hesitation, interjection, review, unfinished word, word repetition, repetition of segments, and repetition of phrases) and the respective percentages were calculated.

The Kolmogorov-Smirnov normality test was performed in all variables. The paired Student's *t*-test (5% significance level) was used for normally distributed variables and the Wilcoxon paired test (5% significance level) otherwise. Because the samples were extracted from the same speaker, dependence of the samples was assumed. Correlation measurements were also performed between variables. Pearson's correlation coefficient (r) was applied for parametric distributions and Spearman's correlation coefficient (x) for nonparametric cases.

DISCUSSION

The key results related to the temporal organization of speech are presented in **Fig. 2**. They show a decrease of Elocution Rate and Articulation Rate on the day of the accident compared to the day preceding the accident. Compared with normality patterns for spontaneous speech in males of the same dialectal variation (Minas Gerais state, Brazil), which is, on average, 5.5 syllables/s (Elocution Rate) and 6.5 syllables/s (Articulation Rate), there was also a decrease.

There was also a significant increase in the disfluency of the speaker on the day of the accident in comparison with the previous days, namely, in days off at home, and in the same working conditions 35 h before the plane crash. This is additionally evidenced by an increase in silent and filled pauses.

The results for the correlational analysis are presented in **Fig. 3**. Elocution Rate and Total Disfluency Time (left panel)

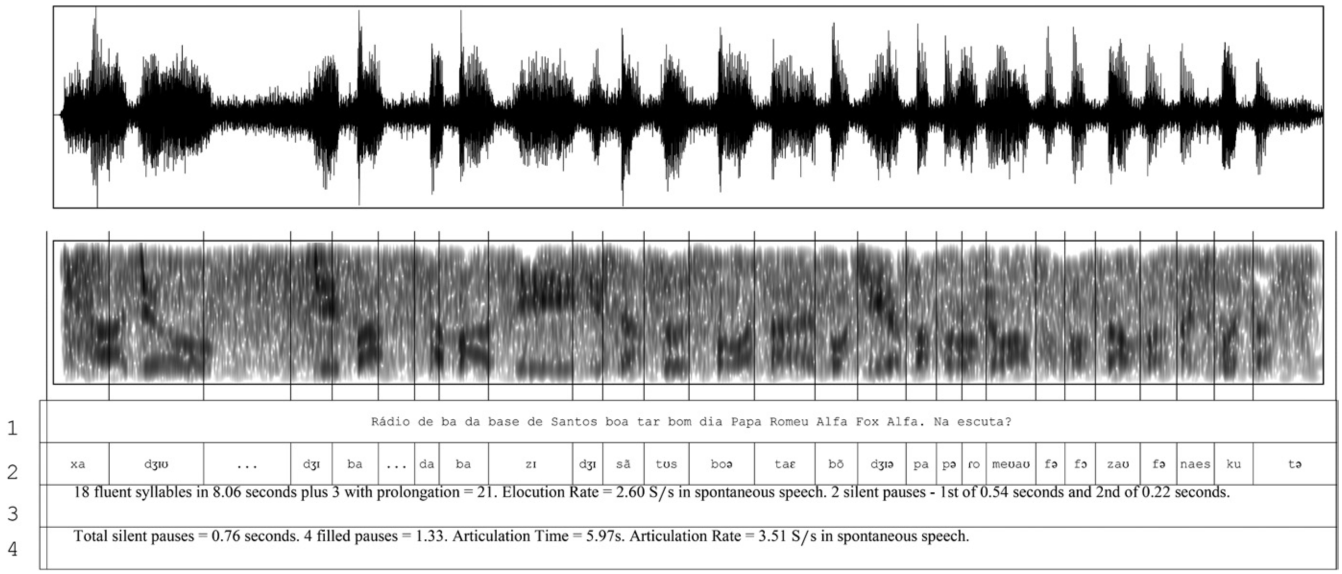

Fig. 1. PRAAT screen containing syllabic segmentation. The top graphic is the waveform and the corresponding spectrogram is shown below. Tier 1 is the pilot´s speech ("Radio… of ba... of the base of Santos, good afternoon, hmm, good morning. Papa Romeo Alpha Fox Alpha. Listening?"). Tier 2 shows the manual syllabic segmentation (syllables border indicated by vertical black lines) and the corresponding phonetic transcriptions; ellipses (...) indicate disfluent silent pauses. Tiers 3 and 4 are summaries of the manual measurements (durations, number of syllables, elocution and articulation rates, etc).

showed a strong ($0.7 \leq |r| < 0.9$) negative correlation ($r = -0.73$, $p < 0.001$). The coefficient of determination (r^2) is a measure of the proportion of variability in one variable that is explained by the variability of the other. In this case, $r^2 = 0.54$, meaning that 54% of the Elocution Rate was explained by the Total Disfluency Time. This demonstrates that not only did disfluency interfere with Elocution Rate, but there were also other causes. Articulation Rate is a likely factor in the 46% nonexplained variability, this parameter being lower on the day of the accident compared to the day before the accident, lowering also the Elocution Rate.

The Number of Pauses and the Number of Syllables (Fig. 3, right panel) showed a weak ($0.3 \leq |x| < 0.5$) Spearman correlation coefficient ($x = 0.47$), indicating that the increase in the Number of Syllables was partly associated with the increase in the Number of Pauses. This allows us to conclude that the pauses did not occur due to the size of the statement, but due to other reasons, leading us to infer that increased disfluency is a consequence of the state of fatigue and sleepiness in this individual.

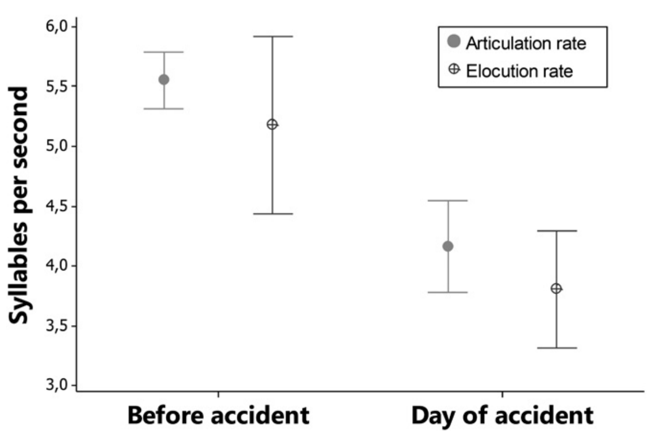

Fig. 2. Representation of the estimated confidence intervals of the Elocution Rate (in syllables per second) and of the Articulation Rate (in syllables per second) for the day prior to the accident and the day of the accident (*t*-test).

There are also interesting voicing (phonatory) aspects in the data. In addition to changes in the temporal organization of speech, the pilot presented vocalizations similar to vocal fry or creaky voice. In such vocalizations, though, the subglottic pressure and the resulting vocal intensity are extremely low and are not likely to occur in high background noise conditions similar to the aircraft cabin. In such cases, speakers are naturally forced to automatically increase their voice intensity, what is known as the Lombard effect.[8] Thus, the occurrence of vocal fry, a voicing aspect of speech, in Lombard effect conditions is not expected because the former demands low subglottic pressure and the latter the opposite.[11] In summary, the hypothesis of low responsive activity of the pilot to control feedback (vocal production–auditory perception) on the day of the accident is plausible.

Changes in the temporal organization of speech that make speech slower are triggered by the state of fatigue.[12] These changes occur in fatigue/sleepiness because in this situation there are changes in the patterns of muscular contraction and in the neurological speech commands.[12] Considering the results on the variations in the speech parameters of the pilot on the day of the accident (changes in temporal organization, disfluency, and voicing aspects) compared with the day before the accident and also speech normality patterns, it is concluded that the types of speech and voice analyses adopted in this study are promising for the detection of fatigue and sleepiness.

The results show significant variation between the speech produced in baseline conditions and just before the accident. Admittedly, information extracted from speech and voice

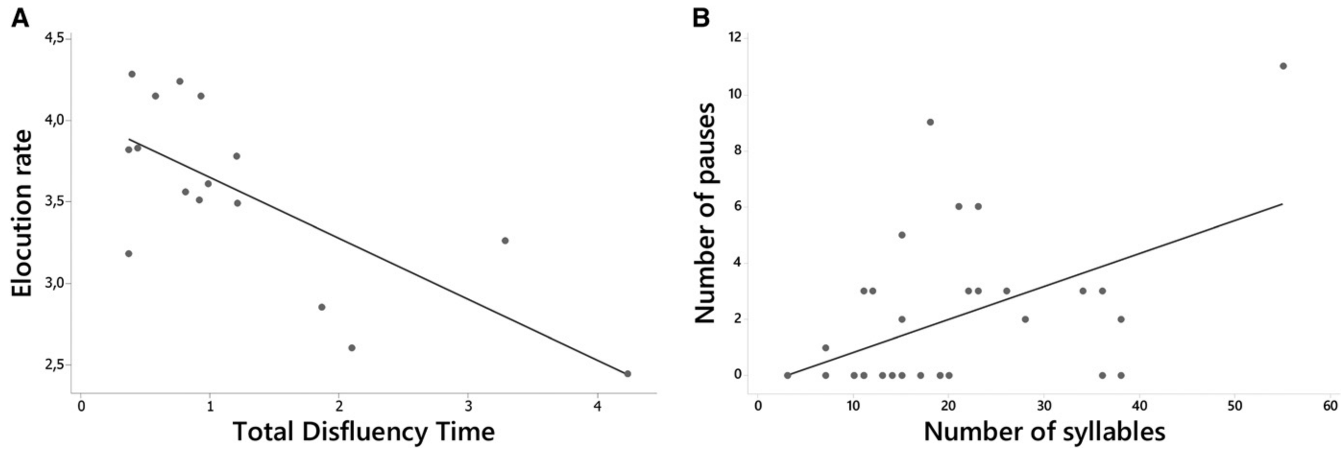

Fig. 3. A) Representation of the correlation between the variables Elocution Rate vs. Total Disfluency Time. B) Representation of the correlation between the variables Number of Pauses vs. Number of Syllables.

analyses is an indirect measure of fatigue and sleepiness. However, other methods based on physiological measures related to assessment of eye movements and electrical brain activity, besides being also indirect, are invasive for the pilots, more difficult to carry out, and less feasible than speech recording and analysis.

To conclude, the use of speech recordings and analysis for fatigue and sleepiness detection in airline pilots is a promising method suitable for situations when verbal communication between crewmembers or air traffic controllers frequently occurs. The analyses discussed in this paper can be used to mitigate accident risks by assessing pilot conditions before and during flights. For such, an automatic analysis of the elocution and articulation rates and of the pauses produced during speech can be done in real time. A Voice Activity Detector can be used to detect pauses, the patterns of which will be processed. Speech recognition techniques can be adapted to measure elocution and articulation rates. Parameters associated with voice quality require a more sophisticated analysis to detect, for example, the occurrence of vocal fry. Last, but not least, full-scale validation studies are needed to determine whether the results obtained in the present case report can be replicated over individuals and in other languages.

ACKNOWLEDGMENTS

The authors thank Paulo Licati from the National Commission of Human Fatigue of ABRAPAC for suggesting the investigation on the impact of mental fatigue on a pilot's speech; CENIPA for providing the data for analysis; and CAPES, CNPq, and FAPEMIG for financial support.

Authors and affiliations: Carla Aparecida de Vasconcelos, M.Sc., Ph.D. candidate, and Hani Camille Yehia, M.Sc., Ph.D., Graduate Program in Neurosciences, and Maurilio Nunes Vieira, M.Sc., Ph.D., Center for Research on Speech, Acoustics, Language and Music (CEFALA), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil; and Göran Kecklund, M.Sc., Ph.D., Stress Research Institute, Stockholm University, and Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden.

REFERENCES

1. Caldwell JA, Mallis MM, Caldwell JL, Paul MA, Miller JC, Neri DF. Fatigue countermeasures in aviation. Aviat Space Environ Med. 2009; 80(1):29–59.
2. Dinges DF. An overview of sleepiness and accidents. J Sleep Res. 1995; 4(S2):4–14.
3. Duez D. Organisation temporelle de la parole et dysarthrie parkinsonienne [Temporal organization of speech and Parkinsonian dysarthria]. In: Ozsancak C, Auzou P. Les troubles de la parole et de la déglutition dans la maladie de Parkinson. Marseille: Solal; 2005:195–211 [in French; abstract in English].
4. Greeley HP, Berg J, Friets E, Wilson J, Greenough G, et al. Fatigue estimation using voice analysis. Behav Res Methods. 2007; 39(3):610–619.
5. Jepsen JR, Zhao Z, Leeuwen WMAV. Seafarer fatigue: a review of risk factors, consequences for seafarers health and safety and options for mitigation. Int Marit Health. 2015; 66(2):106–117.
6. Krajewski J, Kroger BJ. Using prosodic and spectral characteristics for sleepiness detection. In: Proceedings of Interspeech 2007; Aug. 27–31, 2007; Antwerp, Belgium. Baixas (France): International Speech Communication Association; 2007; 8:1841–4.
7. Li X, Duc BD. Functional neuroimaging of circadian fatigue. Int J Comput Appl Technol. 2012; 45(2/3):156–163.
8. Lombard E. Le signe de l'élévation de la voix [The voice elevation sign]. Annales des Maladies de L'Oreille et du Larynx. 1911; 37(Pt. 1):101–109 [in French; abstract in English].
9. Shen J, Barbera J, Shapiro CM. Distinguishing sleepiness and fatigue: focus on definition and measurement. Sleep Med Rev. 2006; 10(1):63–76.
10. Shiomi K, Itano K, Suzuki A. Development and evaluation of the fatigue and drowsiness predictor. Archives of 28th International Congress of the Aeronautical Sciences; 23-28 Sept. 2012; Brisbane, Australia. Amsterdam (Netherlands): International Congress of the Aeronautical Sciences; 2012:1–7.
11. Södersten M, Ternström S, Bohman M. Loud speech in realistic environmental noise: phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers. J Voice. 2005; 19(1):29–46.
12. Vogel AP, Fletcher J, Maruff P. Acoustic analysis of the effects of sustained wakefulness on speech. J Acoust Soc Am. 2010; 128(6):3747–3756.