International Journal of Automotive Technology, Vol. 8, No. 3, pp. 259–268 (2007)

1229–9138/2007/034–01

STATISTICAL ALGORITHMS FOR ENGINE KNOCK DETECTION

A. STOTSKY*

Volvo Car Corporation, Department 97545 HB1S, SE- 405 31 Gothenburg, Sweden

(Received 11 December 2006; Revised 15 March 2007)

ABSTRACT–A knock detection circuit that is based on the signal of an accelerometer installed on the engine block of a spark ignition automotive engine has a band-pass filter with a certain frequency as a parameter to be calibrated. A new statistical method for the determination of the frequency which is the most suitable for the knock detection in real-time applications is proposed. The method uses both the cylinder pressure and block vibration signals and is divided into two steps. In both steps, a new recursive trigonometric interpolation method that calculates the frequency contents of the signals is applied. The new trigonometric interpolation method developed in this paper improves the performance of the Discrete Fourier Transformation, allowing a flexible choice of the size of the moving window. In the first step, the frequency contents of the cylinder pressure signal are calculated. The knock is detected in the cylinder of the engine cycle for which at least one value of the maximal amplitudes calculated via the trigonometric interpolation method exceeds a threshold value indicating a considerable amount of oscillations in the pressure signal; this cycle is selected as a knocking cycle. In the second step, the frequency analysis is performed on the block vibration signal for the cycles selected in the previous step. The knock detectability, which is an individual cylinder attribute at a certain frequency, is verified via a statistical hypothesis test for testing the equality of two mean values, i.e. mean values of the amplitudes for knocking and non-knocking cycles. Signal-to-noise ratio is associated in this paper with the value of *t-statistic*. The frequency with the largest signal-to-noise ratio (the value of *t-statistic*) is chosen for implementation in the engine knock detection circuit.

KEY WORDS : Fourier analysis, Trigonometric interpolation, Filtering techniques, Engine knock, Six sigma, Spark ignition engines

1. INTRODUCTION

Poor knock detectability generates noise and can easily damage the engines. Knock miss-detection, i.e. the state where a knock is mistakenly detected, has a direct impact on engine fuel consumption and drivability performance. Since the engine knock represents one of the major constraints on the engine performance, a large number of papers have been written on the subject of a knock analysis using the cylinder pressure signal and the signal of the accelerometer that is installed on the engine block (knock sensor signal) (Brunt *et al.*, 1998; Burgdorf and Denbratt, 1997; Choi *et al.*, 2005; Sauler *et al.*, 2004; Lee *et al.*, 1998; Milo and Ferraro, 1998; Samimy *et al.*, 1995,1996; Urlab and Bohme, 2004). Very little attention, however, has been paid to the determination of the frequency which is the most suited for engine knock detection since the knock can be detected at a number of frequencies. The knock sensor that measures engine vibrations is combined with the signal processing unit where the band-pass filter is implemented. The knock frequency is an important calibration parameter that should be selected by the calibration engineer. The standard choice of the frequency that is made based on the cylinder geometry often gives poor knock detectability and, as a consequence, has an impact on vehicle drivability performance, noise and fuel consumption. Moreover, possible hardware and knock sensor placement changes during the project development often result in shifts in the frequencies for the engine knock detection. This necessitates the development of computationally efficient algorithms and MATLAB (MATLAB is a registered mark of the Mathworks, Inc of Natick, MA) software for the rapid determination of the knock frequency that is the most suitable for the real-time engine knock control applications.

The approach proposed in this paper introduces a new statistical decision making technology for knock frequency determination. The trigonometric interpolation algorithm applied in a window of a certain size w moving in time, which is used for the approximation of the signal by a trigonometric polynomial, is used as the main tool for detection of the knock frequency. Notice that the recursive DFT (Discrete Fourier Transformation) method

**Corresponding author.* e-mail: astotsky@volvocars.com

in a window of a certain size w moving in time can also be used for frequency analysis (Rizzoni *et al.*, 2005; Stotsky, 2005). However, the orthogonality condition for the trigonometric polynomials in certain intervals is the main restriction to the approximation performance and hence to the application of the DFT method. If the orthogonality condition is violated, then the implementation of the DFT method (in this case a better name is a trigonometric interpolation method) requires a matrix inversion, as is usual for least-squares fitting, and this in turn makes the method computationally expensive. Algorithms proposed in this paper allow trigonometric interpolation for any window size. This in turn allows improvement of the performance of the approximation. A recursive and computationally efficient version of the trigonometric interpolation method is developed.

The knock detection method uses both the cylinder pressure and block vibration signals and is divided into two steps. In the first step, the harmonic contents of the cylinder pressure signal, which has a considerable amount of oscillation during a knock event, is calculated by using the trigonometric interpolation method. Knock is detected in a certain cylinder if one of the maximal amplitudes (the maximum is taken over the knocking window) exceeds a threshold value detecting oscillations in the pressure signal. Notice that the knock does not occur in every cycle during engine steady-state operation even with significantly advanced ignition timing. The output of the first step is the number of engine cycles where the knock is judged to occur.

At the second step, the frequency analysis with a trigonometric interpolation method is performed on the knock sensor signal for the engine cycles selected in the previous step. Mean values of the maximal amplitudes and standard deviations are computed at all the frequencies over engine cycles selected in the previous step. Amplitudes are normally distributed at all the frequencies. In order to distinguish the knock induced vibrations from the vibrations induced by other events the current step should be repeated for the knock free cases. The mean values of the maximal amplitudes and standard deviations are also computed at all the frequencies for non-knocking cycles representing a background noise. A number of events which are not related to the engine knock event such as valve closing events, piston slaps and other mechanical vibrations detected by the block vibration sensor represent a background noise. Knock detectability is associated with the hypothesis test. The *Two Sample T-test* (the name is carried over from the Black Belt Memory Jogger, 2002), which compares two mean values of the maximal amplitudes at all the frequencies for knocking and non-knocking cycles, is performed. The hypothesis in which the two mean values of the maximal amplitudes are equal signifying that the knock is not distinguishable from the background noise is taken as the null hypothesis. The alternative hypothesis is that the mean value of the maximal amplitude in the knock case is larger than in the knock-free case at a certain frequency indicates knock in the corresponding cylinder. The knock is said to be detectable if the alternative hypothesis is valid. The knock is said to be not detectable if the null hypothesis is accepted. Notice that the statistical hypothesis can be verified with a certain risk. The probability of rejecting a null hypothesis when it is true is defined as a level of significance or α risk. The significance level should be chosen relatively small in order to reduce the probability of rejecting the null hypothesis mistakenly. Therefore knock detectability is also defined with a certain α risk.

Usually the *Two Sample T-test* is performed under the assumption that the variances of the two compared variables are equal (see Black Belt Memory Jogger, 2002). The variance of the maximal amplitude for knocking cycles is often significantly larger than the variance of the maximal amplitude for non-knocking cycles. Therefore, the *Two Sample T-test* is modified in this paper for the case of unequal variances. This modification results in the introduction of the correction factor, which represents the ratio of variances. The *Two Sample T-test* for the case where variances of compared variables are unequal is described in Appendix A.

Since the maximal amplitudes are normally distributed, the test statistic is the *t-statistic* which follows a Student distribution. The test statistic is defined as the difference between the mean values of the maximal amplitudes at a certain frequency in the case of knock and in the knock-free case (background noise) divided by the estimated pooled standard deviation multiplied by the number of cycles. The larger the difference between the mean values of the maximal amplitudes and the number of cycles and the smaller the pooled standard deviation, the better the knock detectability at a certain frequency. In other words, the value of the *t-statistic* is introduced as a quantitative measure of the knock detectability at a certain frequency, representing a signal-to-noise ratio. The frequency that is the most suitable for knock detection has the largest value of the statistic. Notice, that the value of the statistic depends also on the sample size, i.e., on the number of cycles, which is different for different cylinders where knock was detected using the cylinder pressure signal, in the first step of the algorithm. A large sample size (number of knocking cycles) gives a large value of the statistic, allowing reliable knock detection in a certain cylinder. On the contrary, a small number of the knocking cycles gives a small value of the statistic for the same difference between mean values of the maximal amplitudes and pooled standard deviation. If a single frequency should be chosen for a certain number of cylinders then the values of the statistics are added together, and the frequency with the largest sum of the statistics is chosen for knock detection.

A Volvo six cylinder prototype engine equipped with the cylinder pressure and block vibration sensors was used in the experiments. Algorithms are implemented in MATLAB and applied to the measured data collected from the engine.

The contribution of the present paper is a new statistical knock detection method which allows the optimal choice of the frequency that has the largest signal-to-noise ratio, i.e. the frequency which is the most suitable for implementation in the filter of the engine knock detection circuit.

2. RECURSIVE TRIGONOMETRIC INTERPOLATION ALGORITHMS

The recursive trigonometric interpolation method which allows the calculation of the frequency contents of an oscillating signal is described in this Section. The method is applied to the calculation of the frequency contents of the cylinder pressure and block vibration signals in Section 3.

2.1. Problem Statement

Suppose that there is a set of measurements of the oscillating signal s_k, $k = 1, 2 ...$, measured at the following points $x_k = k\Delta$, where Δ is a step size and k is the step number. Two signals are considered in the present paper: the cylinder pressure signal and the engine block vibration signal. Suppose that the signal can be exactly approximated by the trigonometric polynomial as follows:

$$\hat{s}_k = \varphi_k^T \theta_k , \tag{1}$$

$$\theta_k^T = [a_{0k} a_{q1k} b_{q1k} a_{q2k} b_{q2k} ,...a_{qnk} b_{qnk}] \tag{2}$$

$$\varphi_k^T = [1 \cos(q_1 x_k) \sin(q_1 x_k) \cos(q_2 x_k) \sin(q_2 x_k) ... \cos(q_n x_k) \sin(q_n x_k)] \tag{3}$$

where θ_k is the vector of the adjustable parameters and φ_k is the regressor, $q=q_1, ...q_n$ are the frequencies, a_{0k}, a_{qk} and b_{qk} are the coefficients that should be found, and $\hat{s}_k$ is an estimate of the signal s_k. The amplitude that plays an important role for the knock detection is defined at a certain frequency as follows:

$$A_{qik} = \sqrt{a_{qik}^2 + b_{qik}^2} \tag{4}$$

where $i=1, ...n$ and n is the number of the frequencies.

Assume that the measured signal can be presented as follows:

$$s_k = \varphi_k^T \theta_* , \tag{5}$$

where θ_* is the vector of true parameters,

$$\theta_*^T = [a_{0*} a_{q1*} b_{q1*} a_{q2*} b_{q2*} ... a_{qn*} b_{qn*}] \tag{6}$$

and a_{0*}, a_{q*} and b_{qn*} are constant unknown coefficients.

The signal s_k is approximated by (1) in the least squares sense in a moving window of a size w. The error to be minimized in every step is:

$$E_k = \sum_{i=k-(w-1)}^{i=k} (s_i - \hat{s}_i)^2, \quad k \geq w \tag{7}$$

2.2. Recursive Algorithms for Trigonometric Interpolation

The vector of adjustable parameters θ_k at step k that minimizes the performance index (7) can be computed as follows:

$$\theta_k = [\sum_{i=k-(w-1)}^{i=k} \varphi_i \varphi_i^T]^{-1} \sum_{i=k-(w-1)}^{i=k} \varphi_i s_i . \tag{8}$$

Notice that, if the orthogonality condition for trigonometric polynomials is satisfied for a certain window size the matrix $\sum_{i=k-(w-1)}^{i=k} \varphi_i \varphi_i^T$ becomes a diagonal matrix. This matrix is easily invertible as the diagonal matrix and the vector of the adjustable parameters θ_k is the vector of the Fourier coefficients. Therefore, the DFT method can be seen as a special case of the trigonometric interpolation method described below.

The vector of the adjustable parameters at step $(k-1)$ of the moving window can be computed as:

$$\theta_{k-1} = [\sum_{i=k-w}^{i=k-1} \varphi_i \varphi_i^T]^{-1} \sum_{i=k-w}^{i=k-1} \varphi_i s_i \tag{9}$$

In step k of the moving window, new data s_k and φ_k enters the window while s_{k-w} and φ_{k-w} leaves the window.

The vector of the adjustable parameters θ_k can be presented as:

$$\theta_k = [(\sum_{i=k-w}^{i=k-1} \varphi_i \varphi_i^T) - \varphi_{k-w} \varphi_{k-w}^T + \varphi_k \varphi_k^T]^{-1} [(\sum_{i=k-w}^{i=k-1} \varphi_i s_i) + \varphi_k s_k - \varphi_{k-w} s_{k-w}] \tag{10}$$

Application of the matrix inversion relation to (10) shows that the vector of the adjustable parameters in step k can be computed via the vector of the adjustable parameters in step $(k-1)$ as follows:

$$\theta_k = (I - \frac{A_{k-1} \varphi_k \varphi_k^T}{1 + \varphi_k^T A_{k-1} \varphi_k}) (\theta_{k-1} + \frac{\Gamma_{k-1} \varphi_{k-w} \varphi_{k-w}^T \theta_{k-1}}{1 - \varphi_{k-w}^T \Gamma_{k-1} \varphi_{k-w}}) + \Gamma_k (\varphi_k s_k - \varphi_{k-w} s_{k-w}) \tag{11}$$

where

$$\Gamma_k = A_{k-1} (I - \frac{\varphi_k \varphi_k^T A_{k-1}}{1 + \varphi_k^T A_{k-1} \varphi_k}) \tag{12}$$

$$A_{k-1} = \Gamma_{k-1}(I + \frac{\varphi_{k-w}\varphi_{k-w}^T\Gamma_{k-1}}{1-\varphi_{k-w}^T\Gamma_{k-1}\varphi_{k-w}}) \tag{13}$$

and I is the identity matrix, Γ_{k-1} is a recursive estimate of $[\sum_{i=k-w}^{i=k-1}\varphi_i\varphi_i^T]^{-1}$, and Γ_k is a recursive estimate of $[\sum_{i=k-(w-1)}^{i=k}\varphi_i\varphi_i^T]^{-1}$.

The elements of the regressor φ_i, $i \geq 3$ are recursively calculated via the following Chebyshev's three term recurrence relations:

$$\varphi_i = 2d_q * \varphi_{i-1} - \varphi_{i-2} \tag{14}$$

$$\varphi_i^T = [1 \cos(q_1 i\Delta)\sin(q_1 i\Delta)\cos(q_2 i\Delta) \\ \sin(q_2 i\Delta),..., \cos(q_n i\Delta)\sin(q_n i\Delta)] \tag{15}$$

$$d_q^T = [1 \cos(q_1\Delta)\cos(q_1\Delta)\cos(q_2\Delta) \\ \cos(q_2\Delta),...,\cos(q_n\Delta)\cos(q_n\Delta)] \tag{16}$$

where Δ is the sampling step, and '*' denotes element-wise vector multiplication, and index i is equal to k, $(i=k)$ and to $k-w$, $(i=k-w)$, $k \geq (w+3)$ for the recursive computations of φ_k and φ_{k-w}, respectively, in (11)–(13). For the recursive computations of the regressor over the whole window that are required in the first step of the moving window, $k=w$ and for calculations of the approximation error (7) the index i is defined as $i=k-w+p$, where $p=3$, $...w$.

The recursive algorithms described in this Section are applied to the calculation of the frequency contents of the cylinder pressure and block vibration signal in Section 3.

3. KNOCK DETECTION ALGORITHMS

The knock detection method proposed in this paper uses both the cylinder pressure and block vibration signals and is divided into two steps corresponding to the two Subsections of this Section.

3.1. Step 1: Knock Detection Using the Cylinder Pressure Signal

In the first step, the oscillations that occur during the knock event in the cylinder pressure signal are analyzed by means of the trigonometric interpolation method described in Section 2. Engine knock was generated by advancing the spark timing from the nominal spark timing. The knock is detected when at least one absolute value of the maximal amplitudes computed via (4) exceeds the threshold value that represents the level of background noise in the pressure signal. The detection method using the cylinder pressure signal is, in turn, divided into three steps.

3.1.1. Elimination of the low frequency trend

In this step, the cylinder pressure signal is approximated

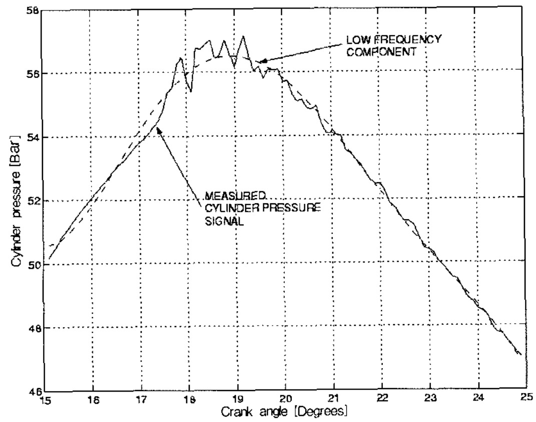

Figure 1. Cylinder pressure signal and the low-frequency component of the cylinder pressure signal are plotted as functions of the crank angle in a selected crank-angle window during a knocking cycle. The engine is operating at 1000 [rpm] and at full load. The cylinder pressure signal is plotted with a solid line, and the low frequency component is plotted with a dashed line. The cylinder pressure is measured with the step of 0.1^0.

via a low-order polynomial with the purpose of elimination of the low frequency component of the cylinder pressure signal. This emphasizes the knock induced signal, which is the input to the next step of the algorithm. Figure 1 shows the cylinder pressure signal and low-order polynomial approximation of the pressure signal. The difference between the cylinder pressure signal and its low frequency component that corresponds to the oscillations due to engine knock is the input to the next step of the algorithm.

3.1.2. Calculation of the frequency contents of the cylinder pressure signal

In this step, the difference between the cylinder pressure signal and its low frequency component is approximated by the trigonometric polynomial (1) aiming to calculate the frequency contents of the signal. The minimal number of terms (frequencies) should be used for approximation. Initial frequencies can preliminarily be estimated from the cylinder geometry. The optimal choice of the number of terms (frequencies) in the approximating polynomial (1) could be done as follows. Assume that the cylinder pressure signal can be exactly approximated by the polynomial (1) and the measured signal has a random measurement noise only. It is assumed that the measurement errors are independent and normally distributed. In this case the errors in the estimated parameters are also normally distributed, and the following ratio $R=E_k/(w-2n-1)$ is an estimate of the variance of the measurement noise, where n is the number of frequencies. The new frequency is included in the model (1), if the ratio R is

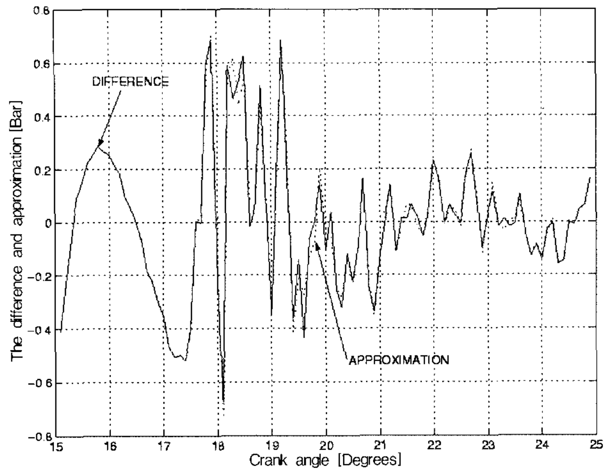

Figure 2. Difference between the cylinder pressure signal and the low-frequency component of the cylinder pressure signal is plotted with a solid line as a function of the crank angle in the selected crank-angle window during a knocking cycle. The approximation calculated via the trigonometric interpolation method is plotted with a dotted line. The engine is operating at 1000 [rpm] and at full load.

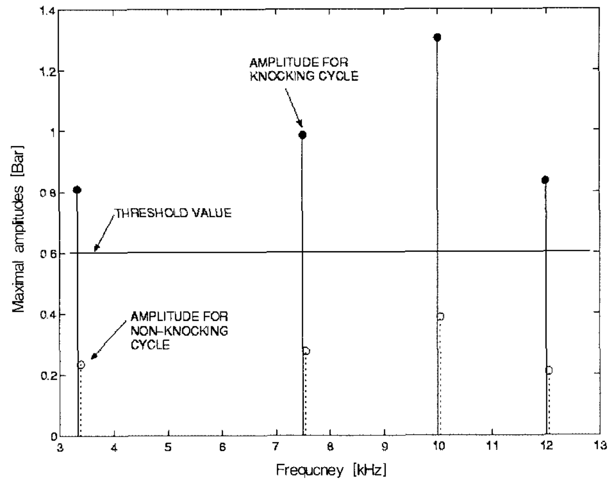

Figure 3. Maximal amplitudes for knocking and non-knocking cycles as functions of the frequency. Maximal amplitudes for knocking cycles are plotted with a solid line and maximal amplitudes for non-knocking cycles are plotted with a dotted line. The threshold value is 0.6 [Bar]. The engine is operating at 1000 [rpm] and at full load.

reduced and this reduction is statistically significant. The choice of the frequencies is optimal if the values of the ratio R and the variance of the measurement noise are approximately the same. A similar algorithm for the automatic term selection for ordinary polynomials is described by Stotsky (2006).

Figure 2 shows the difference between the cylinder pressure signal and the low-frequency component of the cylinder pressure signal and the approximation of the difference computed by the trigonometric interpolation method with four frequencies. Figure 2 shows that the difference can be accurately approximated by a trigonometric polynomial (1) with four frequencies.

3.1.3. Selection of the knocking cycles

The knock does not occur in every cycle of the engine steady-state operation even with a significantly advanced spark timing. A sufficiently large number of the knocking cycles should be selected for further processing of the signal from the knock sensor. The knock event is a non-stationary transient process and the amplitude at the frequencies of interest are time (crank angle) dependent. Maximal amplitudes, where the maximum is taken over the knocking window are used in this paper as a measure of the knock intensity (energy). A knock is detected if one of the maximal amplitudes of the pressure signal exceeds the threshold value. Figure 3 shows maximal amplitudes for knocking and non-knocking cycles as functions of the selected frequencies. If at least one of the maximal amplitudes in a certain cylinder for a certain cycle exceeds the threshold value that is selected by the calibration engineer, this cycle is considered as a knocking cycle and is selected for further knock sensor signal processing. Notice that slight oscillations are present in the pressure signal even during normal combustion. The intensity of these oscillations is significantly less than the intensity of the oscillations induced by the knock event. Therefore the intensity of pressure oscillations can be used as an indicator of the true knock intensity.

The threshold value - that is used for the detection of the knocking cycles could be selected as follows. First, the frequency contents of the cylinder pressure signal are calculated for a sufficiently large number of non-knocking cycles using the method described above (see steps 1 and 2). The mean values of the maximal amplitudes and standard deviations are calculated at all frequencies. The confidence interval for the mean value of the maximal amplitude in a certain cylinder utilizes a Student distribution and can be calculated by using the following formula:

$$\bar{A}_{q\max p} - t(\alpha, n_b)\frac{s_{qp}}{\sqrt{n_b}} \leq A_{q\max p} \leq \bar{A}_{q\max p} + t(\alpha, n_b)\frac{s_{qp}}{\sqrt{n_b}} \tag{17}$$

where $A_{q\max p}$ is the true mean value of the maximal amplitudes of the pressure signal at the frequency q, $\bar{A}_{q\max p}$ is the mean value of the maximal amplitudes calculated from the sample size n_b, s_{qp} is the sample standard deviation and $t(\alpha, n_b)$ is the value in the Student distribution look-up table for a certain α-risk and degrees of freedom (n_b-1). Notice that a larger number of non-knocking cycles results in tighter confidence intervals as expected from the Central Limit Theorem. The upper

$\bar{A}_{q\max p}+t(\alpha,n_b)\frac{s_{qp}}{\sqrt{n_b}}$ bound can be used as a threshold value for the background noise of the pressure signal for a particular cylinder.

3.2. Step2: Knock Detection by Using the Knock Sensor Signal

In this step, the frequency analysis is performed by the trigonometric interpolation method on the knock sensor signal for engine cycles selected in the previous step. Simultaneously, frequency analysis is performed for the non-knocking cycles in order to calculate the amplitudes representing background noise, i.e., vibrations induced by events which are not related to the engine knock. Figure 4 shows the distributions of the maximal amplitudes for knocking and non-knocking engine cycles at a frequency of 10 [kHz].

The distributions of the maximal amplitudes are close to normal distributions for sufficiently large numbers of engine knocking and non-knocking cycles.

Note that if the distribution of the maximal amplitudes is non-normal (see, for example Naber *et al.*, 2006) the data should be converted from a non-normal distribution to a normal distribution using a transformation function (look-up table). Standard transformation functions are presented in the Black Belt Memory Jogger (2002).

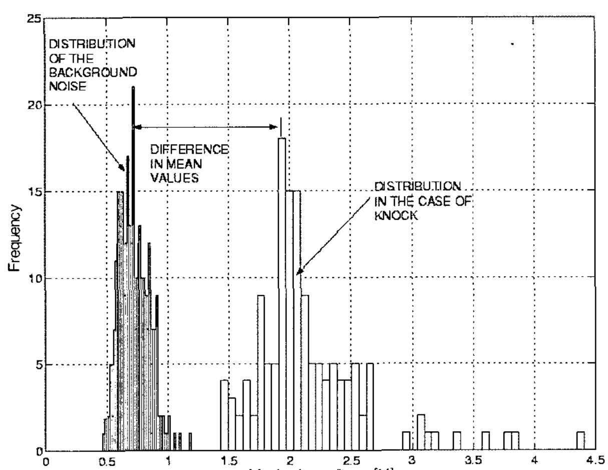

Figure 4. Distributions of the maximal amplitudes for knocking and non-knocking engine cycles at a frequency of 10 [kHz]. Maximal amplitudes are expressed in Voltage [V], due to the Voltage output of the block vibration sensor. The engine is operating at 1000 [rpm] and at full load.

3.2.1. Knock detectability

Knock detectability as a statistical attribute introduced in this paper is associated with the hypothesis test that compares the two mean values of the maximal amplitudes at a certain frequency. The hypothesis that the knock in one of the engine cylinders is not distinguishable from the background noise at a certain frequency is taken as the null hypothesis. The null hypothesis is defined as follows $H_{0qi}{:}\bar{A}_{q\ \max\ ki}=\bar{A}_{q\ \max\ bi}$, where $\bar{A}_{q\ \max\ ki}$ is a maximal amplitude at a frequency q averaged over a certain number of knocking cycles n_{ki}, $\bar{A}_{q\ \max\ bi}$ is a maximal amplitude at the same frequency averaged over a certain number of non-knocking cycles n_{bi}, and i is the cylinder number. Alternative hypothesis that $H_{Aqi}{:}\bar{A}_{q\ \max\ ki}>\bar{A}_{q\ \max\ bi}$ indicates that the knock is recognizable from the background noise. Therefore knock detectability is defined as follows.

Definition: *Engine knock is detectable at a certain frequency q, in the cylinder i with a certain significance level* α_{qi} *if the alternative hypothesis* $H_{Aqi}{:}\bar{A}_{q\ \max\ ki}>\bar{A}_{q\ \max\ bi}$ *is valid.*

The null hypothesis with a certain significance level is tested using the *Two Sample T-test.* Since the variance of a maximal amplitude for knocking cycles is often significantly larger than the variance of the maximal amplitude for non-knocking cycles, the *Two Sample T-test* is modified in this paper for the case of unequal variances (see Appendix A). The test statistic is the *t-statistic* which is calculated as:

$$t_{qi}=\frac{\bar{A}_{q\max ki}-\bar{A}_{q\max bi}}{s_{qi}}\sqrt{\frac{n_{ki}n_{bi}}{(n_{ki}+F_{qi}n_{bi})}} \tag{18}$$

where s_{qi} is the estimated pooled standard deviation which is defined as:

$$s_{qi}=\sqrt{\frac{(n_{ki}-1)s_{qki}^2+F_{qi}(n_{bi}-1)s_{qbi}^2}{F_{qi}(n_{ki}+n_{bi}-2)}} \tag{19}$$

where $s_{qki}=\sqrt{\frac{1}{(n_{ki}-1)}\sum_{j=1}^{j=n_{ki}}(A_{q\max ki}-\bar{A}_{q\max ki})^2}$,

$s_{qbi}=\sqrt{\frac{1}{(n_{bi}-1)}\sum_{j=1}^{j=n_{bi}}(A_{q\max bi}-\bar{A}_{q\max bi})^2}$ are standard deviations for knocking cycles and background noise respectively, $F_{qi}=s_{qki}^2/s_{qbi}^2$ for the case where the difference between variances s_{qki}^2 and s_{qbi}^2 is statistically significant and $F_{qi}=1$ otherwise. The significance of the difference in variances is tested via the *Test for Equal Variances* (see Black Belt Memory Jogger, 2002), where the hypothesis that two variances are equal is taken as a null hypothesis against the alternative hypothesis that the variance for the knocking cycles s^2_{qki} is larger than the variance of the background noise s^2_{qbi}, i.e., $s^2_{qki} > s^2_{qbi}$. The null hypothesis is rejected if the ratio F_{qi} is larger than a value that is taken from the F-distribution look-up table for degrees of freedom $(n_{ki}-1)$ and $(n_{bi}-1)$, respectively, with a certain significance level, which is chosen beforehand. The alternative hypothesis is accepted in this case, and the ratio $F_{qi}=s^2_{qki}/s^2_{qbi}$ is used in (18). This ratio is equal to one, i.e. $F_{qi}=1$, if the null hypothesis is accepted.

The value of t_{qi}-*statistic* is compared to the value in the Student distribution look-up table for $f_{qi}=n_{ki}+n_{bi}-2$ degrees of freedom and a certain significance level α_{qi}. If the value of the statistic t_{qi} is larger than the value in the Student distribution look-up table, the null hypothesis H_{0qi} is rejected in favor of H_{Aqi}. This in turn, implies that the engine knock is detectable in the cylinder i at a certain frequency q. If the value of the statistic is less than the value in the Student distribution look-up table, the number of knocking cycles should be increased. Notice, that the number of knocking and non-knocking cycles should be approximately the same for the hypothesis testing. If increasing of the number of the knocking cycles further is impossible due to some constraints (computer memory constraints, for example), then the significance level α_{qi} could be increased, reducing the values in the Student distribution look-up table to guarantee that the value of the statistic is larger than the value in the Student distribution look-up table. Increasing the significance level increases the probability of rejecting the null hypothesis mistakenly and hence the probability of the erroneous knock detection.

3.2.2. Choice of the most suitable frequency for the knock detection

The statistical knock decision making mechanism described above is used for verification of knock detectability at a certain frequency, but does not allow the choice of the most suitable frequency for knock detection. The knock is often detectable at several frequencies and the most suitable frequency have the largest signal-to-noise ratio. The signal-to-noise ratio is evaluated via the comparison of the two distributions (see Figure 4) representing knocking cycles and background noise. The better the separation between the two distributions at a certain frequency, the larger the signal-to-noise ratio. The value of the statistic (18) is introduced in this paper as a quantitative measure of the separation between the two distributions of interest. The larger the difference between the mean values of the amplitudes, the larger the number of the knocking cycles, and the smaller the pooled standard deviation, the larger the value of the statistic at a certain frequency.

If a single frequency should be chosen for a number of cylinders, then the individual cylinder values of the statistics are added for the number of engine cylinders for which the knock sensor is assigned. The frequency with the largest value of the common statistic is chosen as the best frequency for knock detection in those cylinders for a given sensor. In other words the most suitable frequency q_b for knock detection in M cylinders at a certain engine speed and engine load has the largest value of the following performance index:

$$p_q = \max_q \sum_{i=1}^{M} t_{qi} \tag{20}$$

where t_{qi} is the value of the statistic for a cylinder $i=q,\ldots, M$, and M is the number of the cylinders assigned to the knock sensor at the frequency q, engine speed ω and engine load L. The performance index (20) allows one to choose the best frequency at a certain speed and load.

The values of the performance index $\sum_{i=1}^{M} t_{qi}$ can be added for a number of working points with certain weighting factors assigned for every engine speed and load working point, which allows prioritization of certain areas. The frequency that has a maximal value of the performance index is chosen for knock detection in that particular working area.

Remark 1. The ratio $\frac{(\bar{A}_{q\max ki} - \bar{A}_{q\max bi})}{s_{qi}}$, which represents the difference between the average values of the maximal amplitudes expressed in terms of the pooled standard deviation, can also be used as a separation measure between the two distributions. The difference between the average values of the maximal amplitudes is statistically significant if $(\bar{A}_{q\max ki} - \bar{A}_{q\max bi}) > 2s_{qi}$, and the frequency with the maximal ratio $\frac{(\bar{A}_{q\max ki} - \bar{A}_{q\max bi})}{s_{qi}}$ could be chosen for knock detection. In the case of the same number of knocking and non-knocking cycles in the cylinder i, $n_{ki}=n_{bi}$ and $F_{qi}=1$, the pooled standard deviation $s_{qi} = \sqrt{\frac{s^2_{qki} + s^2_{qbi}}{2}}$ does not depend on the number of cycles and hence the ratio $\frac{(\bar{A}_{q\max ki} - \bar{A}_{q\max bi})}{s_{qi}}$ does not depend on the number of cycles either. The number of knocking cycles selected by cylinder pressure signal in the first step of the method is different for different cylinders. The difference in the number of knocking cycles for different cylinders is not accounted for in this approach which equally treats cylinders with different numbers of knocking cycles which might imply erroneous detection of the frequency. The performance index (18) proposed in this paper takes into account the number of the knocking cycles, prioritizing the cylinders with larger numbers of the knocking cycles.

Remark 2. Another statistical engine knock detection method was proposed by Samimy *et al.* (1995; Samimy and Rizzoni, 1996) using the signal processing technique proposed by Boashash and O'Shea (1990) and Kumar and Carroll (1984). The optimal test statistic is calculated via the correlation function of the knock sensor signal. The calculation of the correlation function is performed via a series expansion and calculations of the eigenfunctions of a corresponding covariance matrix of a given data set. The calculation of the eigenfunctions is performed in each discrete step since the correlation function is time dependent. This makes the method computationally expensive. The errors in calculating the eigenfunctions if the covariance matrix is close to a singular matrix might imply errors in the detection statistic. Moreover, the frequency analysis introduced by Samimy *et al.* (1995); and Saminy and Rizzoni (1996) is based on the DFT method, where the orthogonality condition limits the approximation performance of the method. The knock detection method proposed in this paper is based on the trigonometric interpolation method, which is not restricted by the orthogonality condition, and the simple *Two Sample T-test* ,which is modified for the case of unequal variances. The method is not computationally expensive and allows rapid determination of the frequency which is the most suitable for implementation.

4. CONCLUSION

A new statistical method for determination of the most suitable frequency for engine knock detection is proposed. The cascaded method uses the cylinder pressure signal for detection of knock events and the block vibration signal for statistical determination of the most suitable frequency for knock detection in real-time applications. The method is implemented in MATLAB and allows rapid determination of the frequency after the hardware and knock sensor placement changes. This in turn results in significant savings in calibration time during project development.

ACKNOWLEDGEMENT–This work was done within the Volvo Six Sigma Programme. The author is grateful to Mrs. V. Prokhotskaya for preparation of the manuscript in Microsoft Word format and pointing out typing errors.

REFERENCES

Black Belt Memory Jogger (2002). *A Pocket Guide for Six Sigma Success*. D. Picard Editor. GOAL/QPC. USA.

Boashash, B. and O'Shea, P. (1990). A methodology for detection of some underwater acoustic signals using time-frequency analysis techniques. *IEEE Trans. Acoustic, Speech, and Signal Processing* **38, 11,** 1829–1841.

Brunt, M., Pond, C. and Biundo, J. (1998). Gasoline engine knock analysis using cylinder pressure data. *SAE Paper No.* 980896.

Burgdorf, K. and Denbratt, I. (1997). Comparison of cylinder pressure based knock detection methods. *SAE Paper No.* 972932.

Choi, H., Lim, J., Min, K. and Lee, D. (2005). Simulation of knock with different piston shapes in a heavy-duty LPG engine. *Int. J. Automotive Technology* **6, 2,** 133–139.

Kumar, B. and Carroll, C. (1984). Performance of wigner distribution function based detection methods. *Optical Engineering* **23, 6,** 732–736.

Lee, J.-H., Hwang, S.-H., Lim, J.-S., Jeon, D.-C. and Cho, Y.-S. (1998). A new knock-detection method using cylinder pressure, block vibration and sound pressure signals from SI engine. *SAE Paper No.* 981436.

Millo, F. and Ferraro, C. (1998). Knock in SI engines: A comparision between different techniques for detection and control. *SAE Paper No.* 982477.

Naber, J., Blough, J., Frankowski, D., Goble M. and Szpytman, J. (2006) Analysis of combustion knock metrics in spark ignition engines. *SAE Paper No.* 2006-01-0400.

Rizzoni, G., Guezennec, Y., Soliman, A. and Lee, B. (2005). *Engine Control Using Torque Estimation.* US Patent 6866024 B2.

Samimy, B., Rizzoni, G. and Leisenring, C. (1995). Improved knock detection by advanced signal processing. *SAE Paper No.* 950845.

Samimy, B. and Rizzoni, G. (1996). Mechanical signature analysis using time-frequency signal processing: Application to internal combustion engine knock detection. *Proc. IEEE* **84, 9,** 1330–1343.

Sauler, J. Franke S., Unland, S., Torno, O., Heinstein, A., Kluth, C., Haeming, W., and Baeuerle, M. (2004). *Knock Recognition in Internal Combustion Engines with Modifications by Changing Filter Characteristics or Cylinder Specific Changes.* US Patent Application Publication US/2004/0030486 A1.

Stotsky, A. (2005). Computationally efficient filtering algorithms for engine torque estimation. *Proc. IMechE,* ***219,*** *Part D; J. Automobile Engineering,* 1099–1107.

Stotsky, A. (2006). Data-driven algorithms for engine friction estimation. *Proc. IEEE Int. Conf. Control Applications*, Munich.

Urlab, M. and Bohme, J. (2004). Reconstruction of pressure signals on structure-borne sound for knock investigation. *SAE Paper No.* 2004-01-0521.

APPENDIX A

TESTING THE EQUALITY OF TWO MEANS FOR THE CASE OF UNEQUAL VARIANCES

Usually the *Two Sample T-test* ,which is used for testing the equality of two means, is performed under the assumption that the variances of the two compared variables are equal (see Black Belt Memory Jogger, 2002). The *Two Sample T-test* for the case of unequal variances where the relationship between variances is known is proposed below.

Consider two normally distributed variables x and y described by the mean values $\bar{x}$ and $\bar{y}$ and sample variances S_x^2 and S_y^2 calculated from a size n sample and a size m sample respectively. Consider the null hypothesis to be that the mean values of the two variables are equal, i.e. $H_0{:}\bar{x}=\bar{y}$. Suppose that the variances σ_x^2 and σ_y^2 are different and unknown, but that the relationship between the variances presented in the form of a ratio is known, i.e.,

$$F=\frac{\sigma_y^2}{\sigma_x^2} \tag{21}$$

where F is a known number. Denoting the variance of the variable x as σ^2, i.e., $\sigma^2=\sigma_x^2$ the variance of the variable y is calculated as $F\sigma^2$. Ratio (21) can be estimated for sufficiently large sample sizes by using the two pooled sample variances S_x^2 and S_y^2, i.e.,

$$F=\frac{S_y^2}{S_x^2} \tag{22}$$

where $S_x^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(n-1)}$, $S_y^2=\frac{\sum_{i=1}^{m}(y_i-\bar{y})^2}{(m-1)}$.

The variance of the difference $(\bar{x}-\bar{y})$ is calculated as follows:

$$V(\bar{x}-\bar{y})=V\bar{x}+V\bar{y} =\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}=\frac{(m+Fn)}{nm}\sigma^2 \tag{23}$$

The following variable follows a normal distribution i.e.,

$\frac{(\bar{x}-\bar{y})}{\sigma}\sqrt{\frac{nm}{m+Fn}}\in N(0,1)$, where $\frac{nm}{m+Fn}$ is the "effective sample size".

The next step is evaluation of the variance σ. To this end the following variables are considered:

$$\frac{(n-1)S_x^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_x^2}, \text{ and } \frac{1}{F}\frac{(m-1)S_y^2}{\sigma^2}=\frac{\sum_{i=1}^{m}(y_i-\bar{y})^2}{\sigma_y^2}$$

which follow a χ^2 distribution with $(n-1)$ and $(m-1)$ degrees of freedom, respectively. Hence the variable

$$\frac{F(n-1)S_x^2+(m-1)S_y^2}{F\sigma^2} \tag{24}$$

follows a χ^2 distribution with $(n+m-2)$ degrees of freedom. Taking into account that

$E[S_x^2]=\sigma^2$, and $E[S_y^2]=F\sigma^2$,

where E is the mathematical expectation, consider:

$$E\left[\frac{F(n-1)S_x^2+(m-1)S_y^2}{F(n+m-2)}\right]=\sigma^2. \tag{25}$$

The variable $S^2=\frac{F(n-1)S_x^2+(m-1)S_y^2}{F(n+m-2)}$ which is often called estimated pooled variance, is in fact an unbiased estimate of the variance σ^2. Consider the following statistic:

$$t=\frac{(\bar{x}-\bar{y})}{\sqrt{F\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{m}(y_i-\bar{y})^2}}\sqrt{\frac{Fnm(n+m-2)}{(m+Fn)}}=$$

$$\frac{(\bar{x}-\bar{y})}{S}\sqrt{\frac{nm}{(m+Fn)}}$$

$$=\frac{(\bar{x}-\bar{y})}{\sigma}\sqrt{\frac{nm(n+m-2)F\sigma^2}{(m+Fn)F(n+m-2)S^2}}$$

$$=\frac{(\bar{x}-\bar{y})}{\sigma}\sqrt{\frac{nm(n+m-2)F\sigma^2}{(m+Fn)(F(n-1)S_x^2+(m-1)S_y^2)}}$$

$$=\frac{(\bar{x}-\bar{y})}{\sigma}\sqrt{\frac{nm}{(m+Fn)}} \Big/ \sqrt{\frac{F(n-1)S_x^2+(m-1)S_y^2}{F\sigma^2}\frac{1}{(n+m-2)}}.$$

Since $\frac{(\bar{x}-\bar{y})}{\sigma}\sqrt{\frac{nm}{(m+Fn)}}\in N(0,1)$, and $\frac{F(n-1)S_x^2+(m-1)S_y^2}{F\sigma^2}\in\chi^2_{n+m-2}$, the statistic t follows a Student distribution with $(n+m-2)$ degrees of freedom according to the definition $\left(t\triangleq\frac{N(0,1)}{\sqrt{\frac{\chi^2_{n+m-2}}{n+m-2}}}\right)$.

If the value of the statistic t is larger than the value in the Student distribution look-up table, then the null hypothesis is rejected in favor of the alternative hypothesis. Notice that the statistic t can be used for hypothesis testing if the difference in variances σ_x and σ_y is statistically significant. Therefore the hypothesis about the equality of the two variances should first be tested using the *Test for Equal Variances* (see Black Belt Memory Jogger, 2002). If the difference in the variances is statistically significant, the statistic t with the ratio $F=S_y^2/S_x^2$ can be used for hypothesis testing. If the difference in variances is not statistically significant, the statistic t with $F=1$ which corresponds to the standard *Two Sample T-test* is used for the hypothesis testing.