Foreign Object Detection in Tires by Acoustic Event Detection

Jens Schröder[1], Moritz Brandes[1], Danilo Hollosi[1], Jan Wellmann[1],
Marten Wittorf[2], Oliver Jung[2], Volker Grützmacher[2], Stefan Goetze[1]
[1] *Fraunhofer IDMT, 26129 Oldenburg, Germany, E-Mail: jens.schroeder@idmt.fraunhofer.de*
[2] *Adam Opel AG, 65423 Rüsselsheim, Germany*

Introduction

Nowadays cars are equipped with various different sensors, e.g., radar or ultrasonic sensors for distance measuring and parking assistance, cameras for light-control or lane departure alerting, traffic sign recognition etc. Acoustic sensors find their way into cars mainly for automatic speech recognition (ASR) for in-car communication, entertainment and navigation systems [1]. However, acoustical sensors are capable of recognizing many more car-related events like sirens [2] and car horns [3], lane departure, road or tire conditions, engine failures, broken or worn chassis parts, loose parts, broken tie rods, broken/cracked exhaust, worn clutch etc.

In this paper, acoustically based detection of foreign objects in tires is investigated. This application scenario is important since objects in tires can cause punctures and, thus, lead to accidents in the worst case. The already existing pressure sensors for tires installed at the tires themselves can only detect an already existing critical air loss. However, microphones can detect the tire damage before a significant loss of air pressure is measurable. For this study, a sound database using acoustic sensors at different car positions has been recorded in real environments. Different algorithms to recognize objects in tires will be compared in this paper.

Detection of Nails in Tires

If an object is stuck to the cap of a tire, e.g., a screw, it will cause a periodically repeating ticking sound with the frequency of the spinning tire. Thus, we consider an approach that is based on the comparison of the frequency of the acoustic signal and the tire frequency to detect a foreign object. A block diagram of the frequency estimation based on acoustical signals is depicted in Figure 1.

Therefore, the sampled signal $s(n)$ is processed blockwise, i.e.,

$$y(n,\ell) = s(n+\ell\cdot n_\mathrm{s}) w_N^\mathrm{rect}(n) \tag{1}$$

with n denoting the sampling index, ℓ the block index, n_s the sample shift and $w_N^\mathrm{rect}(n)$ a rectangular window of length N. The influence of N will be investigated in detail. A signal block of 3 s duration is depicted in Figure 2, upper panel.

Since traffic and environmental noises often have much energy at low frequencies, a pre-emphasis filter is applied, yielding

$$y'(n,\ell) = y(n,\ell) - \gamma\cdot y(n-1,\ell) \tag{2}$$

where γ is a factor which controls the relation between high and low frequency energies. The pre-emphasized signal is shown in Figure 2, middle panel.

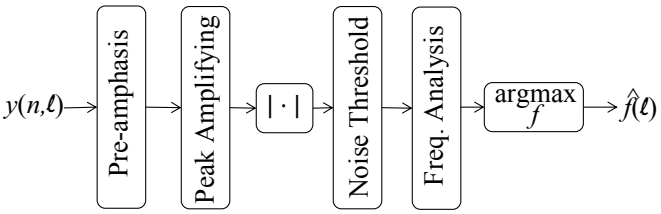

Figure 1: Overview of the frequency estimation procedure. The signal $y(k,\ell)$ is filtered by a pre-emphasis filter. A peak amplifying method is applied that is either based on auto regression or short-time averaging. Noise reduction is applied to the absolute value of the result. A frequency analysis is conducted afterwards. Four approaches have been tested based on DFT, Comb-filtering (Comb), modulation spectrogram (mod. STFT), modulation Mel-spectrogram (mod. Mel). The frequency with maximal energy $\hat{f}(\ell)$ is the outcome of the estimation procedure.

The ticking sound from objects in a moving tire results in periodical peaks in the time representation of the signal. A mechanism to amplify peaks and inhibit other signal components is applied. Two approaches have been tested. For the first approach, the signal $y'(n,\ell)$ is averaged for short windows of length N^av and subtracted from $y'(n,\ell)$, i.e.,

$$\hat{y}^{(\mathrm{av})}(n,\ell) = y'(n,\ell) - \frac{1}{N^\mathrm{av}} \sum_{i=-0.5N^\mathrm{av}}^{0.5N^\mathrm{av}} y'(n+i,\ell). \tag{3}$$

The other approach utilizes the auto regression function yielding

$$\hat{y}^{(\mathrm{reg})}(n,\ell) = y'(n,\ell) - \sum_{i=0}^{N^\mathrm{reg}-1} a_i y'(n-i,\ell) \tag{4}$$

where a_i denotes regression coefficients and N^reg the regression dimensionality. The resulting signal is depicted in Figure 2, lower panel. The noise contained in the absolute value of this resulting signal $|\hat{y}(n,\ell)|$ is reduced by a threshold, i.e.,

$$\tilde{y}(n,\ell) = \begin{cases} |\hat{y}(n,\ell)| & \text{for } |\hat{y}(n,\ell)| > \alpha\cdot\sigma_\ell + \mu_\ell, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

with μ_ℓ and σ_ℓ denoting the mean and standard deviation of the signal at block ℓ, respectively.

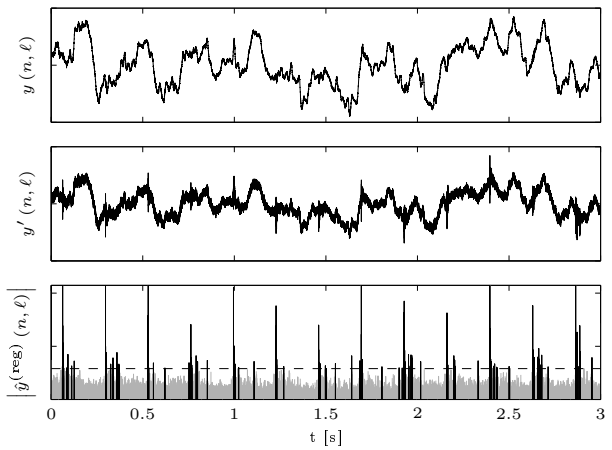

Figure 2: Pre-processing of the signal. In the upper panel, a signal block $y(n,\ell)$ of 3 s duration is shown. The pre-emphasized signal $y'(n,\ell)$ is depicted in the middle panel. The signal $\left|\hat{y}^{(\mathrm{reg})}(n,\ell)\right|$ after applying the auto regression method is depicted in gray in the lower panel. The signal components above the threshold (horizontal, dashed line), that are further processed, are marked in black.

After pre-processing, the frequency modulations are analyzed. The first approach is based on the DFT

$$\tilde{y}(k_1,\ell) = \sum_{n=-\infty}^{\infty} \tilde{y}(n,\ell) w_N^{\mathrm{Hamm}}(n+\ell\cdot n_{\mathrm{s}}) e^{-j2\pi k_1 n/N} \tag{6}$$

with $w_N^{\mathrm{Hamm}}(n)$ denoting a Hamming window of length N and $k_1 = 0\ldots N-1$ indexing the frequency bins. The dB-scaled spectrum is used, i.e,

$$Y_1(k_1,\ell) = 10\log_{10}\left(\tilde{y}(k_1,\ell)\right)^2. \tag{7}$$

The second approach applies a comb filterbank to the spectrogram of Eq. (6). A feedback comb filter is defined in the frequency domain as

$$H^{\mathrm{comb}}(k_1,k_2) = \left(1-\beta e^{-j2\pi k_1/k_2}\right)^{-1}, \tag{8}$$

where k_2 is indexing the comb frequency and β is a gain factor. Hence, the spectrum using a comb filter bank is defined by

$$Y_2(k_2,\ell) = 10\log_{10}\left(\sum_{k_1} H^{\mathrm{comb}}(k_1,k_2)\tilde{y}(k_1,\ell)\right)^2. \tag{9}$$

The third approach computes a spectrogram for each frame ℓ, i.e.,

$$Y'(k',l,\ell) = \sum_{n=-\infty}^{\infty} \tilde{y}(n,\ell) w_L^{\mathrm{Hamm}}(n+l\cdot l_s) e^{-j2\pi k' n/L}. \tag{10}$$

where l is a subframe index, L is the number of samples per subframe ($\equiv$ 25 ms) and l_s the hop size ($\equiv$ 10 ms). The modulation spectrogram is obtained by applying a Fourier transform to the subframes l

$$Y''(k',k_3,\ell) = \sum_{l=-\infty}^{\infty} Y'(k',l,\ell) w_{L/2}^{\mathrm{Hamm}}(l) e^{-j2\pi k_3 k'2/L}. \tag{11}$$

The resulting amplitude modulation spectrogram is summed over the frequency bands, i.e.,

$$Y_3(k_3,\ell) = 10\log_{10}\left(\sum_{k'} Y''(k',k_3,\ell)\right)^2. \tag{12}$$

The fourth approach applies a Mel-filterbank $H^{\mathrm{mel}}(k_4,k')$ to the spectrum $Y'(k',l,\ell)$, i.e.,

$$Y'''(k_4,l,\ell) = \sum_{k'} H^{\mathrm{mel}}(k_4,k') Y'(k',l,\ell), \tag{13}$$

where k_4 is indexing the Mel-bins. The further steps for the Mel approach are equal to that of $Y_3(k_3,\ell)$, i.e., applying Eq. (11) and Eq. (12), leading to $Y_4(k_4,\ell)$. All approaches lead to a spectral power density $Y_v(k_v,\ell)$ with K_v frequency bins per approach v. Thus, the frequency estimation is given by

$$\hat{f}_v(\ell) = \arg\max_{k_v}\{Y_v(k_v,\ell)\}\cdot\frac{f_s}{K_v} \tag{14}$$

with f_s indicating the sampling frequency. If the difference between the estimated frequency $\hat{f}_v(\ell)$ and the reference tire frequency $f(\ell)$ is within a tolerance range Δ, a foreign object is classified, i.e.,

$$b_v(\ell) = \begin{cases} 1 & \text{for } \left|\hat{f}_v(\ell) - f(\ell)\right| \le \Delta, \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

where $b_v(\ell) = 1$ denotes a positive foreign object detection and else no object detection.

Experimental Setup

For evaluation of the foreign object detection algorithms, recordings were done using microphones and accelerometers. The microphones (MICs) and accelerometers (ACCs) were installed at the wheel hub and the wheel guard of the rear right tire. To simulate foreign objects, a screw was attached to the profile of that tire. Recordings were done during driving. The sampling frequency was 96 kHz. The data were downsampled to 44.1 kHz for testing. For ground-truth of the tire frequency $f(\ell)$, the antilock braking system (ABS) signal was accessed and recorded as reference.

Three different metrics were applied for evaluation. The frequency estimations $\hat{f}_v(\ell)$ are compared to the reference frequency $f(\ell)$ by either the mean squared error (MSE)

$$\mathrm{MSE} = \frac{1}{\mathcal{L}}\sum_{\ell=0}^{\mathcal{L}-1}\left(f(\ell)-\hat{f}_v(\ell)\right)^2 \tag{16}$$

or the absolute value of the correlation coefficient

$$|r| = \left|\frac{\sum_{\ell=0}^{\mathcal{L}-1}\left(\hat{f}_v(\ell)-\mu_{\hat{f}_v}\right)\left(f(\ell)-\mu_f\right)}{\sqrt{\sum_{\ell=0}^{\mathcal{L}-1}\left(\hat{f}_v(\ell)-\mu_{\hat{f}_v}\right)^2\cdot\sum_{\ell=0}^{\mathcal{L}-1}\left(f(\ell)-\mu_f\right)^2}}\right|, \tag{17}$$

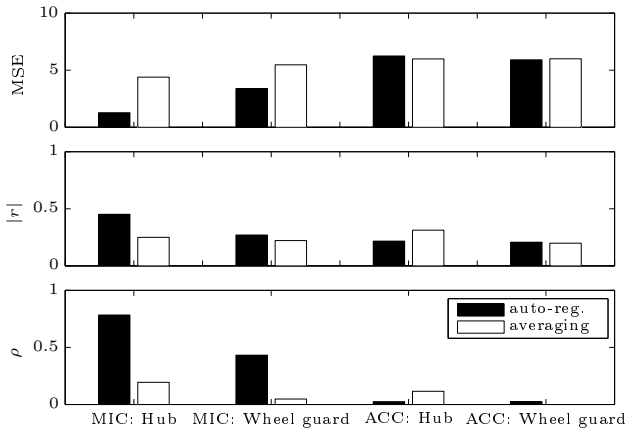

Figure 3: The mean squared error MSE and the correlation coefficient $|r|$ between frequency estimation $\hat{f}(\ell)$ and reference $f(\ell)$ and the positive rate ρ for four different sensor-position-combinations for the peak amplifying methods "auto regression" (black bar, cf. Eq. (4)) and, averaging (white bar, cf. Eq. (3)). The system is based on $Y_3(k_3, \ell)$ (mod. STFT) with N corresponding to 3 s.

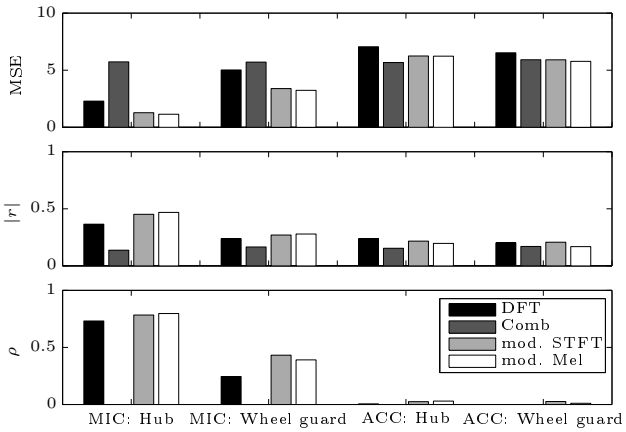

Figure 4: The mean squared error MSE and the correlation coefficient $|r|$ between frequency estimation $\hat{f}(\ell)$ and reference $f(\ell)$ and the positive rate ρ for four different sensor-position-combinations for the frequency analysis methods $Y_1(k_1, \ell)$ (DFT), $Y_2(k_2, \ell)$ (Comb), $Y_3(k_3, \ell)$ (mod. STFT) and $Y_4(k_4, \ell)$ (mod. Mel). The block size N corresponds to 3 s.

where $\mathcal{L}$ denotes the number of all frames used for evaluation and $\mu_{\hat{f}_v}$ and μ_f define the mean of $\hat{f}_v(\ell)$ and $f(\ell)$, respectively. The positive rate ρ is measured by

$$\rho = \frac{1}{\mathcal{L}} \sum_{\ell=0}^{\mathcal{L}-1} b_v(\ell). \tag{18}$$

Results

In the following subsections, the two approaches for the peak amplifying, the four frequency analysis methods and the block length N are investigated.

Peak Amplifying

To amplify peaks in a time signal $y'(n, \ell)$, two approaches have been proposed: one averages frames over a period of 0.03 s (cf. Eq. (3)) and the other uses an auto regression function with $N^{\text{reg}} = 16$ coefficients (cf. Eq. (4)). The Results using the $Y_3(k_3, \ell)$ approach from Eq. (12) are depicted in Figure 3. The results for the microphones are more accurate than for the accelerometers. The best results are achieved for the microphone at the hub. The auto regression approach yield lower MSE and higher $|r|$ and ρ than the averaging approach. Thus, in the following, the auto regression approach is used for peak amplifying.

Frequency Analysis

Four methods $Y_v(k_v, \ell)$ with $v = 1 \ldots 4$ to estimate the frequency $\hat{f}_v(\ell)$ from an acoustic signal are tested. The results are depicted in Figure 4. It can be seen that the comb filterbank approach $Y_2(k_2, \ell)$ fails in detecting objects. The best results are achieved for the modulation methods $Y_3(k_3, \ell)$ and $Y_4(k_4, \ell)$.

Block Size

The frequency estimation $\hat{f}_v(\ell)$ is processed in blocks ℓ. The block size N should be short on the one hand since the car velocity is not constant over a longer period. On the other hand, long windows are needed to have an adequate frequency resolution. If the resolution was low (considering a frequency range between 0 Hz and 20 Hz), i.e., few frequency bins k_v existed, the chance level for misclassification would rise. Furthermore, the smaller the window, the fewer turns of a tire fall into a window. For example, for a car velocity of 15 km/h, that is equivalent to $f \approx 2$ Hz, one turn of a tire is within a period of 0.5 s. Thus, maximally one ticking sound is within a block of 0.5 s duration that makes it quite difficult to get a correct frequency estimation for this window size. Hence, five block sizes between 1 s and 5 s are tested. In Figure 5, the mean squared error MSE, the correlation coefficient $|r|$ and the positive rate ρ of the sensors with continuously present foreign object are plotted for these block sizes. The frequency estimation was based on $Y_3(k_3, \ell)$ and auto regression peak amplifying. It can be seen that longer block sizes yield better results. However, there is a convergence in accuracy, i.e., nearly no gain is achieved for longer block sizes than 3 s.

Conclusion

We presented an algorithm to detect periodically repeating sounds from tires by acoustic sensors. Two kinds of sensors where used at different tire positions to record acoustic data. We showed that microphones work more accurate than accelerometers. Two approaches to amplify peaks from the repeating ticking sounds of objects were tested. The auto-regression method resulted in better performance than the averaging approach. To estimate the frequency, four approaches were evaluated. The

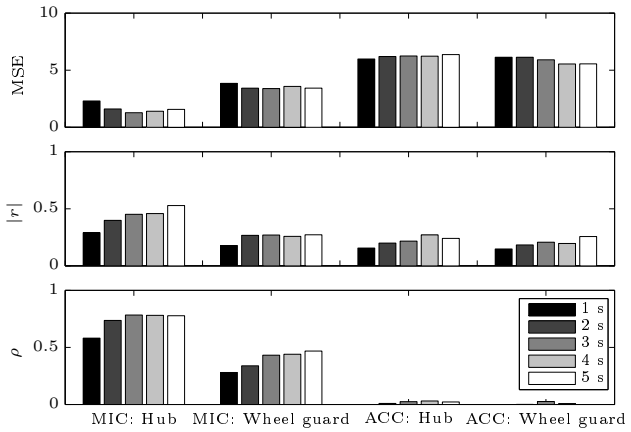

Figure 5: The mean squared error MSE and the correlation coefficient $|r|$ between frequency estimation $\hat{f}(\ell)$ and reference $f(\ell)$ and the positive rate ρ for four different sensor-position-combinations for the analysis method based on auto regression and $Y_3(k_3, \ell)$ (mod. STFT). The block size N varied corresponding to 1 s up to 5 s (see grey shades).

modulation spectrogram approach yielded highest performance. The block sizes were investigated and blocks with 3 s durations achieved highest accuracies.

References

[1] M. Holmberg, D. Gelbart, and W. Hemmert, "Automatic speech recognition with an adaptation model motivated by auditory processing," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 14, no. 1, pp. 43–49, 2006.

[2] J. Schröder, S. Goetze, V. Grützmacher, and J. Anemüller, "Automatic acoustic siren detection in traffic noise by part-based models," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 493 – 497.

[3] R. A. Lutfi and I. Heo, "Automated detection of alarm sounds," *Journal of the Acoustical Society of America*, vol. 132, no. 2, Sep. 2012.