TECHNOLOGY DIGEST

TD-10-032
©OCTOBER 2010

TIMELY TECHNOLOGY TRANSFER

The work described in this document was performed by Transportation Technology Center, Inc., a wholly owned subsidiary of the Association of American Railroads.

Analysis of Speech Metrics to Estimate Crew Alertness Using Fuzzy Logic

Parham Shahidi, Steve C. Southward, and Mehdi Ahmadian
Railway Technologies Laboratory, Center for Vehicle Systems & Safety, Virginia Tech

Summary

A novel approach for estimating alertness levels of train operations personnel by analyzing speech signals during communications between the train crew and train dispatcher has been developed.

Communications during railroad operations were studied in the field during a train ride in an eastern railroad locomotive cab. Valuable insight was gained about the speech production process and its relation to background noise levels and railroad maneuvers.

A real-time deployable version of the algorithm was developed and deployed on a Matlab/Simulink xPC Target-based hardware platform. Initial testing was used to adapt parameters to improve the performance of the algorithm in real-time mode.

Test results show that digital signal processing and fuzzy logic can be implemented to detect changes in speech, which are indicative of alertness levels and that the membership functions for this purpose can be found empirically through iterative testing. Furthermore, this study proves that the framework to run such an analysis continuously as a monitoring function in locomotive cabs is feasible and can be realized with relatively inexpensive hardware. The tests conducted during the development process are explained and their results are presented in this *Technology Digest*.

Previous research has shown that sleep disorders, reduced hours of rest, and disrupted circadian rhythms can lead to increased fatigue levels that may manifest themselves in alterations of speech patterns[1, 2] as compared to alert states of mind. In this study, vocal indicators of fatigue were extracted from the speech signal and fuzzy logic was used to generate an estimate of the cognitive state of a locomotive engineer.

Communication between a locomotive engineer or train conductor and a train dispatcher over the radio provides an unobtrusive way of accessing the speech signal through the existing speech infrastructure. The speech signal is separated and processed through a digital signal processing algorithm, which extracts speech metrics from the signal that are determined to be indicative of fatigue levels. Speech metrics include, but are not limited to, speech duration, silence duration, word production rate, phrase gap duration, number of words per phrase, and speech intensity. A fuzzy logic minimum inference engine maps the inputs to an output through an empirically determined rule base. The rule base and the associated membership functions were derived from real-time testing and the subsequent tuning of parameters to refine the detection of changes in patterns.

©2010 Transportation Technology Center, Inc. Unauthorized duplication or distribution prohibited.

INTRODUCTION

The approach used to address the issue of fatigue in railroad operations described in this *Technology Digest* is based on the proposal that continuous monitoring of alertness levels during on-duty hours can help detect potentially dangerous situations ahead of time; and therefore, prevent accidents due to human errors associated with fatigue. To develop an alertness monitoring system with this capability, the following crucial constraints were identified:

1. Continuous monitoring: Conversations occur on a random basis. This fact requires a monitoring system that is able to continuously run in the background and monitor all conversations that occur. Therefore, the algorithm has to be lean and efficient enough to run in real-time mode.
2. Noninterpreting code: In order to protect the conductors' privacy rights, the algorithm has to be coded such that it does not interpret the speech, but rather uses other speech criteria such as speech energy levels, word production rate, etc. to monitor alertness.
3. Unobtrusive: To prevent the train conductors from feeling "observed" or obstructed during their regular work routine, the solution needs to be unobtrusive. This is a key aspect to minimize the impact of the monitoring system on the actual operation of the train itself.
4. Hardware: Because this work is addressing an issue that is not theoretical in nature, it is mandatory to design the solution in a manner that makes it feasible to deploy the system. Specifically, this means that pre-existing hardware needs to be utilized and computational requirements have to be kept to a minimum.
5. Background noise level. The algorithm included a noise reduction function to minimize the effects of background noise on the performance of the system. Because the background noise in a locomotive cab is relatively static and the conductors use personal handsets, there are multiple options to address this circumstance.

The proposed method (Figure 1) is centered on the train conductors because they communicate with the dispatcher through the locomotive cab radio. The speech signal sent back and forth between the conductor and dispatcher is the input to the algorithm on which the analysis is based.

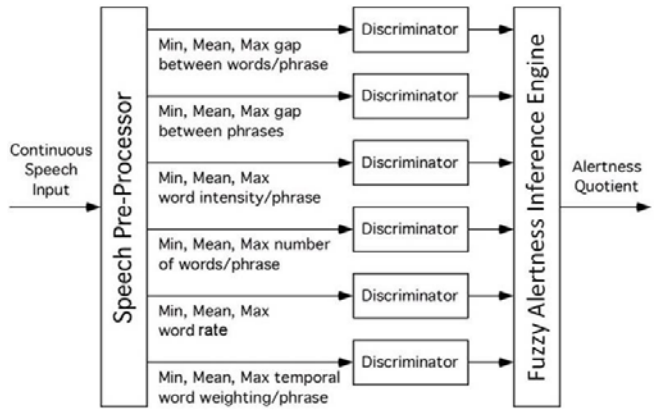

Figure 1. Proposed Crew Alertness Estimation Algorithm Overview

Results

A validation test was designed and conducted to examine the performance of the speech preprocessor subsystem of the algorithm. The input signal for this test was a noise free, hyper-articulated passage of text read aloud from a book. The test signal emphasized the following metrics: speech duration, silence duration, word production rate, and speech intensity (see red boxes in Figure 2). Each passage was read aloud in a manner such that only one metric was articulated at a time. The input signal and the preprocessor results are shown in Figures 2 and 3 respectively.

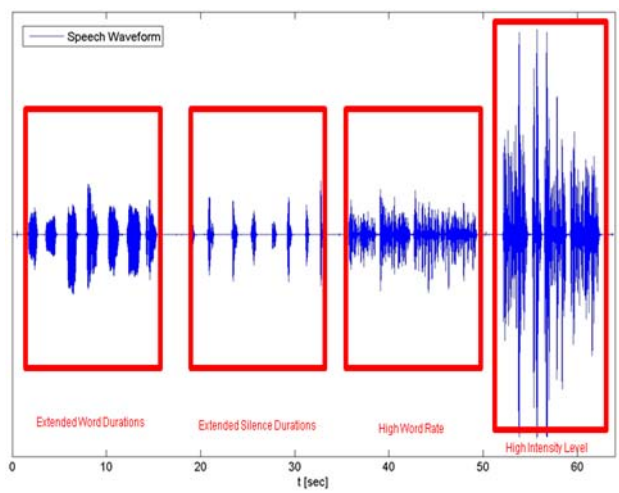

Figure 2. Test Signal for Evaluating the Speech Preprocessor

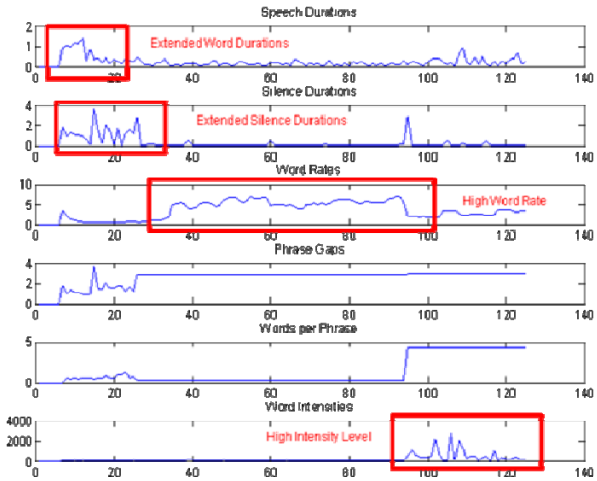

Figure 3. Extracted Segments of Hyper-articulated Speech

Figure 3 shows the results of the validation test. The algorithm detected each region of speech correctly and was able to visualize the changes in the speech metrics. Each detected metric exhibited an increase in either magnitude and/or duration for the segment it was detected in. The output for the test was generated on a per-word basis. Even though two of the metrics generated by the algorithm were not under test, they reflected the changes correctly.

The next steps in the development process were the implementation of the Fuzzy Logic Inference Engine[3] and the deployment of the algorithm on a real-time hardware platform. A 30-element, canonical rule base was established and implemented for the fuzzy logic element of the program. A *Minimum Inference Engine* with *Individual Rule-based Inference*, *Mamdani Minimum Implication*[4] and *max* union operator was used for the Fuzzy Logic Inference Engine.

$$\mu_{B'}(y) = \max_{l=1}^{M}\left[\sup_{x\in U}\min\left(\mu_{A'}(x), \mu_{A'_1}(x_1), \ldots, \mu_{A'_n}(x_n), \mu_{B'_l}(y)\right)\right] \tag{0.1}$$

Figure 4 shows a schematic of the hybrid model.

The membership functions for the inference engine were derived from a statistical analysis of a laboratory environment recording that classified ranges of low, medium, and high for each metric respectively. The metrics were divided into the aforementioned ranges and implemented through Gaussian normal distributions and Sigmoid functions. Then the whole system was implemented in a hybrid model in which all the previously mentioned parts were combined to be able to execute them in real time.

The model was prototyped with xPC Target and the Speedgoat Automation Real-Time Target Machine. The Speedgoat hardware platform is very similar to a standard personal computer. The system that was used for this study ran on an Intel Celeron-M 1 GHz central processing unit and had 256 megabytes of DDR ram and a data acquisition module in a ruggedized chassis, which made it deployable in the railroad operating environment. Apart from the ruggedized chassis, this system was in no way superior to a standard computer. Figure 5 shows the test setup.

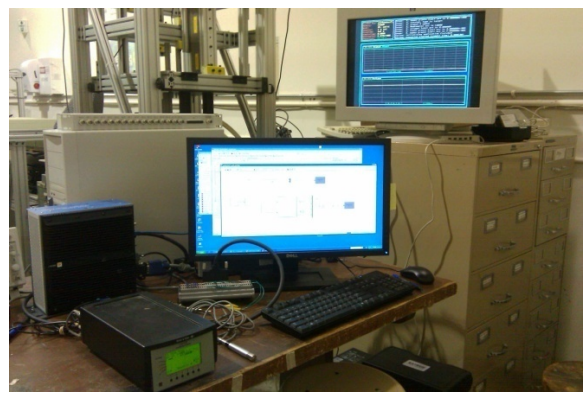

Figure 5. Hybrid Model Real-time Implementation Test Setup

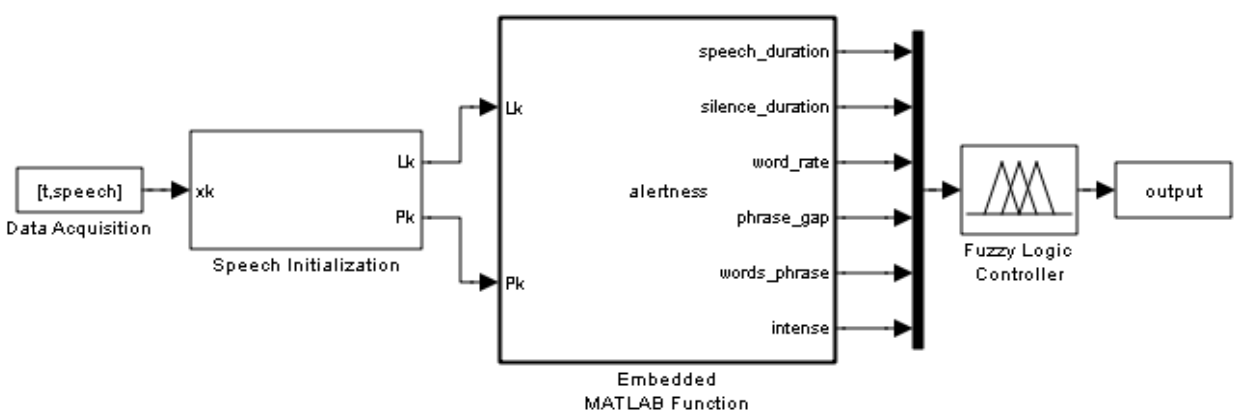

Figure 4. Overview of Hybrid Model Implementation

In addition to the host machine and the Speedgoat Target machine, the test setup included a single channel B&K Nexus conditioning amplifier, one B&K 4189 ½-inch free-field microphone, and one B&K 2669 ½-inch microphone preamplifier.

In Figure 6, the upper Target Scope was used to output the speech signal and the power envelope, and the lower Target Scope was used to output the Fatigue Quotient output from the Fuzzy Logic Controller. The test procedure was set up in a fashion such that once the algorithm was built and deployed on the target, the speech metrics indicative of fatigue for this set of metrics were mimicked to see the target's response. A solid recognition of changes in the speech pattern were observed. Even though the range was limited and never went down to zero, it exhibited larger changes than the batch mode test.

The approach presented here was shown to work, and an initial framework for processing speech to draw conclusions about the alertness level of the subject under test has been established. With tuning, it was possible to improve the algorithm's accuracy to the point where it detects, within the limits of speech processing algorithms,[5] changes in the speech pattern, mimicked through hyper-articulation, and relate these changes to a desired conclusion through fuzzy logic.

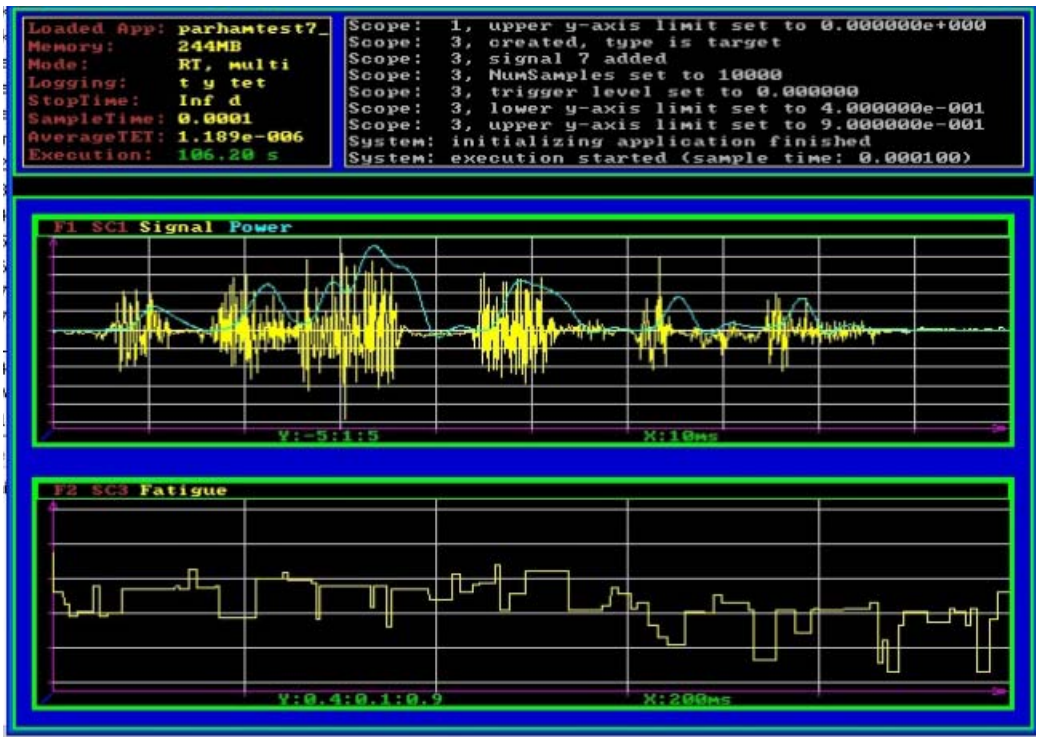

Figure 6. Real-time Implementation Outputs

Future Work

Further proof testing, in terms of speaker independency and accuracy of the system, is needed to establish the system's reliability to assess train crew alertness. The proposed methodology will test system performance with respect to different speakers and different alertness states.

Electroencephalography devices will be investigated as a benchmarking tool to assess the performance of the algorithm under controlled conditions with human subjects. The results of this validation study will be used to further improve the existing rules and add more heuristic rules to fuzzy rule base.

Furthermore, a set of data collected in a clinical trial can be looked upon as a set of stipulated input/output pairs that can be utilized to implement adaptive techniques to make the algorithm self-learning. The incorporation of pitch estimation in real time as another speech metric will be explored as well.

References

1. Morris, Gary O., Harold L. Williams, and Ardie Lubin. 1960. "Misperception and disorientation during sleep deprivation." *Archives of General Psychiatry* 2: 247-254.

2. Harrison, Y. and J. A. Horne. 1997. "Sleep deprivation affects speech." *Sleep* 20(10): 871-877.

3. Zadeh, L. A. 1964. *Fuzzy Sets*. United States: p. 2p.

4. Wang, Li-Xin. 1997. *A Course in Fuzzy Systems and Control.* Upper Saddle River, N.J.: Prentice Hall PTR: 424.

5. Lippmann, R. P. 1997. "Speech recognition by machines and humans." *Speech Communication* 22(1): 1-15.

Disclaimer: Preliminary results in this document are disseminated by the AAR/TTCI for information purposes only and are given to, and are accepted by, the recipient at the recipient's sole risk. The AAR/TTCI makes no representations or warranties, either expressed or implied, with respect to this document or its contents. The AAR/TTCI assumes no liability to anyone for special, collateral, exemplary, indirect, incidental, consequential or any other kind of damage resulting from the use or application of this document or its content. Any attempt to apply the information contained in this document is done at the recipient's own risk.