Advanced Diagnostic System for Piston Slap Faults in IC Engines, Based on the Non-stationary Characteristics of the Vibration Signals

Jian Chen1, 2*, Robert Bond Randall2
1. Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, UK
2. School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney 2052 Australia

Telephone: +44-758-7456129
Email: jc2d14@soton.ac.uk;

Abstract

Artificial Neural Networks (ANNs) have the potential to solve the problem of automated diagnostics of piston slap faults, but the critical issue for the successful application of ANN is the training of the network by a large amount of data in various engine conditions (different speed/load conditions in normal and with different locations/levels of faults). On the other hand, the latest simulation technology provides a useful alternative in that the effect of clearance changes may readily be explored without recourse to cutting metal, creating enough training data for the ANNs. In this paper, based on some existing simplified models of piston slap, an advanced multi-body dynamic simulation software was used to simulate piston slap faults with different speeds/loads and clearance conditions. Meanwhile, the simulation models were validated and updated by a series of experiments. Three-stage network systems are proposed to diagnose piston faults: fault detection, fault localization and fault severity identification. Multi Layer Perceptron (MLP) networks were used in the detection stage and severity/prognosis stage and Probabilistic Neural Networks (PNN) was used to identify which cylinder has faults. Finally, it was demonstrated that the networks purely trained on simulated data can efficiently detect piston slap faults in real tests and identify the location and severity of the faults as well.

Keywords: Piston slap fault diagnosis, Internal Combustion engines, Artificial Neural Networks, simulation

1. Introduction

Many researchers have already investigated the principle of the motion and the impact between the piston and inner wall of Internal Combustion engines (IC engines). Zinchenko [1] and Meier [2] were pioneers in the study of the noise induced by piston slap. Ungar & Ross [3] studied the dynamics of lateral piston motions across the cylinder clearance spaces and developed the relationship between the side thrust force and the inertia force and combustion force. They found that the times of occurrence of piston slap in one cycle are dependent on the operating conditions. Haddad [4] and Haddad & Fortescue [5] appear to be the earliest researchers to use a computer to simulate the motion of piston/liner on the basis of a simplified mathematical model. On the other hand, many works have been contributed to the lubrication effects between the piston and inner wall into account in recent years [6-15]. They tried to solve the Reynolds equation and calculate the impact force. A disadvantage of solving the impact force through the Reynolds equation is that the approach needs a lot of computing resources [16]. In fact, when the piston hits the inner wall, the main effect of the lubricating oil on the impact force should be equivalent to a squeezed film damper. Therefore, when studying the impact force of piston slap, some researchers proposed to use squeeze film dampers to simplify the effects of the lubricating oil, rather than to solve Reynolds equation. Based on Haddad’s method [4], Cho et al. [17] further studied the modelling of piston slap. They modelled the piston and inner wall as lumped masses, spring and damper elements, and the parameters were estimated from point mobility experiments. They simplified the lubrication oil as a viscous damper in the model. After the impact forces were calculated, they predicted the engine block surface vibration response by convolving the impact forces with measured
impulse responses between the cylinder inner walls and the engine block outer surfaces. More recently, Zheng & Ohta [18] proposed a theoretical model of piston slap which considering the movement and coupled vibration of piston, connecting rod, crankshaft and engine block, so as to accurately estimate the piston slap induced vibration of IC engine. Dolatabadi et al. [19] developed a coupled simulation model of piston dynamics and engine tribology (tribodynamics) to study piston slap events.

In general, the aims of these existing studies about piston slap were focussed on the piston design (with normal piston-inner wall clearance), including the geometrical and lubrication aspects. However, IC engines have severe operating conditions, involving high temperatures and pressures, with large variations in internal forces within each cycle, and as such the wear mechanisms are complex [20]. The piston, with high speed motion and complicated kinematic characteristics, are particularly vulnerable to the harsh engine operating conditions. Under the conditions of an external shock load or contaminated lubricant, metal to metal contact between the piston and the inner wall can occur and will lead to excessive wear and quickly create oversized clearances. With oversize clearance, the impact forces between the piston and cylinder inner wall increase greatly, which further accelerates the wear. Therefore, piston slap faults are considered to be a common mechanical fault in engines. As mentioned in preceding paragraph, even though many works have contributed to piston design with respect to piston slap, there is still much to be done on diagnosis of piston slap faults using vibration signals (but it is worth pointing out that all these existing studies, especially the modelling of piston slap, set up the foundation for the vibration based simulation model of piston slap faults.).

Up to now, only a limited number of researchers have investigated the technology of using the vibration signals for the diagnostics of piston slap faults. Geng & Chen [21] studied the piston slap phenomenon inside the reciprocating engine from experiments, and then proposed a simulation model to predict the slap-induced vibration response and analysed the relationship between the piston slap faults and speed/load and clearance sizes. Li et al. [22] used the ranged angle empirical mode decomposition method to extract the vibrations induced by piston slap for oversized clearance.

On the other hand, when these vibration-based techniques are applied in a real situation, the piston slap faults cannot automatically be diagnosed from the analysed vibration signals. ANNs have the potential to solve the problem of automated diagnostics of piston slap faults. But the critical issue for the application of ANNs is having sufficient input data to train the networks and it is uneconomical to achieve this by performing large numbers of experiments with piston slap faults for all possible combinations of fault size and location, with large speed and load ranges. The latest simulation technology provides a useful alternative in that the effect of clearance changes may readily be explored without recourse to cutting metal, creating enough training data for the ANNs. In this paper, based on some existing simplified models of piston slap, advanced simulation software (LMS Virtual.Lab, now Siemens PLM Virtual.Lab) was used to simulate piston slap faults with different speeds/loads and clearance conditions. The simulation models were validated and updated by a series of experiments. As found from the experimental data, the useful information for the piston slap fault diagnosis lies in the processed envelope signals, so the simulations only have to produce the correct envelope signals, rather than the raw acceleration signals. The automated system consists of three stages: Multi Layer Perceptron (MLP) for the first stage (detection stage), Probabilistic Neural Networks (PNN) for the second stage (localization stage) and MLP for the third stage (severity identification stage). The selected amplitude features were used in the detection stage and severity identification stage and the selected phase features were used in the PNNs to localize which cylinder has a fault. Updated on our former works [23, 24], during the severity identification stage, separate MLP were designed for each speed condition (but with different loads) rather than using one MLP to cover all speed and load conditions; moreover a pure linear function was selected as the transfer function of the output layer of the MLPs for the fault severity identification stage, therefore the networks can linearly classify the fault levels and the output agrees better with the real situation.

2. **Simplified mechanisms of piston slap**

Piston slap is an inherent phenomenon in a reciprocating engine or compressor. As shown in Figure 1, when the resultant transverse force exerted by the connecting rod changes direction, the piston leaves the side of the inner wall it was previously held against, and moves to the other side of the inner wall. This motion has also been called the “secondary movement of the piston”. If the lateral acceleration of the piston (induced by the transverse force) is big enough and the interval until the next time the force direction changes is long enough, the piston will move across the clearance and collide against the opposite side of the inner wall. The
impacts will then excite vibration of the engine block and cause structure-borne noise around the engine. If we neglect the piston offset (between the centre of gravity of piston and the joint centre of piston and piston pin in $Y$ direction) and the friction between the piston and piston pin, we can find a simplified equation of piston motion and also solved the lateral side-thrust force.

Here, $F_g$ is combustion force, $F_{cr}$ is the reaction force of the connecting rod and its direction is in line with centre line of connecting rod, $F_{w}$ is the reaction force of the inner wall, $F_z$ is the axial inertia force of the piston and $F_y$ is the lateral inertial force of the piston, $c_p$ is the radial clearance, $\omega$ is the almost constant angular speed of the crankshaft, $b$ is the length of the connecting rod and $R$ is the crank radius. The reaction force of the connecting rod $F_{cr}$ can be further decomposed into two component forces: vertical force $F_{z11}$ and side-thrust force $F_{y11}$. Essentially, the piston slap is due to the periodic change of the transverse component of the reaction force (side-thrust force $F_{y11}$) by the connecting rod. The side-thrust force can be solved by equation below [21] (please note there was a mistake for eq.3 in reference [21], the denominator of eq.3 should be $\sqrt{1 - (\lambda sin \theta)^2}$ instead of $\sqrt{1 + (\lambda sin \theta)^2}$.)

\[
F_{y11} = \frac{(F_g - F_w) R \sin \theta}{\sqrt{1 - (\frac{R}{b} \sin \theta)^2}} \\
= \frac{[F_g - m_p R \omega^2 (\cos \theta + \frac{R}{b} \cos 2 \theta)] \frac{R}{b} \sin \theta}{\sqrt{1 - (\frac{R}{b} \sin \theta)^2}}
\]

$m_p$ is the mass of the piston (also includes the mass of piston pin and 1/3 mass of connection rod), $\theta$ is the crank angle. Whenever $F_{y11}$ passes through zero and changes direction it may cause an occurrence of piston slap at some position along the inner wall. Piston slap normally happens around the Top Dead Centres (TDCs) and Bottom Dead Centres (BDCs). This analysis can help us to understand the basic mechanism of the occurrence of piston slap, but on the other hand, some 3D multi-body simulation software developed in recent years, provides a capability to accurately simulate the motion of piston, and also solve more sophisticated kinematic and dynamic equations involved in piston slap.

3. Experiments and signal processing
3.1 Experiment set-up and data acquisition

The basis of using a simulation model to create sufficient data is that the simulation models have to be updated and evaluated by experimental results, therefore a series of vibration signals with piston slap faults and without piston slap faults (normal or with combustion faults) were recorded on a Toyota 3S-FE 4-cylinder gasoline engine at the University of New South Wales. The firing sequence of the engine is 1-3-4-2. The engine was connected to a hydraulic dynamometer which works as an external load. By controlling the dynamometer, three constant speed conditions were selected: 1500rpm, 2000rpm and 3000rpm. For each speed, there were three different load conditions: 50Nm, 80Nm, 110Nm. The normal piston clearance is 0.05mm. After the test with normal and misfire conditions, the cylinder 1 was separately bored with two oversized clearances (3 times and 6 times normal piston clearance) to simulate piston slap faults. The two oversized clearances were advised by the engineers in the engine reconditioning workshop: 3 times normal clearance can represent moderate piston slap faults and 6 times normal clearance can represent sever piston slap faults. The inner wall was born around the whole circumference. It is worth mentioning that geometry shape of piston and inner wall will indeed affect the lubrication oil distribution and the integration of impact force, but in this paper, we studied the vibration signal change (precisely, envelope signal change) to diagnose the oversized clearance between the piston and inner wall, rather than solving accurate impact force. Therefore, even though the inner wall was born around the whole circumference instead of the side-thrust impact sections, we can consider that the impact force concerned here is mainly dominated by the clearances in the plane of motion and the mass of the piston. The clearances in the crankshaft axial direction and the shape of piston & inner wall have no big influence on the impact force we concerned too.

Regarding the data acquisition system, the sampling frequency was 25600 samples/sec. An accelerometer was stuck on the block surface, and the location is close to the major thrust side of cylinder 1 (as shown in Figure 2). A proximity transducer was used to pick up a once-per-rev tacho signal which corresponds to the TDC of cylinder 1. Additionally, a pressure sensor, which is integrated with spark plug, was used to measure the cylinder chamber pressure. The synchronously averaged pressure signal was used to update the combustion chamber pressure of the simulation models in the simulation section of this paper. Meanwhile, because there are two tacho signals corresponding to the TDCs in the firing stroke and expansion stroke in each cycle, but only the tacho signal in the firing stroke is useful for the fault diagnostics (especially for identifying the localizations of faults), so the pressure signals were also used to remove the tacho signals in the expansion strokes from the measured tacho signals.

![Figure 2. Accelerometer location.](image)

Examples of raw vibration signals in normal condition and with different severity levels of piston slap faults are shown in Figure 3. The tacho signals are shown at the bottom of each figure and they correspond to the firing stroke TDC (tacho) of the cylinder 1. It can be seen that the main piston slap events occurred around the TDC and the piston slap events in the firing strokes are much stronger.
3.2 Cylinder pressure balance check

Because the piston clearance was machined to six times the normal clearance in the second stage piston slap fault experiments, it was necessary to check whether there was combustion leakage occurring in the cylinder 1. During the second stage piston experiment, the pressures in the cylinder 1, 2 and 4 were measured at different speeds/loads. The pressure signals were synchronously averaged for better comparison. Regarding the number of averages, if the signals were averaged over a certain period for two firing cycles, and there is no big difference between the two cycles of the averaged results, it means the number of averages is sufficient. Four firing cycles are shown in Figure 4 so as to indicate the sufficiency of average numbers.

It was found that pressure differences (peak values) between cylinder 1 and other cylinders were reasonably small (less than ±5%). Notably, the pressure in cylinder 1 is bigger than for the other cylinders at 2000rpm/80Nm. Therefore, the excessive piston clearance does not necessarily cause combustion pressure loss as some authors have previously found [25].
3.3 The “kurtogram” for band filter and envelope analysis

In order to find how the impulsiveness of the measured mechanical faults signals varies with frequency and remove some interfering noise, the “kurtogram” was applied as a prelude to the envelope processing. The “kurtogram”, proposed by Antoni and Randall [26], is a powerful way of using spectral kurtosis to detect and characterise transient signals for envelope analysis. The spectral kurtosis is calculated by taking the fourth power of the time/frequency envelope $H(t, f)$ of the signal at each time and frequency bin, and averaging its value by moving a window along the record, then normalising it by the square of the mean square value [26]. The subtraction of 2 in the Equation (2) is because the definition is in terms of cumulants, and is explained in [26].

$$K(f) = \frac{H^4(t, f)}{[H^2(t, f)]^2} - 2$$ (2)
In the equation, the denominator is independent of the window length chosen, but the numerator is affected by it. For any non-stationary signal, its spectral kurtosis is not only dependent on the frequency \( f \), but also the frequency resolution \( \Delta f \). The “kurtogram” can represent the spectral kurtosis in the \((f, \Delta f)\) plane and can find the best window for a maximum value of kurtosis, which indicates the highest level of impulsiveness of the filtered signal. Antoni also defined a “fast kurtogram” in [27], and this has been used in this work. The “fast kurtogram” is a based on a tree-like multirate filter-bank structure (1/3-binary tree structure). As shown in Figure 5, for different decomposition levels \( k \), the frequency band from 0Hz to Nyquist frequency (12800Hz in this paper) will be divided into multiple sequences with central frequency \( f_i \) and with bandwidth (frequency resolution) \( \Delta f_k \). It is worth pointing out that the value of kurtosis for one impulse in a cycle is larger than that for several impulses in a cycle, but that does not mean that the former case has more useful information than the latter for the detection of engine faults. So the frequency band with the highest level of kurtosis is not necessarily the best selection to find the resonance bands most excited by the faults, but the results from the kurtogram are useful as a guide to the most excited frequency bands for the following feature extraction.

![Figure 5. Fast kurtogram paving of the (frequency/frequency resolution) plane [27]](image)

When the kurtogram was applied to the acceleration signals at different speeds/loads, it was found that the frequency bands with highest impulsiveness vary in a small range (both central frequency and bandwidth), and mostly lie below the half the Nyquist frequency (ie, 6400Hz). Four examples of the results from the kurtogram for the piston slaps are shown in Figure 6, where the brown area indicates the most impulsive frequency range. In order to have a fixed basis for comparison, the frequency band selections were adjusted and a common frequency band was found. The most important premise of the adjustment was that there is no significant feature change or loss in the subsequent Fourier analysis of the envelope. A common frequency range of (400-3200Hz) was selected for the piston slap fault detection and this range is agreed with the finding (around 2000Hz) by reference [28]. It is worth mentioning that engine knock, as a common combustion fault in engine, also excites high frequency resonances, but earlier research [29] has demonstrated that engine knock excites high frequency resonances with a typical range of 4-10 kHz. The vibration caused by other mechanical faults, such as journal bearing knock faults, may also affect the frequency selection, but firstly the location of accelerometer is very close to major thrust side of piston (as shown in Figure 2), so the contamination is small; secondly our former work [30] also shown that the calculated frequency range (400–4400 Hz) with bearing knock faults is not total overlap with the selected frequency range with piston slap faults.
As mentioned before, owing to their characteristics of second order cyclostationarity, the diagnostic information in the signals for piston slap fault should reside in the repetition frequency and pattern of the bursts. The envelope signals should be deterministic (actually first order cyclostationary) and therefore can be synchronously averaged. The envelope signal was also squared to represent it in power units (proportional to variance). An example of the raw signals and filtered squared envelopes for piston slap fault are shown in Figure 7. Two firing cycles are presented in Figure 7. Compared to the normal condition, it is easy to see that the squared envelope for piston slap fault increased a lot in magnitude. Even though the amplitude of the raw signal in the faulty condition is approximately two times larger than that in the normal condition, after filtration and squaring the envelope, the combustion noises and other mechanical noises were removed and the piston slap fault information got much clearer, therefore it implies that it should be quite easy to use the information from the squared envelopes to detect the piston slap faults. At same time, it can be seen that the most obvious piston slap happens just after TDC of the firing stroke. This is because in the firing stroke the horizontal reaction force from the con-rod was increased by the combustion, so the impact between piston and inner wall became stronger. Meanwhile there is up to 20 degrees lag of major piston slap from TDC and the degrees depend on the ignition time at different speeds and loads. Note that the chamber pressures reach the maximal values after TDC and there is an interval for the piston movement from one side of inner wall to another side.
From Figure 8, it can be seen that the peak values of the squared envelope for piston slap fault increase with the load; actually for each speed condition, the increase of the peak value has approximately linear relationship with the increase of load. From Figure 8, it can be found that the magnitude of the squared envelope also increases with the speed, but without linear relationship between them. Comparing the squared envelopes at the same speed/load but different clearance stages (for instance, the two stages of piston slap fault at 1500rpm/110Nm in Figures 8 and 9), it can be found that the peak value has obviously increased (about 2 times increase for 1500rpm/110Nm), but the increase ratio has big variation and is strongly related to the speed and load; therefore, in order to accurately identify the severity level of piston slap faults, it is necessary to consider the influence of speed and load.
Figure 8. Squared envelopes for piston slap faults at same speed but different load
3.4 Feature extraction and selection

The most important thing for the diagnosis of piston slap faults is to find parameters to interpret the process squared envelope signals. With such parameters, the diagnosis system can automatically monitor the operation of any engine by pattern recognition technology. Therefore it is necessary to further analyse the envelopes in normal and with different piston slap faults in the frequency domain. Inspired by the pattern recognition research of Desbazeille et al. [31], polar diagrams were employed to interpret the harmonics of the waveform patterns from the processed vibration waveforms. The Fourier coefficients were studied and presented in polar diagrams. Due to the transient characteristics, the amplitude and phase of the first forty harmonics of the envelope signal were considered.

However, the candidate features have different sensitivities to the different conditions of the engine and only some of them may be useful for the pattern recognition of the mechanical faults. Moreover, it requires a large amount of computing resources if all features are to be input into the ANNs. Thus it is important to select the optimal features to input into the ANNs, especially for the amplitude feature selection. A GA algorithm [32] is applied to find the best amplitude features for the diagnostics of piston slap faults. The detail of GA’s selection on the amplitude feature can be obtained in reference [33]. The best ten amplitude features was selected by the GA as the input to the MLPs. For the piston slap faults, the selected amplitudes of the harmonic numbers are: 14, 21, 23, 19, 1, 13, 25, 27, 6 and 16 (the sequence is based on the ranks of their fitness values in GA).

Because the Fourier transform interprets one cycle (two rotations) as 360º, the phase change of the 1st harmonic can be directly recognized in a 360 degree polar plot. When the faults happen in a certain cylinder for different speed/load conditions, if the phases of the nth harmonic of all cases (all speed and load conditions) are gathered together or near fixed, it implies that this phase is a potential feature for identifying the localization of faults. When the phases of the first twenty harmonics were considered firstly, it was found that the phases of higher harmonics are more scattered. That is because the phase difference of the nth harmonic of any two cases is equal to n times of the phase difference of the 1st harmonic. So the conclusion is that the best phase features should be selected from the first ten harmonics rather than higher harmonics. The phases of the first ten harmonics with piston slap faults are shown in Figure 10 (including all three speeds and three loads).

Figure 10. Phase distribution of the first ten harmonics of piston slaps
Specifically, if the phases of the 1st harmonic of all cases are clustered, or near fixed, it means that there will be 90 degree difference for the faults in individual cylinders (60 degree difference for a six cylinder engine); therefore the fault localization can be directly and solely identified by the phases of the 1st harmonic. Moreover, if a fault occurs in the cylinder 1 or cylinder 3 (firing sequence 1-3-4-2), the phase difference of the 1st harmonic is 90 degrees, and the phase difference of the 2nd harmonic is 180 degrees (90×2), so using the clustered phases of the 2nd harmonic can differentiate the faults in cylinder 1 from those in cylinder 3. On the other hand, if a fault occurs in cylinder 1 or cylinder 4, the phase difference of the 1st harmonic is 180 degrees, and the phase difference of the 2nd harmonic is 360 degrees (180×2), so using the clustered phases of the 2nd harmonic cannot differentiate the faults in cylinder 1 from those in cylinder 4. Ideally, the phase difference of the 2nd harmonic can only identify the location of fault in 50% of cases. As shown in Figure 10, the phases of the 1st harmonic for the piston slap faults are not well fixed but the phases of the 2nd harmonic are nearly fixed, it means the phases of the 2nd harmonic can be used to differentiate the piston slap faults in the cylinders 1 and 4 from those in the cylinders 3 and 2 (firing sequence is 1-3-4-2). Meanwhile, the phases of 3rd, 5th and 7th harmonics are nearly fixed and they are odd number harmonics (90 degrees by any odd number is not an integer multiple of 360 degrees), so they can be added to the phases of the 2nd harmonic to identify which cylinder has a fault. Note that the 4th harmonic represents the firing frequency, so the phases of the 4th harmonic and the (4*n)th harmonics are useless to identify the localization of the mechanical faults.

To better understand the phase selection method, we generated very simple perfect periodic impulse signals with ideal phase shifts in time, so can study the corresponding phase change of their Fourier coefficients. These signals have same period of 800 samplings, but we need two periods to do further Fourier analysis (we will explain why need two periods later on). Signal 1 is similar to the dominant squared envelope impacts happening in cylinder 1 (as shown in Figure 11), Signal 2 for the dominant impacts happening in cylinder 3(200-sampling shift from Signal 1), and Signal 3 for the dominant impact happening in cylinder 4 (further 200-sampling shift from Signal 2). Again, remember that the firing sequence is 1-3-4-2 and we only study the phase characteristic of Fourier coefficients here (the amplitude information is not our concerned here). It is worth mentioning that if other minor thrust side impacts (small impulses) are included into the time signals, it only affects the amplitudes of Fourier coefficients, has no influence on phases. Next, the Fourier coefficients of three signals were studied and first 5 harmonics are shown in polar diagrams in Figure 12. It can be seen that there are 90-degree differences for the 1st harmonic from Signal 1 to Signal 2, further to Signal 3; 180-degree differences for the 2nd harmonic (the phases of Signal 1 and Signal 3 are same); 270-degree differences for the 3rd harmonic and 360-degree differences for the 4th harmonic (same phase for all three signals and this explains why 4th harmonic is useless in the preceding paragraph). But for real piston slap fault cases, there are lots of noise, variations and uncertainties during the engine operation, and the time signal definitely is not perfect periodic, therefore we need study the clustering characteristic of phases.
It is also worth mentioning if only one period of time signal was studied and its Fourier coefficients was presented by the “polar” command of Matlab in a polar diagram, it was found that the vector cannot return to zero point of polar diagram, so in order to better present the amplitude and phase information, we need do Fourier analysis on two periods of time signals and also need double the line number selection for the corresponding harmonics in frequency domain (so the analysis can return to one period basis), for instance, the amplitudes and phases of the first harmonic and second harmonic should be respectively get by Matlab commands:

\[
[\text{theta1}, \text{rho1}] = \text{cart2pol}(\text{real(ACC(2:3,1))), imag(ACC(2:3,1)))};
\]
\[
[\text{theta2}, \text{rho2}] = \text{cart2pol}(\text{real(ACC(4:5,1))), imag(ACC(4:5,1)))};
\]

here \(\text{theta}\) is the phase, \(\text{rho}\) is the amplitude, and \(\text{ACC}\) is the Fourier analysis value. This is also another reason why all raw and processed vibration signals in this paper were presented in two-period form.

4. Simulation of piston slap fault

Owing to the high speed linear movement of the piston, direct metal to metal contact between the piston and inner wall would inevitably lead to huge friction loss and engine seizure. So in reality, lubrication oil is used to reduce the level of contact and the friction resistance to a minimum to ensure maximum mechanical efficiency. If the lubricating oil functions properly and is sufficient thick that there is no asperity contact between the piston and inner wall, the reaction force from the inner wall against the piston can be solved by the classic Reynolds equation. However, during the engine operation, especially with very high load or contaminated lubricant, the oil film will become extremely thin and the load is mainly (or even completely) supported by metal to metal asperity contacts. This condition is called boundary condition or dry contact [34]. Under this condition, the impact force can be solved by the well known Hertzian contact theory. Between the good lubrication and boundary lubrication, there is a transition condition known as mixed lubrication. Under mixed lubrication, both the hydrodynamic effect from the lubricating oil and the elastic effect from asperity contact should be taken into account and the total impact force should be shared by both effects [10, 34].

If there is dry contact between the piston and inner wall, even though the translation and rotation of the piston are complicated during the engine operation, in general, there are three contact modes for the piston
and inner wall: two point contacts, line contact and single point contact. As shown in Figure 13, the line contact mode (the middle one in Figure 13) can be simplified as upper and lower points on one side of the piston simultaneously in contact with the inner wall (although the Hertzian stiffness exponent would be slightly different). Therefore, all the contact modes can be considered as point to point contact and the classical Hertz contact theory can be used in the modelling of piston slap faults.

Figure 13. Contact modes of piston/inner wall

If the lubrication system works properly, the oil film is always 100% full within the radial clearance over the whole 360 degree cycle and the axial length of the piston. By neglecting the inertia forces of the lubricating oil, the lubrication pressure around the piston skirt can be solved by the Reynolds equation, based on the pressure of the lubricating oil, the normal force in the radial direction and the friction force along the central axis of piston can be integrated as well. Even though the Reynolds equation can solve the impact force under good lubrication conditions (no oil film break), it needs a lot of computing resources to solve those equations (cumbersome numerical integration) [16]. On the other hand, the main effect of the lubricating oil on the impact force should be equal to a squeezed film damper. Therefore, when studying piston slap, some researchers proposed to use squeeze film dampers to simplify the effects of the lubricating oil, rather than to solve Reynolds equation [17, 21]. Very similar to their works, a model was proposed to represent the dynamic characteristics between the piston and inner wall when there is an oversized clearance. In this model, the piston and inner wall were modelled as lumped masses. The effect of the piston ring and lubricating oil was modelled as a viscous damper and spring between the piston and inner wall (shown in Figure 14). The spring and damper are fixed length independent of the clearance, so that they only come into play when the piston is close to the inner wall.

Figure 14. Interaction model between the piston and inner wall

The end of the spring/damper was fixed on the engine block. When the piston contacts the surface of the inner wall, the dynamic model enters a boundary condition (dry contact) and the effects of spring and damper are negligible. The impact force caused by the elastic contact will be solved by the Hertz theory.

Figure 14. Interaction model between the piston and inner wall

In real engines, normally the inner walls are tightly pressed into the bore of the engine block and become a cylinder liner. So it is necessary to model the inner wall separately in the simulation. The accurate mass of the inner wall is not critical, because hydrodynamic effects and dry contact do not involve the mass of the inner wall, but the dry contact needs the material properties of the inner wall. Meanwhile, as mentioned
above, the spring and damper were fixed to the block, instead of the inner wall, so the spring can continue to be compressed after dry contact and the length of spring is always positive. Therefore, in the simulation model, the inner wall was modelled separately but needed to be connected with the engine block with a rigid joint.

The simulation model of piston slaps was built in the Motion module of Virtual.Lab. The dynamics of piston slap mainly involves the piston, connecting rod, crankshaft (plus flywheel), inner wall and engine block, so modelling these components in Motion is sufficient to simulate the piston slap faults. The piston and inner wall were treaded as rigid (lumped masses). Meanwhile, for the purpose of simplification, the crankshaft (flywheel) and connecting rod were also modelled as rigid. The revolute friction type was introduced into the joint between the piston and piston pin. Actually, no real spring or damper elements were used in the real simulation model built in the Motion. Instead, the spring and damping forces were expressed as a function.

As shown in Figure 15, the top axis system and bottom axis system were defined at the centre points of the top and bottom surfaces of the piston respectively (the bottom surface is hollow, and it is also worth mentioning that the direction difference between the coordinate of simulation model and that in Figure 1 will not affect the target simulation results). Two axis sensors were then inserted, one at the top axis system and one at the bottom axis system of the piston. One axis sensor was located at the centre line axis system of the inner wall. During the simulation, the axis sensors can record the relative displacement ($u_p$) and relative velocity ($\dot{u}_p$) between the piston and the inner wall. The reaction force from the piston ring and the lubrication can be calculated from:

$$F_o = ku_p + cu_p$$  \hfill (3)

The values of the damping and stiffness were obtained from reference [21] (but they were adjusted in accordance with the final simulated envelope signals). The calculated force $F_o$ was exerted at the centre point on the top surface and centre point on the bottom surface of the piston. Therefore, only virtual spring and damper elements were used in this model.

In order to model the spring/damper units as fixed (free) length, i.e independent of the clearance, a mathematical cubic step function was used (shown in Figure 16). Compared to the Boolean function, the advantage of the cubic step function is that it is free from the problem of instantaneous parameter change, which may cause oscillations in the simulated forces and accelerations. The cubic step function is a blending function. For values of $u$ less than $x_{\text{min}}$, step has the value $y_{\text{min}}$. For the values of $u$ greater than $x_{\text{max}}$, step has the value $y_{\text{max}}$. For values of $u$ in between $x_{\text{min}}$ and $x_{\text{max}}$, step returns the value of the cubic function fitted.
between the two points \( y_{\text{min}} \) and \( y_{\text{max}} \). In the piston slap fault simulation model, the variable \( u_y \) is the displacement in the \( Y \) direction between the axis sensors on the piston and the axis sensor on inner wall and the outputs of the step functions are a stiffness coefficient and a damping coefficient. The \( y_{\text{min}} \) was set to zero. Meanwhile, the top and bottom points shown in Figure 15 (4 points for each piston) were defined as spherical contact points. When the piston skirt touches the surface of the inner wall, the Hertz contact mode will come into operation.

The Hertz contact in simulation of piston slap fault is based on the contact force:

\[
F_{\text{cont}} = k_{\text{cont}} \delta^{3/2}
\]

(4)

Where \( k_{\text{cont}} \) is the contact stiffness, \( \delta \) is total deformation of both contact surfaces. The contact stiffness \( k_{\text{cont}} \) depends on the material properties of the impacting structures and the contact surface geometry and is dimensional so it depends on the units used.

\[
k_{\text{cont}} = k \left( 1 - \frac{1-e^2}{1+e^2} \right) \tanh \left( \frac{2.5V_{\text{pen}}}{V_{\text{eps}}} \right)
\]

(5)

Where \( e \) is the coefficient of restitution, \( V_{\text{pen}} \) is penetration velocity, \( V_{\text{eps}} \) is minimum penetration velocity and the stiffness parameter \( k \) can be obtained from Virtual.Lab manual 2012 [35].

Because the vibration signals are recorded at the points outside of the engine block (as mentioned in Section 3.1), the engine block was meshed as a flexible model. Consequently, the vibration signal at any node points of the meshed model can be directly simulated. The block was meshed by the embedded FEA tool of Virtual.Lab, and the mesh type is tetrahedron, with the global mesh size of 40mm and local mesh size of 10mm. The flexible FE model of the block was updated using the experimental transfer function. Using Maxwell's Theory of Reciprocity, during the measurement of the transfer functions, the shaker excitation force was applied on the external points (the locations of accelerometer 5 and 6 in Figure 2) and the accelerometer was attached at the thrust points on the inner wall of cylinder 1 and 2 (4 points for each cylinder). Totally, there are 10 dofs from experiment wireframe. The experimental frequency is up to 6400 Hz. The model updating is carried out by the Correlation tool and Optimization tool of Virtual.Lab. After the nodes in FEA were connected to the nodes in the test wireframe, the modal correlation value can be evaluated by Correlation. In the Optimization, the geometry parameters (thickness) of engine block can be adjusted based on the modal correlation value from the Correlation, and the optimization target is to make the modes between 400Hz to 4000Hz approximately match with the experimental modal frequencies.

The cylinder chamber pressures for different speeds/loads were first calculated in Matlab. The compression pressure can be derived from the polytropic formula and the combustion pressure can be calculated by classic Wiebe’s functions [36]. The details of pressure calculation can be obtained from our previous work [37-40]. The pressure calculation was evaluated and updated using the averaged measured cylinder pressure. The pressure model was updated (parameter adjustment) is based on only one speed/load condition, then the model was reasonable accurate for the pressure calculation of remaining speed/load conditions. The calculated cylinder pressures were the inputs to the combustion element in the model.

After the model was computed, by changing the inner diameter of the inner wall, two stages oversized piston clearances (three-time normal clearance and six-time normal clearance) were created. The vibration signals for a range of piston slap faults were simulated. The reaction force of the inner wall against the piston can be calculated and an example at 1500rpm/80Nm with 2\textsuperscript{nd} stage piston slap fault is shown in the Figure 17. The impact forces around 0 and 360 degrees of crank angle act on the major thrust side and another two around 180 and 540 degrees of crank angle act on the minor thrust side. As studied before, the impact always occurs when the piston moves from one side to the other. The lateral motion of the piston at 1500rpm/80Nm with 2\textsuperscript{nd} stage piston slap fault is shown in Figure 18. From the graph of displacement, it can be seen that the piston is sometimes in horizontal “flying” modes when it moves from one side of the inner wall to the other side, so the reaction force is zero under this condition. After 0 degrees of crank angle, owing to the combustion force, the piston moves very rapidly from the minor thrust side to the major thrust side and the lateral velocities are very high, so the impact force is much bigger than for the other three impacts in each cycle.
For the piston slap faults at different speeds/loads or with different clearances, the simulated vibration signals are very similar, except that the magnitudes change and there are small phase shifts. Obviously, the numbers of samples for a whole cycle at different speeds change considerably, but they can all be re-sampled on the basis of the crank angle. An example of the comparison between the simulated acceleration and experimental acceleration at 2000rpm/110Nm with 2nd stage piston slap faults is shown in Figure 19. Note that, in the Figure 17, there is no big difference among the reaction forces around 180, 360 and 540 degrees of crank angle, but because the location of the simulated vibration point is close to the major thrust side, the acceleration around 360 degrees of crank angle is bigger in Figure 19. Evidently, the simulated acceleration signals have less noise contamination. But as mentioned above, for the piston slap fault diagnostic system, the simulations only need to produce the correct envelope signals, so the raw acceleration signals do not have to be reproduced exactly (the experimental acceleration signals have more noise contamination).
Before the envelope processing, the simulated acceleration signals were also passed through a band-pass filter. Examples of experimental and simulated envelope signals are shown in Figure 20. It can be seen that the simulated envelope signals have good agreement with experiment. In particular, the ratios of the largest amplitudes at two different clearance stages are correctly simulated.
5. Automated diagnosis system

Recently, the architectures of ANNs have evolved in various forms and have been broadly applied in many industrial applications, such as classification and pattern recognition. Limited works have been devoted to the application of ANNs for the fault diagnosis of IC engines but they have demonstrated successful results. Jenner [41] used ANNs to detect and classify misfire faults in a certain cylinder of an engine. As mentioned before, Desbazeille et al. [31] developed a three stage ANN system; the first stage is to detect whether the engine is good or not, the second stage is for when the engine is faulty, to point to which cylinder; thirdly how serious is the fault in that cylinder. Ortmann et al. [42] proposed a system that consists of a two-level feature extraction step followed by a neural network detector to detect the engine knock. Moreover, these existing works were concentrated on the aspect of combustion faults; almost no work has been done on ANN-based fault diagnosis of piston slap faults.

The simulated signal for a certain speed/load condition is deterministic, but in reality the measurement signals have small deviations in different test scenarios, so a variation in the simulated signals should be instituted to create representative cases for the ANN training. The experimental period of each stage piston slap fault ranged over two months, and the ambient temperature had over 20 degrees variation (day time & night), therefore it was found that there are some variations in the processed experimental signals (up to 5% variation in the same condition but different experiment times). Therefore, the standard deviation of the variations was set by the analysis the experiments signal in normal conditions and was applied to the simulated envelope signals. The piston slap faults were also simulated in different cylinders for each speed and load condition and all the simulated signals were instituted with variations. For each simulated case, four variations were created. Three-stage network systems were designed for the automated diagnosis of piston slap faults. The first stage is the piston slap detection stage and the inputs to the networks are the selected amplitude features of the envelope signals. In the second stage, the neural network localizes which cylinder has piston slap faults and the inputs to the network are the selected phase features of the envelope signals. In the third stage, based on the detection and location results, the severity of the piston slap faults is identified. As mentioned in Section 3, the experimental cases cover all three speed and three load conditions for two clearance stages, normally it took 3 to 5 days to finish the measurements for three speed and three load conditions, so in terms of the operation speed and load, the spread of cases should be robust enough to guarantee the reliability of ANN results.

Feed-forward MLP neural networks were used to detect whether there are piston slap fault and to identify the severity of piston slap fault. The diagram of an MLP is shown in Figure 21. Two MLPs were designed: one for the fault detection (MLP1) and the other for the severity identification (MLP2). The MLPs consist of three layers: input, hidden and output. The number of hidden neurons of the MLPs was determined using a trial and error procedure (that shown in Figure 21 has 30 hidden neurons). $w$ and $b$ are respectively the weight and bias factors distributed to the individual elements of the input feature vectors. Identifiers 1 and 2 indicate the different weight vectors and bias vectors for the hidden layer and the output layer. During the training stage, MLPs were led to a specific target output by adjusting the values of the connections (weights and bias) among the elements of the input vectors.
Due to its advantage for classification problems, a PNN was used to identify which cylinder has misfires. The PNN is based on the weighted-neighbour method and was proposed by Specht [43]. The distance is computed from the point being evaluated to each of the other points, and a radial basis function is applied to the distance to compute the weight for each point. The structure of a typical PNN is shown in Figure 22. The outputs of the PNN for the misfire diagnostics are the integer numbers 1, 2, 3 and 4, which directly indicate the cylinder number.

As studied in Section 3.3, when the amplitude features of piston slap faults are compared to those for normal condition, the amplitudes with piston slap fault are much larger than those in normal condition. Therefore, the amplitude information can definitely be used for the detection of piston slap faults. The MLP used in the detection stage was noted as MLP1. Because log sigmoid function can efficiently classify all cases into two groups (converging to two boundary conditions), it was used as the transfer function of both layers of MLP1; for the output, 0 represents normal condition and 1 for piston slap fault. But as mentioned before, because the increase ratio of the amplitude peak value of squared envelopes from the first stage fault to the second stage fault is not fixed and strongly related to the speeds and loads, it is not easy to accurately detect the fault severity identification under all speed/load conditions by one MLP; meanwhile it also found that the increase of the peak value has approximately linear relationship with the increase of load, but no linear relationship with the increase of speed. Therefore, different from our former work [21, 22], in order to get finer and accurate classification result in severity identification stage, three MLPs (noted as MLP2) were designed for each speed condition, and the input amplitude features of the individual MLP were scaled (divided) by the load values. Moreover, a linear transfer functions purelin(x), rather than a nonlinear log sigmoid function, was introduced into output layer of new MLP2 as its transfer function (only the hidden layer of new MLP2 used log sigmoid function, but both transfer functions in the hidden layer and output layers of old MLP2 used log sig functions). This function is linear in the interval (-∞, +∞) and returns the value of x in the range 0 to +∞. So the output of MLP2 can be any value linearly distributed from 0 to +∞ and output results of severity identification agree more logically with the real situation. Even though only maximal 6-time normal clearance was seeded in the experiment, there is a strong likelihood that larger clearance faults appear in reality. Therefore, during the training process, same as the old MLP2, the output of 6-time normal clearance was set as 1 and the output of 3-time normal clearance was set as 0.5 (note that 7-time normal clearance could be indicated by the output 1.17 in the new MLP2, but could not by the output in the old MLP2). The number of hidden neurons of all MLPs was determined using a trial and error procedure. A fitness criterion was introduced to evaluate the performance of the MLP:
\[
\text{Error} = \frac{1}{N} \sum_{i=1}^{N} |\text{ANN}(i) - \text{VAL}(i)|
\]

where, \(\text{ANN}\) is the output of the MLP and \(\text{VAL}\) is the corresponding target number. \(N\) is the total number of the test group. A higher fitness criterion means poorer MLP performance. The trial and error procedure about the number of hidden neurons is also based on the values of fitness criterion. PNNs (Probabilistic Neural Networks) were used to identify which cylinder has faults. The outputs of the PNN for the misfire diagnostics are the integer numbers 1, 2, 3 and 4, which directly indicate the cylinder number.

All training data for the automated system are from the simulated (squared) envelope signals. The training cases included 288 piston slap faults (with oversize clearance) and 112 cases with normal piston clearance. 49 experimental cases were used to test the networks, of which 28 cases are with piston slap faults and 21 cases are with normal piston clearance (some of them having combustion faults). For the PNN, the number of training cases is reduced to 288 and the number of test cases is reduced to 28. In the MLP2 of the severity identification stage, the training and test group are reduced to piston slap fault cases for the individual cylinder only. The outputs of MLP1 are shown in Figure 23. In reality, there are piston slap faults for cases 1 to 18 and 33 to 42. The fitness criterion in Equation (6) was also introduced to test the performance of MLP1. The final fitness criterion is 4.598*10^{-3}, so the MLP1 100% correctly detected all piston slap faults.

![Figure 23. Output of MLP1 for the piston slap faults](image-url)
Using the selected phase features as inputs, the PNN also achieved a good classification result during the fault location stage. The results 100% match with the real situation of the test cases: all piston slap faults (28 cases) occurred in cylinder 1 in the experiments. After the piston slap fault detection stage, the severity identification became more specific. As mentioned before, besides the new transfer function, the new MLP2 were designed as three separated MLP networks for three different speeds, rather than using one MLP to cover all speeds in the old MLP2 design; in order to demonstrated how is the improvement of the new designed MLP2, the number of the total training cases (288) and the number of the total test cases (28) are same for both old and new MLP2 designs (for the new MLP2, the numbers of training cases for each speed are 96, 96 and 96, and the numbers for test cases are 9, 9 and 10). For comparison, the three MLP output results of the new MLP2 are combined together and shown in Figure 24 (a), and the results of the old MLP2 are shown in Figure 24 (b). It can be seen that the new MLP2 100% correctly identified the severity classification of the piston slap faults with a small deviation. The new MLP2 need more computing time (more epochs) to achieve the training goal and the final fitness criterion (3.937*10^{-2}) is larger compared to the detection stage, but lower than the fitness criterion (4.761*10^{-2}) of the old MLP2; therefore the new designed MLP2 (with linear transfer function and separated networks for individual speed) obtained better results and the linear outputs agreed more logically with the real situation.

6. Conclusion

This paper expands the former works and updates the network design of automated piston slap diagnosis (a model based automated diagnosis system). In order to train the networks by adequate data in different engine conditions (normal and faulty conditions in different speeds/loads), an advanced multi-body simulation model was built in this paper. Beforehand, a small number of experiments were carried out so as to evaluate and update the simulation models. Signal envelope processing with band pass filtering was used to process experimental signals to extract potential features. The standard deviations derived from experimental measurements in normal condition were applied to the simulated envelope signals to create suitably varied cases. The simulated acceleration signals were processed by the same signal processing techniques as the measurements and the selected features from the processed data were used to train the networks. The automated system consists of three stages: MLP1 for the detection stage, PNN for the localization stage and MLP2 for the severity identification stage. In particular, for the severity identification stage, separate MLPs with pure linear transfer functions were designed for the three speed conditions, so the outputs of the MLP2 can be any value linearly distributed from 0 to +∞ and the results can more logically indicate different severe levels of faults in the real situation. The networks was trained purely by simulation data and tested by the experiment data. Finally, it has been demonstrated that even though the networks were only trained by simulation data, they can efficiently detect the piston slap faults in real tests and identify the location and severity of the faults. More importantly, compared to results from the old MLP2, the new designed MLP2 had finer classification on the severity of the piston slap faults.
Acknowledgements

The authors would like to convey special gratitude to the Australian Research Council and LMS International (now Siemens PLM Software) for sponsoring this research under Linkage Project LP0883486. Special thanks to Professor Asoke Nandi at Brunel University London, for his helpful discussions and suggestions on the design of MLPs during Surveillance 7 conference in Chartres, France.

References
