# Speed Change Classification for Engines in Vehicles

Elżbieta Kubera[1], Alicja Wieczorkowska[2], Tomasz Słowik[3], Andrzej Kuranc[3], and Krzysztof Skrzypiec[4]

[1] University of Life Sciences in Lublin, Department of Applied Mathematics and Computer Science, Akademicka 13, 20-950 Lublin, Poland
elzbieta.kubera@up.lublin.pl
[2] Polish-Japanese Academy of Information Technology, Koszykowa 86, 02-008 Warsaw, Poland
alicja@poljap.edu.pl
[3] University of Life Sciences in Lublin, Department of Energetics and Transportation Akademicka 13, 20-950 Lublin, Poland
[tomasz.slowik][andrzej.kuranc]@up.lublin.pl
[4] Maria Curie-Skłodowska University in Lublin, Pl. Marii Curie-Skłodowskiej 5, 20-031 Lublin, Poland
krzysztof.skrzypiec@poczta.umcs.lublin.pl

**Abstract.** Vehicle speed is an important factor influencing highway traffic safety. Radars are applied to control the speed of vehicles, but the drivers often decelerate when approaching radar, and then accelerate after passing by. In this research, we address automatic recognition of speed change from audio data, based on recordings taken in controlled conditions. Data description, acquisition details, and classification experiments illustrate both changing speed and maintaining constant speed. This is a starting point to investigate what percentage of drivers actually maintain constant speed, or slows down only to speed up immediately afterwards. Automatic classification and building an appropriate database can help improving traffic safety.

**Keywords:** Intelligent transport system; Road Traffic Safety; Audio signal analysis

## 1 Introduction

Development of transport brings numerous benefits, but it also brings problems, including pollution of the environment, and decreased safety. According to WHO (World Health Organization), more than one million people die each year as a result of road traffic crashes, and people aged 15-44 years account for almost a half of global traffic deaths [11]. Pedestrians, cyclists and motorcyclists are especially vulnerable road users. Road crashes also cause economic losses. Additionally, people from low- and middle-income backgrounds and countries are more often involved in road crashes.

Improving the safety of roads and enhancing the behavior of road users became priority in all actions related to transport in Poland and around the world. Crashes and dangerous situations are carefully analyzed, and the society is informed about the results, which sometimes evoke heated discussions. The use of photo radars, traffic calming zones, speed bumps and other means of improving traffic safety are often criticized by the drivers. However, the statistics confirm that road safety improves, and the number of fatalities in road crashes decreases. The statistics of fatalities in road traffic crashes for selected European countries are shown in Figure 1 [4]. According to Central Statistical Office of Poland, the number of fatalities in road crashes decreased in last years, with 3348, 3190 and 2933 fatalities in 2013, 2014 and 2015 respectively.

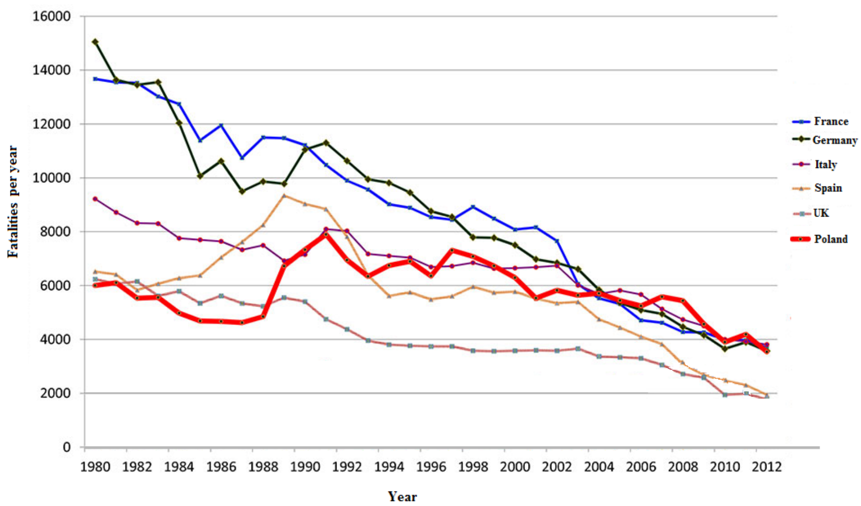


**Fig. 1.** Fatalities in road traffic crashed in selected European countries [4]

The road traffic safety can be improved. The main factors that decrease safety are

- low quality of many roads, and of vehicles;
- lack of protection for pedestrians and bikers,
- deficiencies of the road traffic safety systems - prolonged implementation of laws relating to risk factors, insufficient financial support, low public awareness;
- reckless behavior of drivers and other road users - speeding, driving and walking while impaired; not using seat belts.

Therefore, road safety can be improved through setting and enforcing appropriate regulations, through public awareness campaigns, and through interventions targeting the road users behavior. The reports from the operators of navigation systems show that drivers increase speed when passing a speed camera, after slowing down when approaching the camera. However, these report describe the behavior of the users of these navigation systems, and further research on other drivers' behavior is needed. The data on drivers' behavior can be collected from audio recording systems, if such systems are developed and deployed. Therefore, the goal of our research is to investigate if changing (or maintaining) speed can be recognized from audio data, as this can be the first step to prepare a system to collect data on drivers' behavior around speed cameras. Additionally, vehicle speed influences the noise level generated by the vehicle, as the speed increase also increases the noise level. Another factors increasing the noise level are:

– road inclination - vehicle traveling uphill generates more noise, but average traffic speed decreases; traveling downhill generates less noise than traveling on flat road,
– road surface, and its dampness,
– the technical condition of vehicles; especially old vehicles generate more noise and more exhaust fumes, and these vehicles also decrease the traffic road safety.

Out of all the factors influencing the road traffic safety and noise generated, the speed of vehicles can be more easily addressed than some other factors. This is why we decided to address the behavior of drivers when speed is controlled, and prepare audio data representing acceleration, deceleration, and constant speed. Next, we trained classifiers on speed data. This is the initial step of our research, aiming at automatic classification of speed change, and collecting a bigger dataset of drivers' behavior. These data are very complex, since audio signal recorded depends on many factors, and it changes with time. The recorded data represent changes of the amplitude of the audio waveform with time; close-up of a segment of such audio data is shown in Figure 2. As we can see, data change quickly, and even though we can observe some periodicity, irregular variations are also clearly visible. Time domain data are usually parameterized, based on frequency content (i.e. on the spectrum). The spectrum is usually calculated for short segments of data, and changes of the spectrum over time can be can be observed in so called spectrogram, where the amplitude of particular frequencies is represented using a selected color scale. Exemplary spectrogram in grayscale are shown in Figure 3.

As we can see, representing vehicle sound contains noise and harmonic components. This is why we designed a feature vector that allows capturing noise features, harmonic features, and their changes in time. Speed is usually monitored by radars, but usually a single measurement is done, and the drivers are aware of this fact. We hope that monitoring how drivers change speed in these circumstances and launching public awareness campaign can increase the road traffic safety.

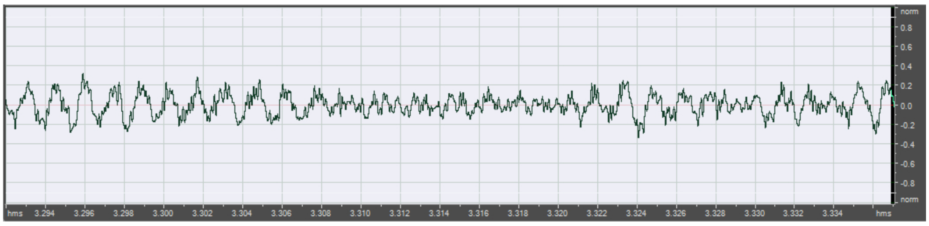


**Fig. 2.** Audio data in time domain: horizontal axis represents time, vertical - amplitude. Recording for Ford Focus (plot from Adobe Audition [1])

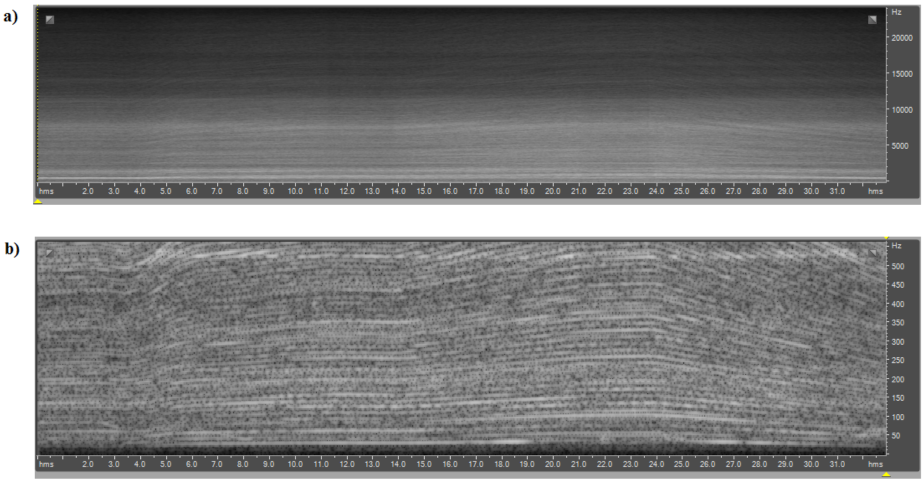


**Fig. 3.** Spectrogram of Ford Focus accelerating, then decelerating, accelerating again and decelerating, for 1 Newton load (a). A close-up for low frequencies is shown in (b). Higher luminance represents higher amplitude. Spectrograms obtained from Adobe Audition [1]

## 2 Data

In order to perform the data in controlled conditions, dyno test bench (see Figure 4) recordings were performed in May and June 2016, at the University of Life Sciences in Lublin. The position of the vehicle, OBD (On-Board Diagnostics) acquisition and audio recorder is shown in Figure 5. Eight vehicles were recorded:

– Smart ForFour - car with gasoline engine,
– Ford Focus - car with gasoline and LPG (Liquid Petroleum Gas) engine,
– Hyundai i30 - car with Diesel engine,
– Toyota Corolla Verso - car with Diesel engine,
– Daewoo Lublin - van with Diesel engine,
– Fiat Ducato - van with Diesel engine,
– Volkswagen (VW) Transporters, 2004 and 2007 year - vans with Diesel engines.

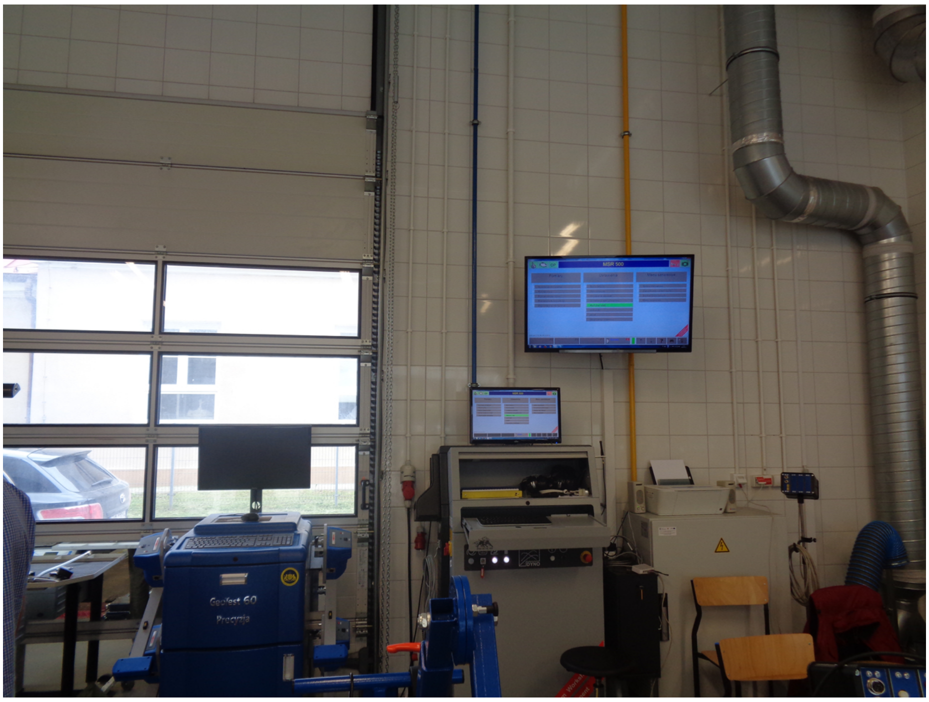

**Fig. 4.** Dyno test bench equipment at the University of Life Sciences in Lublin

All vehicles were equipped with manual transmission. The vehicles accelerated to 110 km/h (with the exception of Daewoo Lublin - to 90 km/h only), then decelerated to low speed at fifth gear, to about 40-45 km/h, when the gear was

**Fig. 5.** Dyno test bench data recording, with OBD acquisition shown. The audio recorder is placed on the tripod

changed. When accelerating, the driver changed gear attempted to maintain a constant speed for a few seconds at 50 km/h, 70 km/h, and 90 km/h. Also, at 110 km/h constant speed was maintained for a few seconds. This way we obtained data representing three categories: accelerating, decelerating, and maintaining constant speed. Two versions were recorded: with 450 Newton load applied, with the exception of Ford Focus, and load adjusted to on-road conditions (with the exception of Daewoo Lublin and Hyundai i30), i.e. depending on vehicle speed, weight, and road coefficients.

## 3 Feature Set

In order to capture changes of audio signal properties with speed changes, if any, we decided to observe these properties within 1-second audio frames. For each frame, we calculate spectral features for the starting 330 ms sub-frame, and for the ending 330 ms sub-frame. Next, the calculated features for the starting part, together with the vector of differences between features for the starting and ending part, are placed in the feature vector. Additionally, one time-domain parameter is calculated, i.e. zero-crossing rate, together with its change between the value for the beginning and the ending part. Also, a parameter capturing the spectrum change between the starting and ending part is added to the feature set.

Features describing spectrum and time domain for the starting 330m ms of a 1-second frame include:

- *Zero Crossing Rate* (ZCR) in the time-domain of the sound; a zero-crossing is a point where the sign of the function (amplitude vs. time) changes;
- *Audio Spectrum Envelope* - 33 features, SE0, ..., SE32 calculated according to [8] as sums of the energy of the power spectrum within logarithmically spaced frequency bands;
- *SUM_SE* - sum of the spectrum envelope values;
- *MAX_SE_V, MAX_SE_IND* - value and index of the spectrum envelope maximum;
- *Audio Spectrum Flatness*, $flat_1, \ldots, flat_{25}$ - a vector describing the flatness property of the power spectrum [8], i.e. the deviation of the signals power spectrum from a flat shape, for each band of the spectrum envelope; bands up to SE25 were used, so we ignore higher frequencies, as in the research on audio data for vehicles the spectrum is usually limited, and even the sampling rate is often decreased in order to limit the amount of data analyzed;
- *MFCC* - 13 mel frequency cepstral coefficients. The cepstrum is calculated as the logarithm of the magnitude of the spectral coefficients, and then transformed to the mel scale, reflecting the properties of the human perception of frequency. Twenty-four mel filters were applied, and the obtained results were transformed (averaged) to 12 coefficients. The $13^{th}$ coefficient is the 0-order coefficient of MFCC, corresponding to the logarithm of the energy [6];

- *F0_ACor,* - fundamental frequency, calculated from the autocorrelation function. This parameter is included in the feature set, since the frequency of wheels rotation is visible in the spectrum (with its multiples), and this frequency changes with the speed change;
- *EnAb4kHz* - proportion of the spectral energy above 4kHz to the entire spectrum energy;
- *Energy* - energy of the entire spectrum;
- *Audio Spectrum Centroid* (SC) - the power weighted average of the frequency bins in the power spectrum. Coefficients were scaled to an octave scale anchored at 1 kHz [8];
- *Audio Spectrum Spread* (SS) - RMS (root mean square) of the deviation of the log frequency power spectrum with respect to *Audio Spectrum Centroid* [8];
- *RollOff* - the frequency below which 85% (experimentally chosen threshold) of the accumulated magnitudes of the spectrum is concentrated,
- *BW_10dB, BW_20dB, BW_30dB* - bandwidth of the frequency band comprising the spectrum maximum (in dB scale) and the level drop by 10, 20 and 30 dB, respectively, towards both lower and upper frequencies.

All these features are placed in the feature vector. Next, they are also calculated for the ending 330 ms of the 1-second frame, and

- the differences between these values and the values for the starting part are added to the feature vector;
- additionally, *Flux1* parameter is added to the feature set. *Flux1* is the sum of squared differences between the magnitudes of the spectrum points calculated for the starting and ending 330 ms sub-frames within the one-second frame.

Altogether, the feature vector consists of 169 features.

Audio data were recorded in stereo, with 48 kHz sampling rate, and 24 bit resolution. Fast Fourier transform was used for spectrum calculation, and sliding frame with 330 ms hop size was applied when analyzing the audio data and calculating the feature vector. The applied 330 ms analyzing frame yields good frequency resolution, which is needed when calculating low frequencies, representing wheels rotation.

### 3.1 Feature Selection

Since our feature vector is relatively large, we also performed feature selection, as it is recommended in such a case [5]. For each of the classifiers investigated, cross validation procedure was applied. We applied 8-fold cross validation, and each fold represented data from one vehicle. Since we used random forests as classifiers, we decided to apply feature importance from these classifiers. We tested 2 versions: with constant number of features to be selected (10 features; number arbitrarily chosen), and with feature importance above various threshold based of mean decrease of Gini criterion. The threshold 0.01 was selected, as giving a small number of features. After threshold based selection, the number of features left varied from several to more than 20 features.

## 4 Classification

The experiments on classification of the acquired data were performed using random forests (RF, [2]), deep learning (DL) architecture (neural network), and support vector machines (SVM), using R [9]. The data were classified into three classes:

- deceleration, with speed decrease more than 5 km/h per second; this class represents fast deceleration, often happening when the driver must quickly reduce speed when speeding before radar registers speed;
- stable speed, with speed change within 3 km/h range per second, originating for speed increase below 1.5 km/h per second or speed decrease below 1.5 km/h per second;
- acceleration, with speed increase more than 2.5 km/h per second.

The data were labeled according to OBD data and other information recorded at the dyno test bench. As we can see, the speed ranges do not represent neighboring intervals, since we wanted to capture clear cases of intent acceleration or deceleration. The remaining data were not taken into account in our experiments. Altogether, we had 101 examples (1-second frames) representing deceleration, 423 examples for acceleration, and 579 examples for stable speed. Since deceleration was underrepresented, compared to the other classes, we also performed upsampling during training (i.e. replicated examples, to match the number of examples in each fold with the biggest class), in order to balance classes.

### 4.1 Classifiers

Random forest (RF) is the classifier which yielded the best results in our experiments. RF is a set of decision trees, and bias and correlations between the trees are minimized during the classifier construction. Each tree is built without pruning, using a different $N$-element bootstrap sample (i.e. obtained through drawing with replacement) of the $N$-element training set. For a $K$-element feature set, $k$ features are randomly selected for each node of any tree ($k \ll K$, often $k = \sqrt{K}$). The best split on these $k$ features is applied to split the data in the node. Gini impurity criterion is applied (minimized) to choose the split. This criterion measures of how often an element would be incorrectly labeled, if random labeling according to the distribution of labels in the subset is applied. The forest of $M$ trees is obtained by repeating this procedure $M$ times; M=500 in our experiments, which is a standard setting in R. Classification using RF is performed by simple voting of all trees.

DL neural network in our experiments is a multi-layer feedforward neural net, with many hidden layers, with data standardization. Training is performed through back propagation with adaptive learning. We used R package h2o [7], with standard settings; training parameters include large weight penalization and drop-out regularization (ignoring a random fraction of neuron inputs). In training, weights are iteratively updated in so-called epochs, with grid-search of the parameter space. DL classifiers yielded good results in our previous research on audio data for vehicles [10].

We also applied SVM classifiers, which look for a decision surface (hyperplane) maximizing the margin around the decision boundary. The training data points are called support vectors. SVM projects data into a higher dimensional space, using a kernel function. We applied radial basis functions (RBF), with $c$ and $\gamma$ parameters. In our experiments, we applied $\gamma$=0.03125 and $c$=2.

## 5 Experiments

In our experiments, we trained RF, DL and SVM classifiers to recognize acceleration, deceleration and stable speed for our data. Since the feature vector is large, we also performed feature selection in 2 versions, i.e. with 10 best features kept, and with features above 0.01 threshold. The results are shown in Table 1. As we can see, the results for SVM without feature selection are low, but after feature selection the accuracy is better than for DL.

**Table 1.** Classification accuracy for speed changes

| Classifier | RF | DL | SVM |
|---|---|---|---|
| No feature selection | 70.6% | 60.9% | 37.2% |
| Top 10 features | 68.8% | 58.7% | 62.6% |
| Features above threshold | 70.1% | 55.5% | 64.9% |

We repeated these experiments after upsampling of the data. The results are shown in Table 2. As we can see, upsampling increases accuracy in some cases, and our best result was obtained for RF with feature selection above the threshold. However, upsampling decreases accuracy for DL with feature selection with top 10 features, which means that such a small amount of features is insufficient to classify such complex data.

**Table 2.** Classification accuracy for speed changes after data balancing (upsampling of training data)

| Classifier | RF | DL | SVM |
|---|---|---|---|
| No feature selection | 66.9% | 64.4% | 36.4% |
| Top 10 features | 67.3% | 46.1% | 60.7% |
| Features above threshold | 72.6% | 68.9% | 65.7% |

The obtained results show how difficult it is to classify such data, but looking into details in confusion matrices (see Table 3) we can see that classifiers most often mistake stable speed with acceleration. Accelerating is a relatively slow process, so such mistakes are not surprising. Also, deceleration is mistaken with stable speed, but the deceleration we recorded was a safe one, with engine braking, without applying brakes. The acceleration vs. deceleration mistakes are rare, and deceleration was never mistaken with acceleration. Confusion matrix for DL is shown in Table 4 for comparison; as we can see, the performance of DL is much worse for deceleration class.

**Table 3.** Confusion matrix for RF with feature selection above threshold, with up-sampling in training

| Class/Identified as: | Acceleration | Stable speed | Deceleration |
|---|---|---|---|
| Acceleration | 300 | 118 | 5 |
| Stable speed | 106 | 447 | 26 |
| Deceleration | 0 | 47 | 54 |

**Table 4.** Confusion matrix for DL with feature selection above threshold, with up-sampling in training

| Class/Identified as: | Acceleration | Stable speed | Deceleration |
|---|---|---|---|
| Acceleration | 264 | 152 | 7 |
| Stable speed | 454 | 120 | 5 |
| Deceleration | 25 | 34 | 42 |

## 6 Summary and Conclusions

In this paper, we aimed at automatic recognition of accelerating and decelerating of vehicles. The data for analysis were recorded at dyno test bench, so we had access to the speed data and rpm (revolutions per minute) information. This is the initial stage of the research investigating how drivers behave when approaching radars, and then after leaving the speed controlling zone. The experimental results show that classifiers can distinguish accelerating, decelerating and stable speed with about 70% accuracy, with accelerating and decelerating rarely mistaken. Also, we recorded safe braking (engine braking) in dyno test bench. Dramatic braking when drivers are speeding and must quickly decelerate before approaching a speed camera is much faster, and we anticipate easier recognition of fast deceleration when such data are recorded.

**Acknowledgments.** This work was partially supported by the Research Center of PJAIT, supported by the Ministry of Science and Higher Education in Poland.

## References

1. Adobe, `http://www.adobe.com/#`

2. Breiman, L.: Random Forests. Machine Learning 45, pp. 5–32 (2001)
3. Central Statistical Office of Poland, `http://stat.gov.pl/en/`
4. EuroRAP, European Assessment Programme `http://www.eurorap.pl/index.php`
5. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, Springer (2009)
6. Niewiadomy, D., Pelikant, A.: Implementation of MFCC vector generation in classification context. J. Applied Computer Science, Vol. 16, No. 2, pp. 55–65 (2008)
7. Package 'h2o', `http://cran.r-project.org/web/packages/h2o/h2o.pdf`
8. The Moving Picture Experts Group: MPEG-7, `http://mpeg.chiariglione.org/standards/mpeg-7`
9. The R Foundation, `http://www.R-project.org`
10. Wieczorkowska, A., Kubera, E., Słowik, T., Skrzypiec, K.: Spectral Features for Audio Based Vehicle Identification. In: New Frontiers in Mining Complex Patterns, 4th International Workshop, NFMCP 2015, 163–178. Springer, LNAI 9607 (2016)
11. World Health Organization: Road Traffic Injuries, `http://www.who.int/mediacentre/factsheets/fs358/en/`