J. Acoust. Soc. Jpn. (*E*) 7, 6 (1986)

Accurate estimation of the surface roughness on the rolling ring in a ball bearing by vibration analysis

Hiroshi Kanai, Masato Abe, and Ken'iti Kido

Research Center for Applied Information Sciences, Tohoku University, 2-1-1, Katahira, Sendai, 980 Japan

(*Received 24 January 1986*)

This paper describes a vibration based diagnostic method for estimating the surface roughness on the race in ball bearings. The surface roughness has been measured by a stylus which directly traverses the surface of the ring obtained by taking apart the ball bearing. We developed a new method to accurately estimate the surface roughness by analyzing the short-length vibration signal which is excited when balls contact with the surface of a defective rotating ring in a ball bearing. Our experimental results confirm that the roughness estimated by the proposed method agrees with that measured directly by using a stylus even in the case of crack μm wide. We applied this new method to the diagnosis of surface roughness in small-sized ball bearings and inferior samples were detected with a 94.7% accuracy rate.

PACS number: 43. 85. Ta, 43. 60. Gk

1. INTRODUCTION

Vibration is induced when a ball bearing has a rough surface or flaws on the race. For example, in the case of the ball bearings used to support the spindle of the head of a video tape recorder, such defects debase the quality of the played-back picture. There have been three diagnostic methods to detect the surface roughness or flaws on the raceway:

(A) The surface roughness has been measured by a stylus which directly traverses the surface of the ring. However, in this case it is necessary to take apart the ball bearing. Thus, it is impossible to measure the surface roughness on the race of the assembled ball bearing.

(B) Therefore, at the final step in the manufacturing process, defects in ball bearings have been detected and classified aurally by inspectors listening to vibration signals picked up using the Anderon meter[1] as described in Chap. 2. However, it requires a great deal of time to train a good inspector. Additionally, the physical and mental condition of inspector affects the results of detection and classification.

(C) In response, several methods have recently been proposed for the automatic detection and classification of the defects in ball bearings.[1-8] These methods utilize the pulses excited when the balls contact with flaws on races. Defects are detected and classified by checking the peaks observed at the expected intervals which are calculated from the position of the flaws, the shape of the bearing and the rotation speed of the inner ring. We proved by experiments that the defects can be classified with an accuracy rate of 97.9%.[8] However, the shape of the surface roughness or the number of flaws on the raceway cannot be detected by this kind of periodicity analysis.

Therefore, we have developed a new diagnostic method to accurately estimate the surface roughness by analyzing the short-length vibration signal excited by the contact of balls with the flaws on the surface of the rotating ring. The principle is described simply in Chap. 3.1 and is described in detail in Chap. 3.2. The processes to determine the parameters used in the proposed method are described in Chap. 4. Since there is scattering in the repetition

intervals of the flaw pulses generated by many balls in the practical case, it is difficult to estimate the surface roughness using the method proposed in Chap. 3. For this reason, the method is modified in Chap. 5 to estimate surface roughness more consistently. In Chap. 6, it is confirmed with the experimental results that the roughness estimated by the proposed method agrees with that measured directly by using a stylus even in the case of crack of μm. In Chap. 7, we apply this new method to diagnosis of the surface roughness in small-sized ball bearings and it is shown that the inferior samples are detected with a 94.7% accuracy rate.

2. APPARATUS FOR THE EXPERIMENTS

Figure 1 shows a block diagram of the experimental system in which the vibration signal is picked up and recorded on a digital audio tape to carry back to the laboratory.[3,4] The inner ring rotates at a constant speed of 1,800 rpm, and the outer one is fixed by imposing an axial pressure in an Anderon meter.[1] Under such the conditions, a surface roughness causes a radial movement of the outer ring and the signal resulting from this movement is picked up by a vibration pick-up attached to the outer ring. The signal is amplified and filtered through a high-pass filter to attenuate the primary frequency component

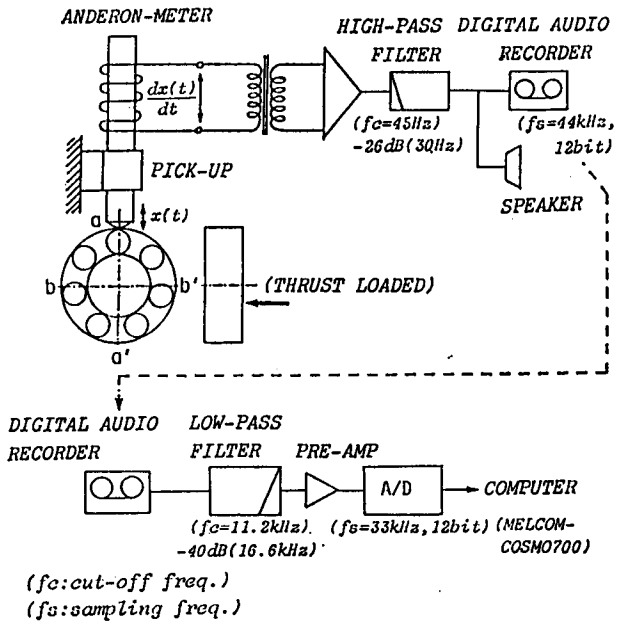


Fig. 1 The block diagram of the procedure for measuring the vibration signal of a ball bearing.

($F_i = 30$ Hz) corresponding to the rotation of the inner ring. The filtered signal is stored in a high-fidelity digital audio tape. In the laboratory, the signal is played back and is A/D converted with a 12 bit A/D converter at a sampling period of 30 μs.

3. ESTIMATION OF THE SURFACE ROUGHNESS BASED ON THE VIBRATION ANALYSIS OF BALL BEARINGS

3.1 Theoretical Explanation Using a Simplified Example

When the inner ring is revolved at a constant speed of $\omega_i = 2\pi F_i$ [rad/s] as described above, the ball also revolves at a constant speed of $\omega_b = 2\pi F_b$ [rad/s]. The value of ω_b is derived theoretically using the rolling speed ω_i and the shape factors of all parts[1,2] as follows:

$$\omega_b = \frac{\omega_i}{2} \cdot \left\{1 - \left(\frac{d}{D}\right) \cdot \cos\delta\right\}, \tag{1}$$

where d is the ball diameter, D the pitch diameter, and δ the contact angle. When a revolving ball arrives at the edge of the flaw on the raceway, the ball rotates around the edge, and the ball hits the other edge of the flaw. It is known that the amplitude $s(x)$, $(0 \leq x < 2\pi)$ of the impulse generated by the collision is in proportion to the 6/5-th power of the width of the flaw at the point x on the race.[4-6] Therefore, the function $s(x)$ used below agrees with the surface roughness.

Figure 2 shows a simplified ball bearing having only two balls, α and β, which are revolving π rad apart. At a time $t = t_1$, the balls α and β contact with the points a and b on the inner race, respectively, as shown in Fig. 2 (a). The resonant vibrations are driven by the impulses $s(a)\cdot\delta(t-t_1)$ and $s(b)\cdot\delta(t-t_1)$, where $s(x)$ denotes the amplitude of the impulse and $\delta(t)$ denotes the unit impulse:

$$\delta(t) = \begin{cases} 1 & (t=0) \\ 0 & (t \neq 0). \end{cases} \tag{2}$$

Since the characteristics of the impulse response $h(t)$ of the resonant vibration are determined both by the size and by the material of the outer ring, the two resonant vibrations driven by the impulses $s(a)\cdot\delta(t-t_1)$ and $s(b)\cdot\delta(t-t_1)$ have the same resonant characteristics except for the amplitude of each signal (see Fig. 3 (a)). Therefore, when the picked up vibration signal $y(t)$ is deconvoluted by the inverse filter

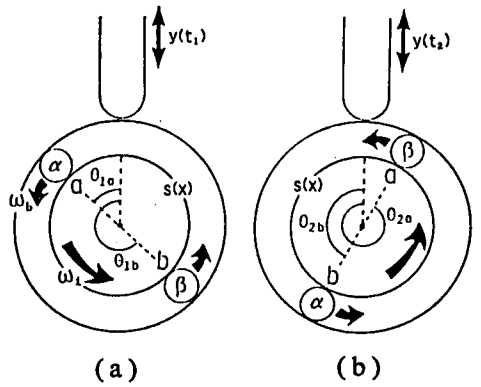


Fig. 2 Illustration of the simplified ball bearing showing the relation between the surface roughness $s(x)$ and the equivalent driving impulse $z(t)$. (The signal $z(t)$ denotes the weighted sum of the impulses caused by the flaws of the points a and b on the race.) (a) at a time $t=t_1$. (b) at a time $t=t_2$.

$$z(t_1)=y(t)*h(t)^{-1}|_{t=t_1}$$
$$=W(\theta_{1a})\cdot s(a)+W(\theta_{1b})\cdot s(b)$$
$$z(t_2)=y(t)*h(t)^{-1}|_{t=t_2}$$
$$=W(\theta_{2a})\cdot s(a)+W(\theta_{2b})\cdot s(b)$$

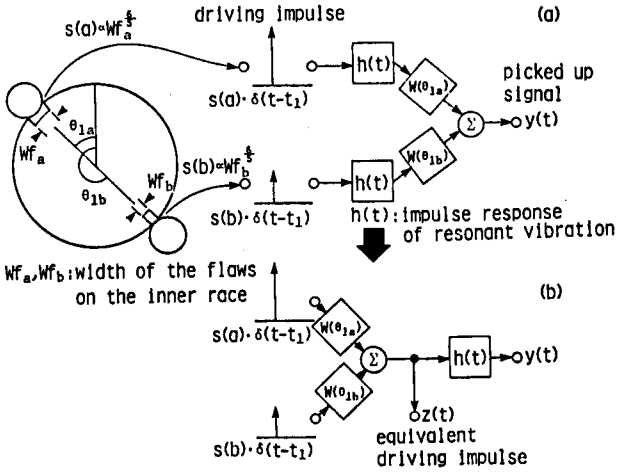


Fig. 3 (a) The picked up signal $y(n)$ is the weighted sum of the resonant vibrations which are driven by the impulses $s(a)\cdot\delta(t-t_1)$ and $s(b)\cdot\delta(t-t_1)$. (b) When the picked up signal $y(n)$ is deconvoluted by the inverse filter $h(n)^{-1}$ of the resonant vibration, the resultant driving impulse $z(t)$ denotes the weighted sum of the driving impulses due to the collisions as shown in Eq. (4).

$h(t)^{-1}$ of the resonant vibration

$$h(t)^{-1}*h(t)=\delta(t)\,, \tag{3}$$

where $*$ denotes the convolution, the resultant driving impulse $z(t)$ denotes the weighted sum of the driving impulses due to the collisions on the points a and b as follows:

$$z(t)=y(t)*h(t)^{-1}$$
$$=W(\theta_{1a})\cdot\{s(a)\cdot\delta(t-t_1)\}$$
$$+W(\theta_{1b})\cdot\{s(b)\cdot\delta(t-t_1)\}\,, \tag{4}$$

where θ_{jx} denotes the angle between the pick-up and the point x with which a ball contacts on the race at a time $t=t_j$, and $W(\theta_{jx})$ is the weight function representing the amplitude ratio of the excited impulse $s(x)\cdot\delta(t-t_j)$ at the point x to the equivalent driving impulse $z(t)$ at the pick-up (see Fig. 3 (b)).

Next, at a time $t=t_2$ the balls α and β contact again with the points b and a on the race, respectively. The impulses are excited due to flaws on the points a and b, and the equivalent driving impulse $z(t)$ of Eq. (4) is modified as follows:

$$z(t)=\{W(\theta_{1a})\cdot s(a)+W(\theta_{1b})\cdot s(b)\}\delta(t-t_1)$$
$$+\{W(\theta_{2a})\cdot s(a)+W(\theta_{2b})\cdot s(b)\}\delta(t-t_2)\,. \tag{4'}$$

The equivalent driving impulse $z(t)$ is expressed at $t=t_1$ and $t=t_2$ as follows:

$$z(t_1)=W(\theta_{1a})\cdot s(a)+W(\theta_{1b})\cdot s(b) \tag{5a}$$

and

$$z(t_2)=W(\theta_{2a})\cdot s(a)+W(\theta_{2b})\cdot s(b)\,. \tag{5b}$$

All of the terms other than $s(a)$ and $s(b)$ involved in the above equations are determined when the following three conditions are satisfied:

(A) The values of the angular velocity ω_i and ω_b are determined previously and the initial angle θ_0 of a ball relative to the pick-up at a time $t=0$ is determined. By using these parameters, the following two values are calculated: the point x with which the i-th ball contacts on the race at a time $t=t_j$, and the angle θ_{jx} of the i-th ball relative to the pick-up at a time $t=t_j$.

(B) The weight function $W(\theta)$ is determined previously.

(C) The inverse filter $h(t)^{-1}$ of the resonant vibration is determined. Then, the equivalent driving impulse $z(t)$ is calculated from the convolution between $h(t)^{-1}$ and the observed vibration $y(t)$ as shown in Eq. (4).

Therefore, the surface roughness quantities $s(a)$ and $s(b)$ are determined by solving the set of simultaneous Eqs. (5a) and (5b). Since the vibration signal $y(t)$ is generally corrupted by additive noise, more accurate surface roughness is estimated by using the least mean square fitting to the P equivalent driving impulses $\{z(t_j)\}$, $(j=1, 2, \ldots, P)$, where P denotes the number of collisions between the points a and b and balls in the observation interval.

3.2 Estimation of the Surface Roughness

This paragraph describes the principle of the proposed method when a ball bearing has N balls, each of which is $\Omega_0=2\pi/N$ [rad] apart on the race (see Fig. 4). The observed signal $y(t)$ is the weighted sum of the resonant vibrations driven by the impulses excited when each ball contacts with the point on the race as shown in Fig. 3 (a). As described in Chap. 3.1, each resonant vibration driven by the impulse has the same characteristics except for the amplitude of each signal, the observed signal $y(t)$ is the resonant vibration driven by the equivalent impulse $z(t)$ which is the sum of the impulses excited when each ball contacts with the point x on the race (see Fig. 3 (b)). When the initial angle of the first ball ("0") relative to the pick-up is θ_0 [rad] at a time $t=0$ as shown in Fig. 4 (a), the angle of the i-th ball relative to the pick-up is equal to $\theta(i, t: \theta_0)=\omega_b \cdot t+\Omega_0 \cdot i+\theta_0$, $(i=0, 1, \ldots, N-1)$ at a time t as shown in Fig. 4 (b). At a time $t=0$, the point '0' on the race is just under the pick-up as shown in Fig. 4 (a). Since the inner ring is revolved at a speed of ω_i, the angle between the point '0' on the race and the point with which the i-th ball contacts on the race is equal to $x(i, t: \theta_0)=(\omega_b-\omega_i)t+\Omega_0 \cdot i+\theta_0$ [rad] at a time t as shown in Fig. 4 (b). At a time t, the angle between the pick-up and the point $x(i, t: \theta_0)$ with which the i-th ball contacts is equal to $\theta(i, t: \theta_0)=\omega_b \cdot t+\Omega_0 \cdot i+\theta_0$. Thus, the observed signal $y(t)$ is expressed as follows:

$$y(t)=z(t)*h(t) \tag{6a}$$

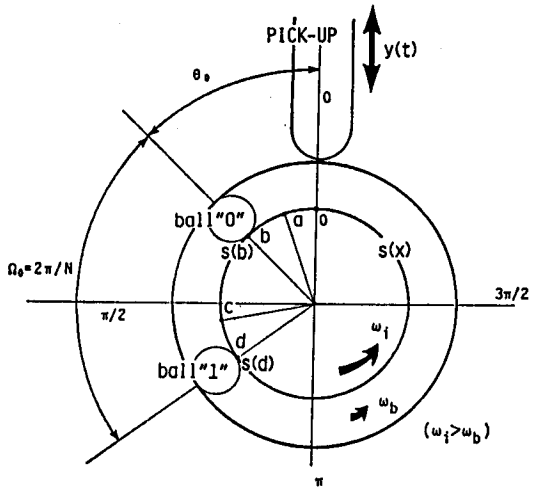
and

$$z(t)=\sum_{i=0}^{N-1} W((\theta(i, t: \theta_0))_{2\pi}) \cdot s((x(i, t: \theta_0))_{2\pi}), \tag{6b}$$

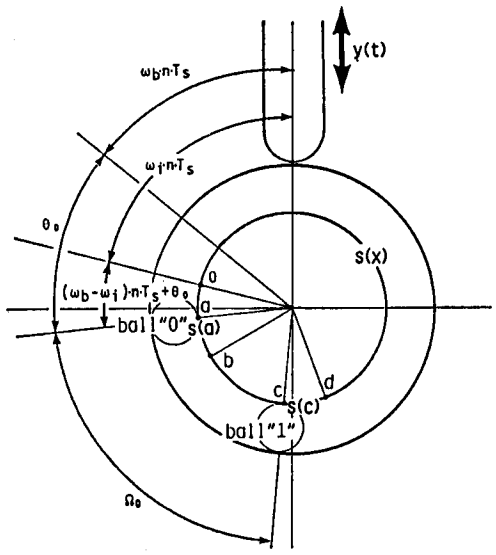
where

$$\theta(i, t: \theta_0)=\omega_b \cdot t+\Omega_0 \cdot i+\theta_0,$$
$$x(i, t: \theta_0)=(\omega_b-\omega_i)t+\Omega_0 \cdot i+\theta_0,$$

and $((\Omega)_0)$ denotes the remainder when Ω is divided by the modulus c so that the quantity Ω used in the function $W(\Omega)$ or $s(\Omega)$ is defined in the range $0 \leq \Omega < 2\pi$.



(a) $t=0$



(b) $t=n \cdot T_s$

Fig. 4 Illustration of the ball bearing showing the relation between the surface roughness $s(x)$ and the equivalent driving impulse $z(t)$ as shown in Eq. (6b). (in general case) (a) at a time $t=0$. (b) at a time $t=n \cdot T_s$.

In the discrete time system, the equivalent driving impulse $z(n: \theta_0)$ described by the model in Eq. (6b) is represented as follows:

$$z(n: \theta_0)=\sum_{i=0}^{N-1} W(([\theta_d(n, i: \theta_0)])_M) \cdot s(([x_d(n, i: \theta_0)])_M), \tag{7a}$$

where

$\theta_d(n, i: \theta_0)=MF_b \cdot nT_s+Mi/N+M\theta_0/2\pi$,

$x_d(n, i: \theta_0)=M(F_b-F_i)\cdot nT_s+Mi/N+M\theta_0/2\pi$,

T_s: sampling period,

$2\pi F_b=\omega_b$, $2\pi F_i=\omega_i$,

$[\Omega]$ represents the integer next to the number Ω, and the integer M denotes the number of points on the race in the range of $0\sim 2\pi$.

By using the inverse filter $h(n)^{-1}$ of the resonant vibration, the equivalent driving impulse $z(n)$ is obtained from the discrete observed signal $y(n)$ as follows:

$$z(n)=y(n)*h(n)^{-1} . \tag{7b}$$

However, in the practical case, the signal $\underline{z}(n)$ obtained from the observed signal $y(n)$ must be used instead of $z(n: \theta_0)$ of Eq. (7a). Then, $\underline{z}(n)$ is represented as follows:

$$\begin{aligned}\underline{z}(n)&=y(n)*h(n)^{-1}\\&=z(n: \theta_0)+\Delta z(n: \theta_0) ,\end{aligned} \tag{7c}$$

where $\Delta z(n: \theta_0)$ is the error function. If the above model shown in Eq. (7a) just represents the signal $z(n)$, the error $\Delta z(n: \theta_0)$ is equal to zero. Therefore, the surface roughness $s(x)$ on the rolling ring is determined directly from the vibration signal $y(n)$ by applying a least mean square criterion to the error function $\Delta z(n: \theta_0)$. The total squared error $\alpha(\theta_0)$ is defined by

$$\begin{aligned}\alpha(\theta_0)&=\sum_{n=0}^{L-1}\Delta z(n: \theta_0)^2\\&=\sum_{n=0}^{L-1}|z(n)-z(n: \theta_0)|^2 ,\end{aligned} \tag{8}$$

where L defines the interval where the error minimization occurs. Here, the following two vectors and one matrix are introduced: $\boldsymbol{Z}$ and $\boldsymbol{S}$ are L-dimensional and M-dimensional vectors, respectively, whose transposes are given by

$$\boldsymbol{Z}^{\mathrm{T}}=[z(0),\ z(1),\ \ldots,\ z(L-1)]$$

and

$$\boldsymbol{S}^{\mathrm{T}}=[s(0),\ s(1),\ \ldots,\ s(M-1)] .$$

W is the L-by-M matrix as

$$W=\begin{pmatrix} w_{00} & w_{01} & \cdots & w_{0,M-1}\\ w_{10} & w_{11} & \cdots & w_{1,M-1}\\ \cdot & \cdot & \cdots & \cdot\\ w_{L-1,0} & w_{L-1,1} & \cdots & w_{L-1,M-1}\end{pmatrix},$$

where w_{nm} is equal to $W(([\theta_d(n,i: \theta_0)])_M)$ if the point m on the race contacts with the i-th ball at a time $t=n\cdot T_s$: otherwise w_{nm} is equal to zero.

By using these vectors and the matrix, the total squared error of Eq. (8) to be minimized is represented as follows:

$$\alpha(\theta_0)=(\boldsymbol{Z}-W\boldsymbol{S})^{\mathrm{T}}(\boldsymbol{Z}-W\boldsymbol{S}) . \tag{8'}$$

Minimization of $\alpha(\theta_0)$ is obtained by setting the partial derivatives of $\alpha(\theta_0)$ with respect to $s(m)$, $(m=0, 1, 2, \ldots, M-1)$ to zero as follows:

$$\frac{\partial\alpha(\theta_0)}{\partial s(m)}=0 . \quad (\text{for } m=0, 1, 2, \ldots, M-1)$$

Then, the set of M linear simultaneous equations are obtained as:

$$W^{\mathrm{T}}W\boldsymbol{S}=W^{\mathrm{T}}\boldsymbol{Z} . \tag{9}$$

When the length L of the interval for the error minimization is so long that every point on the race contacts with balls at least one time, the M-by-M matrix $W^{\mathrm{T}}W$ is nonsingular due to the following reasons:

(A) The points with which a revolving ball contacts on the race of the rolling ring vary every moment.

(B) The weight function $W(\theta)$ is expressed by a nonlinear function as shown below in Fig. 5.

Therefore, the surface roughness $\boldsymbol{S}$ is solved uniquely as follows:

$$\boldsymbol{S}=(W^{\mathrm{T}}W)^{-1}W^{\mathrm{T}}\boldsymbol{Z} , \tag{10}$$

where $(W^{\mathrm{T}}W)^{-1}$ denotes the inverse matrix of $W^{\mathrm{T}}W$.

4. THE PROCESS TO DETERMINE THE PARAMETERS USED IN THE PROPOSED METHOD

As described in Chap. 3.1, it is necessary to determine the following three parameters for the estimation of the surface roughness: (A) the weight function $W(\theta)$, (B) the values of the angular velocity ω_i and ω_b of the inner ring and balls, and (C) the characteristic of the inverse filter $h(n)^{-1}$. By using the two-pulse model proposed in the separate paper,[4] the poles of the resonant vibration $h(n)$ are accurately estimated from the vibration signal $y(n)$ and the inverse filter $h(n)^{-1}$ is determined from the poles. In the following paragraphs, the weight function $W(\theta)$ (in (A)) and the angular velocity (in (B)) are determined from the vibration signal.

(A) The weight function $W(\theta)$

It is difficult to measure the weight function $W(\theta)$ using the vibration caused by the flaw on the rolling ring because it is necessary to know (a) the exact location of the flaw and (b) the time when the flaw contacts with a ball. Therefore, by using the vibration due to a flaw on the outer ring, the weight function $W(\theta)$ is determined as follows: When the outer ring is fixed in an Anderon meter so that a flaw on the race is set at one of M_0 angles $\{k\cdot\theta=2\pi k/M_0\}$, $(k=0,1,2,\ldots,M_0-1)$ relative to the pick-up, the vibration signal $y(n:k)$ caused by the flaw is detected as explained in Chap. 2. The vibration $y(n:k)$ is band-limited so that the significant frequency band, by which the best detection of flaw is achieved, is selected.[3] The envelope signal is obtained by squaring the narrow band-passed signal and passing it through a low-pass filter. The periodicity of the vibration $y(n:k)$ caused by the flaw at the angle $k\cdot\theta$ is detected from the envelope signal. The magnitude $P(k)$ of the peak at the repetition frequency is calculated from the power spectrum of the vibration signal $y(n:k)$. By varying the angle $k\cdot\theta$ of the flaw relative to the pick-up, the magnitudes $\{P(k)\}$, $(k=0,1,2,\ldots,M_0-1)$ of the peaks are obtained from the above processes ($M_0=16$). Then, the squared weight function $|W(k)|^2$ is calculated from the peaks $\{P(k)\}$ divided by $P(0)$ as follows:

$$|W(k)|^2=P(k)/P(0)\,.$$
$$(\text{for } k=0,1,2,\ldots,M_0-1) \qquad (11)$$

By using the digital interpolation process[7] to increase the number M_0 of the points, the interpolated weight function $|W(k)|^2$, $(k=0,1,2,\ldots,M-1)$ is obtained at M points ($M_0<M=64$) on the circumference as shown in Fig. 5.

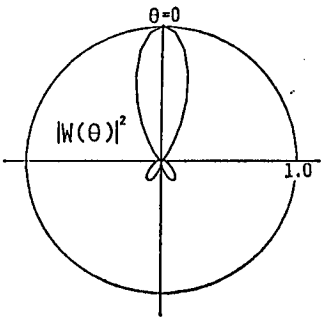


Fig. 5 The squared weight function $|W(\theta)|^2$ representing the amplitude ratio of the excited impulse at a point to the signal $z(t)=y(n)*h(n)^{-1}$, where the angle of the point relative to the pick-up ($\theta=0$) is θ.

(B) The angular velocity ω_i and ω_b

When the inner ring is revolved at a constant speed of $\omega_i=2\pi F_i$ [rad/s], the angular velocity $\omega_b=2\pi F_b$ of the balls may be theoretically derived from Eq. (1). Since the contact angle used in Eq. (1) depends on the amount of the axial pressure, it is advisable to determine the values of ω_i and ω_b directly from the vibration signal as follows: The power spectrum of the vibration signal $y(n)$ shows clear peaks at the frequencies around F_i and $N\cdot F_b$, even if the power spectrum is obtained from the vibration of the normal ball bearing having the smooth races. Therefore, accurate values of the rolling speed F_i and $N\cdot F_b$ are calculated from the power spectrum obtained using a time window of sufficient length. Then, by dividing $N\cdot F_b$ by the number N of the balls, the revolution speed F_b of the balls is obtained.

5. REDUCTION OF THE EFFECT OF THE IRREGULARITY IN THE BALL INTERVALS ON THE ESTIMATION OF THE SURFACE ROUGHNESS

5.1 The Effect of the Irregularity in the Ball Intervals on the Estimation of the Surface Roughness

It is known that there is scattering in the repetition intervals of the flaw pulses generated by many balls[8] due to the following reasons: (a) the diameter of the ball is a little smaller than that of the hole in the cage and (b) the intervals of the balls on the race vary

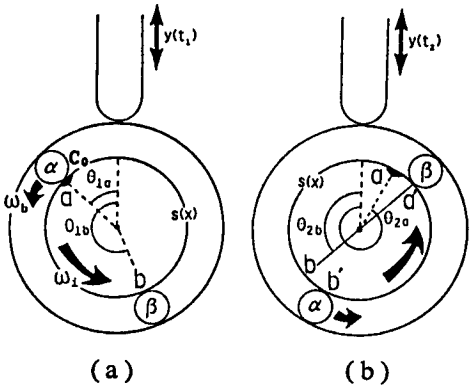


Fig. 6 Illustration of the simplified ball bearing showing the effect of the irregularity in the ball intervals on the estimation of the surface roughness. (a) at a time $t=t_1$. (b) at a time $t=t_2$.

$$z(t)=y(t)*h(t)^{-1}$$
$$\begin{cases} z(t_1)=W(\theta_{1a})\cdot s(a)+W(\theta_{1b})\cdot s(b)=W(\theta_{1a})\cdot C_0 \\ z(t_2)=W(\theta_{2a})\cdot s(a)+W(\theta_{2b})\cdot s(b)=0 \end{cases}$$

every moment due to the slipping of the balls in the cage.

Figure 6 shows the simplified ball bearing having two balls α and β and they are not π rad apart on the race. Figure 6 (a) shows that the balls α and β contact with the points a and b on the race, respectively, at a time $t=t_1$. Under the condition where the angle between α and β is π, the time t_2 when the balls α and β contact with the points b and a, respectively, is calculated from Eq. (7a). However, since the two balls are not π rad apart, the balls α and β do not contact with the points b and a, respectively; at a time $t=t_2$ as shown in Fig. 6 (b). Even if the accurate values of the angular velocity ω_t and ω_b are obtained, the above condition makes it impossible to estimate the surface roughness accurately as follows:

Let the times when the balls α and β contact again with the points b and a be $t=t_2+\Delta t_b$ and $t=t_2+\Delta t_a$, respectively: then, the equivalent driving impulse $z(n)$ of Eq. (4') is modified in the discrete-time system as follows:

$$z(n)=\sum_{x=a,b}\{W(\theta_{1x})\cdot s(x)\cdot\delta(nT_s-t_1) + W(\theta_{2x})\cdot s(x)\cdot\delta(nT_s-t_2-\Delta t_x)\}. \quad (12)$$

Let the points with which the balls α and β contact at a time $t=t_2$ be a' and b', respectively, as shown in Fig. 6 (b). Since the weight function quantities $W(\theta_{2a})$ and $W(\theta_{2b})$ are very close to $W(\theta_{2a'})$ and $W(\theta_{2b'})$, respectively, when the irregularity in the intervals of the balls is slight, the equivalent driving impulse $z(n)$ of Eq. (12) is approximated using the weight function $W(\theta_{2x'})$ of the contact points at the calculated collision time $t=t_2$ as follows:

$$z(n)\simeq\sum_{x=a,b}\{W(\theta_{1x})\cdot s(x)\cdot\delta(nT_s-t_1) + W(\theta_{2x'})\cdot s(x)\cdot\delta(nT_s-t_2-\Delta t_x)\}. \quad (12')$$

When the inner race has a flaw only on the point a as shown in Fig. 6, the surface roughness $s(x)$ is described as:

$$s(x)=\begin{cases}C_0 & (x=a)\\ 0 & (x\neq a),\end{cases} \quad (13)$$

where the positive constant C_0 denotes the 6/5-th power of the width of flaw on the point a. Thus, $s(a')$ is equal to zero and the equivalent driving impulses $z(t_1)$ and $z(t_2)$ of Eq. (5) are expressed as follows:

$$\begin{cases}z(t_1)=W(\theta_{1a})\cdot C_0\\ z(t_2)=0.\end{cases} \quad (14)$$

Using these values, the surface roughness quantities $s(a)$ and $s(b)$ are estimated from the set of simultaneous Eqs. (5a) and (5b) as follows:

$$\underline{s}(a)=W(\theta_{1a})\cdot W(\theta_{2b})\cdot C_0/D$$

and

$$\underline{s}(b)=-W(\theta_{1a})\cdot W(\theta_{2a})\cdot C_0/D, \quad (15)$$

where

$$D=W(\theta_{1a})\cdot W(\theta_{2b})-W(\theta_{2a})\cdot W(\theta_{1b}),$$

and $\underline{s}(a)$ and $\underline{s}(b)$ are the estimated values of $s(a)$ and $s(b)$, respectively.

Thus, the true value of the surface roughness expressed by Eq. (13) cannot be obtained due to the irregularity of the ball intervals. Even if the vibration signal $y(t)$ or $z(t)$ is not corrupted by additive noise and the least mean square fitting technique is used, the surface roughness is not obtained accurately because of the irregularity in the intervals of the balls. Therefore, it is necessary to decrease the sampling rate to reduce the effects of the irregularity, as described in the following two methods.

5.2 Decrease of the Sampling Rate Using the Digital Heterodyne Method

The digital heterodyne method decrease the sampling rate by the following processes[8]:

(1) The narrow band signal $z_B(n)$ is obtained from the signal $z(n)$ using the narrow band filter $h_B(n)$. The band-pass filter $h_B(n)$ is described by the product of the low-pass filter $h_L(n)$ having the cut off frequency ΔF and the cosine signal having the frequency component of F_0 as follows:

$$h_B(n)=h_L(nT_s)\cdot\cos(2\pi F_0\cdot nT_s). \quad (16)$$

(2) To make the signal $z'(n)$, the narrow band signal is multiplied by the cosine wave of frequency F_0, $(F_0<F_0-\Delta F)$.

(3) Only the low frequency components involved in the signal $z'(n)$ calculated in the above process are separated by the low-pass filter.

(4) Since the obtained signal has frequency components only in the range less than $F_0+\Delta F-F_0$, the sampling frequency can be decreased to $F_0+\Delta F-F_0$.

However, this method does not reduce the effects of the irregularity of the ball intervals as described in the following: Using the impulse $z(n)$ of Eq. (12'), the signal $z'(n)$ obtained from the second process above is described as follows:

$$z'(n)=\{z(n)*h_B(n)\}\cdot\cos(2\pi F_0\cdot nT_s)$$
$$=\sum_{x=a,b}[W(\theta_{1x})\cdot s(x)\cdot h_L(nT_s-t_1)\cdot\cos\{2\pi F_0(nT_s-t_1)\}+W(\theta_{2x'})\cdot s(x)\cdot h_L(nT_s-t_2-\Delta t_x)\cdot\cos\{2\pi F_0(nT_s-t_2-\Delta t_x)\}]\cdot\cos(2\pi F_0\cdot nT_s)\,. \quad (17)$$

Using the formula

$$\cos(\Omega_1)\cdot\cos(\Omega_2)=\{\cos(\Omega_1-\Omega_2)+\cos(\Omega_1+\Omega_2)\}/2\,,$$

the signal $z'(n)$ is decomposed into the low frequency components and the high frequency components as follows:

$$z'(n)=\sum_{x=a,b}W(\theta_{1x})\cdot s(x)\cdot h_L(nT_s-t_1)\cdot\cos\{2\pi((F_0-F_0)nT_s-F_0t_1)\}/2+\sum_{x=a,b}W(\theta_{2x'})\cdot s(x)\cdot h_L(nT_s-t_2-\Delta t_x)\cdot\cos\{2\pi((F_0-F_0)nT_s-F_0(t_2+\Delta t_x))\}/2+[\text{high frequency components around } (F_0+F_0)\text{ Hz}]\,. \quad (18)$$

In the third process described above, the signal obtained after passing though the low-pass filter has only the first and the second components of Eq. (18). The second component involves the fluctuation terms Δt_a and Δt_b, which are multiplied by $2\pi F_0$. Therefore, the effect of the irregularity in the ball intervals cannot be reduced by the digital heterodyne method. Thus, the surface roughness cannot be estimated accurately using the method described above.

5.3 Estimation of the Surface Roughness Based on the Envelope Signal of the Squared Magnitude of the Narrow Band Signal

To reduce the effect of the irregularity in the ball intervals, we propose a new, reliable method. The input signal $z(n)$ of Eq. (12′) is band-limited. Then, the envelope signal $z_L(n)$ is obtained by squaring the narrow band-passed signal $z_B(n)$ and passing it through a low-pass filter. The surface roughness is estimated from the envelope signal $z_L(n)$. The details of this processing are explained below.

(1) The signal $z(n)$ of Eq. (12′) is band-limited using a band-pass filter described by Eq. (16). When the length of the impulse response of $h_B(n)$ is short enough, the band-passed signal $z_B(n)$ around t_2 is obtained by neglecting the components occurring around t_1 as follows:

$$z_B(n)=\sum_{x=a,b}W(\theta_{2x'})\cdot s(x)\cdot h_L(nT_s-t_2-\Delta t_x)\cdot\cos\{2\pi F_0(nT_s-t_2-\Delta t_x)\}\,. \quad (\text{for } nT_s \text{ around } t_2) \quad (19)$$

(2) The squared signal $|z_B(n)|^2$ of the band-passed signal $z_B(n)$ is calculated.

(3) The envelope signal $z_L(n)$ is obtained by passing the squared signal through a low-pass filter as follows:

$$z_L(n)=[\text{low frequency components of } |z_B(n)|^2]$$
$$=\sum_{x=a,b}\{W(\theta_{2x'})\cdot s(x)\cdot h_L(nT_s-t_2-\Delta t_x)\}^2/2+W(\theta_{2a'})\cdot W(\theta_{2b'})\cdot s(a)\cdot s(b)\cdot\cos\{2\pi F_0(\Delta t_a-\Delta t_b)\}\cdot h_L(nT_s-t_2-\Delta t_a)\cdot h_L(nT_s-t_2-\Delta t_b)\,. \quad (\text{for } nT_s \text{ around } t_2) \quad (20)$$

In this equation, the envelope signal $z_L(n)$ has two fluctuation terms: (a), the time lag Δt_x involved in the first term of the low-passed signal, and (b), the constant $\cos(2\pi F_0(\Delta t_a-\Delta t_b))$ involved in the second term. However, the effects of both the terms can be disregarded due to the following reasons:

(a) Since the low-passed signal does not involve the high frequency components, the phase shift due to the time lag Δt_x in the first term is negligible.

(b) Since the central frequency F_0 of the band-pass filter is high, $2\pi F_0\cdot(\Delta t_a-\Delta t_b)$ in the second term takes a large value. However, when a ball bearing has N balls, there are a number of combinations of the term in the second term because each of the balls contacts with the point on the raceway. Since there is no correlation between Δt_x and Δt_v $(x\neq v)$ because the irregularity of the ball intervals is generated by the slipping of the balls in the cage, the constant $\cos\{2\pi F_0|\Delta t_x-\Delta t_v|\}$ $(x\neq v)$ in the second term takes a random value in the range of $-1\sim+1$. Then, the component value of the second term is smaller than that of the first term. Therefore, the low frequency component of the squared magnitude of the narrow band signal is expressed as follows:

$$z_L(n)\simeq h_L(nT_s-t_2)^2\sum_x|W(\theta_{2x'})|^2\cdot s(x)^2/2\,. \quad (\text{for } nT_s \text{ around } t_2) \quad (21)$$

From Eq. (21), the envelope signal $z_L(n)$ denotes the sum of the squared signal of the impulses generated by N balls. Equation (21) is the same as Eq. (7a) except for the squared low-pass filter $|h_L(n)|^2$; that is, $(|h_L(n)|^2\cdot|W(\theta)|^2/2)$ and $|s(x)|^2$ are used instead of $W(\theta)$ and $s(x)$, respectively.

Thus, by neglecting the characteristic of the low-pass filter, the squared surface roughness $|s(x)|^2$ is estimated accurately from the envelope signal $z_L(n)$ of the squared magnitude of the narrow band signal. The squared surface roughness $|s(x)|^2$ obtained here is in proportion to the $(6/5)^2$-th power of the width of the flaw on the race as described in Chap. 3.1.

6. EXPERIMENTS AND THE RESULTS

To estimate the surface roughness based on the proposed method in the practical case, the following processes are carried out in the following sequences:

(1) The squared weight function $|W(k)|^2$, $(k=0, 1, \ldots, M-1)$ is obtained at M points on the race $(M=64)$ by the method described in Chap. 4.1.

(2) The inverse filter $h(n)^{-1}$ is calculated from the impulse response of the resonant vibration estimated using the two-pulse model.[4] The equivalent driving impulse $z(n)$ is obtained by convoluting $h(n)^{-1}$ with the observed vibration signal $y(n)$.

(3) The signal $z(n)$ is band-limited by using the FIR narrow-band pass filter (16 points), and the narrow-band signal $z_B(n)$ having the components around the central frequency of the resonant vibration is calculated.[3] The band-limited signal $z_B(n)$ is squared and moving-averaged using a 16 point Hanning window so that only low frequency components are extracted. Then, an envelope signal $z_L(n)$ is generated by picking up one sample every third segment.

(4) The squared surface roughness $|s(m)|^2$ of Eq. (21) is estimated from the envelope signal $z_L(n)$ using the squared weight function $|W(k)|^2$ and Eq. (7a).

(5) The unknown value of the initial angle θ_0 of Eq. (7a) is decided as follows: Since the initial angle θ_0 has a value in the range of $0 \leq \theta_0 < 2\pi/N$, where N denotes the number of ball, the total squared error $\alpha(\theta_0)$ in Eq. (8) is calculated for various θ_0 within the range. The optimum angle of θ_0 is decided so that the squared error $\alpha(\theta_0)$ is minimum. Using the optimum angle of θ_0, the surface roughness is obtained.

Figure 7 shows the experimental results obtained by the proposed method for a ball bearing (JIS696) having flaws on the inner ring. Figure 7 (a) shows the envelope signal $z_L(n)$ calculated from the vibration signal. The squared surface roughness $|s(m)|^2$, $(m=0, 1, 2, \ldots, 63)$ estimated from the envelope signal is shown in Fig. 7 (b). The two main flaws are observed in this figure in the direction of 40° and 350°.

By using the estimated surface roughness $|s(m)|^2$ in Fig. 7 (b), the weight function $|W(k)|^2$ in Fig. 5 and initial angle θ_0, the envelope signal $\underline{z}_L(n)$ of the vibration is synthesized based on Eq. (7b) as shown in Fig. 7 (c). Since the synthesized envelope signal $\underline{z}_L(n)$ just coincides with the envelope signal $z_L(n)$ shown in Fig. 7 (a), it is found that the surface roughness is estimated accurately. The synthesized envelope signal $\underline{z}_L(n)$ obtained by the proposed method is evaluated by the following normalized mean square error (NMSE) as

$$\mathrm{NMSE}=\frac{\sum_n (z_L(n)-\underline{z}_L(n))^2}{\sum_n z_L(n)^2}, \qquad (22)$$

and we find that the value of the NMSE for the above synthesized vibration is about 0.73% (−21 dB). Figure 7 (d) shows the surface texture of the

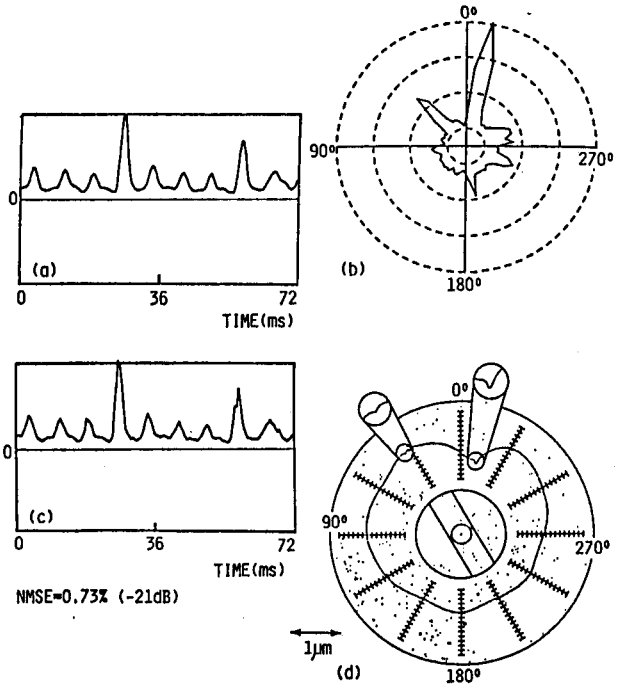


Fig. 7 Experimental results of the estimated surface roughness, the synthesized vibration signal and the surface texture measured by a stylus. (a) the envelope signal $z_L(n)$ of the vibration signal $y(n)$ generated by flaws on the inner ring. (b) estimated surface roughness $s(m)$, $(m=0, 1, 2, \ldots, 63)$ from the envelope signal $z_L(n)$. (c) the envelope signal $\underline{z}_L(n)$ synthesized from the estimated surface roughness $s(m)$, the weight function $|W(\theta)|^2$ and the initial angle θ_0. (d) the profile of the surface texture measured using a stylus. (cross-sectional view of the inner ring of the samples used above).

inner race measured directly by a stylus. The positions of the two main flaws in Fig. 7 (d) agree with those of the two main flaws in Fig. 7 (b).

7. SORTING OF BALL BEARINGS

Figure 8 shows the distributions of the maximum values of the squared surface roughness $|s(m)|^2$ estimated from the vibration of 75 ball bearings. At first, these ball bearings were classified into the following three categories aurally by an inspector: 15 normal bearings, 41 ball bearings having flaws on the inner race, and 19 ball bearings having rough races. Table 1 shows the results of sorting the ball bearings using the threshold shown in Fig. 8. The number of correctly sorted samples and that of incorrectly sorted samples are 171 and 4, respectively. The ball bearings are sorted correctly at the rate of 94.7%.

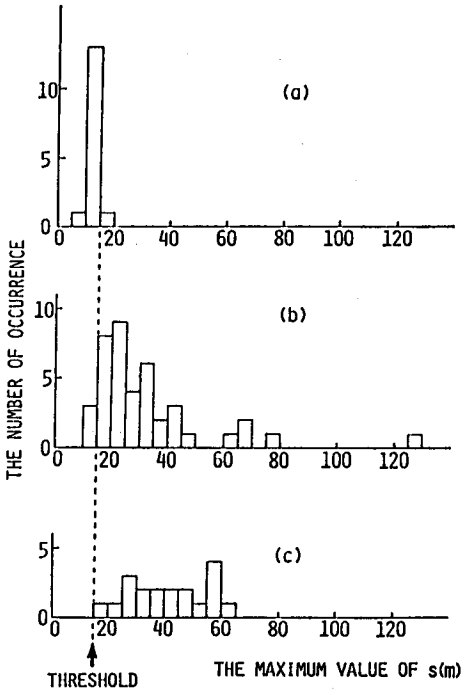


Fig. 8 The distribution of the maximum values of the estimated surface roughness $s(x)$ and threshold for sorting bearing samples. (a) 15 normal samples. (b) 41 samples having flaws on the inner races. (c) 19 samples having rough races.

Table 1 Recognition results of ball bearings.

In	Out: The number of samples recognized to be normal	Out: The number of samples recognized to be inferior	Total
Normal	14	1	15
Flaw on the inner ring	3	38	41
Rough race	0	19	19

Total number of used samples=75, recognition rate=94.7%.

8. CONCLUSIONS

A new method, utilizing vibration signals, is proposed for the estimation of the surface roughness of the races of the rolling ring in ball bearings. The degree of the surface roughness of the races and the number and the width of the main flaws on the races are obtained from the estimated surface roughness. It is confirmed with the experimental results presented herein that the roughness estimated by the proposed method agrees with that measured directly by using a stylus. Based on the estimated surface roughness, ball bearings having inferior races are detected by the proposed method with a 94.7% accuracy rate.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. Akihiro Yuasa, the graduate student of Tohoku University, for his contribution to the measurements of the squared weight function $|W(\theta)|^2$ in Fig. 5.

REFERENCES

1) T. Igarashi, "Sound of rolling bearings," Lubrication **22**, 751–756 (1977) (in Japanese).
2) S. Braun and B. Datner, "Analysis of roller/ball bearing vibrations," Trans. Am. Soc. Mech. Eng., J. Mech. Design. **101**, 118–125 (1979).
3) H. Kanai, M. Abe, and K. Kido, "Detection and discrimination of flaws in ball bearings by vibration analysis," J. Acoust. Soc. Jpn. (E) **7**, 121–131 (1986).
4) H. Kanai, M. Abe, and K. Kido, "Detection of slight defects in ball bearings by non-periodic analysis," J. Acoust. Soc. Jpn. (E) **7**, 219–228 (1986).
5) G. Nishimura and K. Takahashi, "Ball bearing noise," J. Jpn. Soc. Precis. Eng. **30**, 475–489 (1964) (in Japanese).
6) M. Noda, "A study of the inferior foresight in roller/ball bearing," Proc. Spring Meet. Lubr. Soc. Jpn. **25**, 125–128 (1981) (in Japanese).
7) L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signal* (Prentice-Hall, Englewood Cliffs, 1978), Chap. 2.4.
8) Y. Takebayashi, T. Itoh, and K. Kido, "Measurement of transfer function characteristics of the system with long reverberation (Application of narrowband digital processing)," J. Acoust. Soc. Jpn. **35**, 595–601 (1979) (in Japanese).