Non-invasive Monitoring of Knee Pathology based on Automatic Knee Sound Classification

C. G. Song, K. S. Kim, and J. H. Seo

Abstract—Knee joint sounds, generated during the active flexion and extension of the leg, represent acoustic signals caused by joint vibration and can be used as useful indicators of the roughness, softening, breakdown, or the state of lubrication of the articular cartilage surface. This paper describes an efficient algorithm in order to improve the classification accuracy of the features obtained from the time-frequency distribution (TFD) of normal and abnormal knee sounds. Knee sounds were correctly segmented by the dynamic time warping and the noise within the TFD of the segmented knee sounds was diminished by the singular value decomposition method. The classification of the knees as normal or abnormal was evaluated using a backpropagation neural network (BPNN). 1408 knee sound segments (normal 1031, abnormal 377) were used for evaluating our devised algorithm by a BPNN and, consequently, the mean accuracy was 91.4±1.7%. This algorithm could help to enhance the performance of the feature extraction and classification of knee sounds.

Index terms -knee pathology, back-propagation neural network; singular value decomposition; dynamic time warping

I. INTRODUCTION

Non-invasive techniques such as X-ray imaging, computed tomography (CT) and magnetic resonance imaging (MRI) [1-2] and invasive techniques such as arthroscopy [3] have been widely used as a diagnostic tool in order to detect articular pathologies. Arthroscopy, where the cartilage surface is inspected with a fiber-optic cable, has emerged as the "gold-standard" for the relatively low-risk assessment of joint surfaces, in order to determine the prognosis and treatment for a variety of conditions. However, image-based techniques could provide only anatomical images of the joint cartilage and, consequently, they failed to characterize the functional integrity of the cartilage, in terms of its softening, stiffness, or fissuring [4]. Also, these methods may be limited to the early detection of cartilage pathologies [5]. Therefore, the drawbacks and limitations of imaging techniques have motivated researchers to look for alternative tools such as vibroarthrography (VAG).

VAG signals, generated during the active flexion and extension of the leg, represent acoustic signals caused by joint vibration and are considered to be associated with pathological conditions, such as the degeneration of the cartilage surface. Also, it is known that the acoustical characteristics of the VAG signals of a patient suffering from cartilage degeneration or a ruptured anterior cruciate ligament (ACL) can differ from those of a normal subject. Therefore, they can be used as useful indicators of the roughness, softening, breakdown, or the state of lubrication of the articular cartilage surface [6]. The detection of knee joint problems via the analysis of VAG signals could help avoid unnecessary exploratory surgery, as well as enabling the better selection of patients who would benefit from surgery.

Several researchers have continued to classify the VAG signals of normal and abnormal knees according to the pathological conditions using modeling techniques such as autoregressive [7], least square [8] and linear prediction modeling [9]. Besides, various techniques such as time-frequency analysis [10], wavelet decomposition [11] and the non-linear strict 2-surface proximal classifier based on statistical parameters, including the skewness, kurtosis and entropy [12], have been tested.

The aim of this paper is to improve the classification accuracy of the features obtained from the time-frequency distribution (TFD) of normal and abnormal VAG signals using segmentation by the dynamic time warping (DTW) and denoising algorithm by the singular value decomposition (SVD) and to evaluate the performance of our algorithm using an artificial neural network (ANN). The SVD algorithm was applied in order to divide the TFD of the VAG signals into interesting and noise subspaces and to remove the noise component. Also, we determined the optimum learning rate and the number of hidden nodes of an ANN, which can provide the best classification accuracy.

II. MATERIALS AND METHODS

A. Participants

Thirty one subjects were examined. Twenty subjects, having no anamnesis on either the knee or femoral joint, were enrolled as the normal group (7 males and 13 females, age: 33.3±10.6 years). Eleven patients (7 males and 4 females, age: 39.5±13.2 years), diagnosed as having degenerative arthritis by physical examination and MRI scan, were enrolled as the abnormal group. Patients with a surgical history such as total knee replacement or myopathy and neuromuscular disorders of the lower limbs, such as local spasticity, were excluded.
For a given signal, joint representation in both the time and frequency domains. the Wigner-Ville distribution (WVD). The TFD of a signal is a TFD. By setting dimensional kernel that determines the characteristics of the frequency movement artifacts and prevents aliasing effects.

(BPF), whose cut-off frequency was between 20 and 2000 Hz, MCI, were removed by a 4th-order Butterworth band-pass filter low frequency components, caused by motion artifacts and characteristics of the VAG signals. Thirdly, the harmonic and velocity of knee movement could affect the frequency in order to minimize the possibility that the variation of the signals was normalized using the DTW algorithm at 2 sec/cycle

A. Segmentation, Preprocessing and Transformation

The VAG signals were transferred to a PC through an analog-to-digital converter (MP-100™, Biopac system, U.S.) for storage. The data were then digitized with a sampling rate of 6 KHz and 12 bits/sample. The electro-goniometer used to measure the knee joint angle was placed on the lateral aspect of the patella with the axis of rotation at the joint line.

The contact area selected for auscultation was the medial condyle on the patella, in order to take into consideration the physiological structure and position which can remove the muscle contraction interference (MCI) [13]. A motion from full flexion to full extension is defined as ‘knee extension’, whereas ‘knee flexion’ is defined as a motion from full extension to full flexion. ‘Full flexion’ means an angle of 90° between the femur and tibia, while ‘full extension’ means an angle of 0°. One cycle consisted of one flexion and one extension over an approximate angle range of 90°→0°→90° in 2 seconds.

Each subject sat on a rigid chair in a relaxed position with the leg being tested freely. In the sitting position, each subject underwent active ‘knee flexion’ and ‘extension’ for 20 seconds, while keeping a constant velocity of 2 sec/cycle using a metronome. Measurements were performed three times at the same position with at least 10 minutes rest between sessions in order to avoid muscle fatigue.

III. FEATURE EXTRACTION

A. Segmentation, Preprocessing and Transformation

Fig. 1 shows the procedure used for extracting the features, in order to classify the VAG signals into the normal and abnormal groups using the ANN. Firstly, the start and end points of the VAG signals during one cycle of knee movement were detected for the purpose of segmentation using the knee angle signals recorded simultaneously by an electro-goniometer. Secondly, the time axis of the segmented VAG signals was normalized using the DTW algorithm at 2 sec/cycle in order to minimize the possibility that the variation of the velocity of knee movement could affect the frequency characteristics of the VAG signals. Thirdly, the harmonic and low frequency components, caused by motion artifacts and MCI, were removed by a 4th-order Butterworth band-pass filter (BPF), whose cut-off frequency was between 20 and 2000 Hz, with the transfer function as Eq. (1). The BPF minimizes low-frequency movement artifacts and prevents aliasing effects.

The Wigner-Ville distribution (WVD) of a signal can be expressed as Eq. (2), where \( g(\tau, \nu) = \int_{t} e^{i 2 \pi f t} x(t) x^*(t) e^{-i 2 \pi (\nu - \tau) t} dt \). The filtered VAG signals were transformed into the TFD by the Wigner-Ville distribution (WVD). The TFD of a signal is a joint representation in both the time and frequency domains. For a given signal, \( x(t) \), the TFD that belongs to the quadratic class can be expressed as Eq. (2), where \( g(\tau, \nu) \) is a two-dimensional kernel that determines the characteristics of the TFD. By setting \( g(\tau, \nu) = 1 \) we get the WVD.

\[
X(t, f) = \int_{t} e^{i 2 \pi f t} x(t) x^*(t) e^{-i 2 \pi (\nu - \tau) t} dt
\]

B. Denoising in TFD

In the case of VAG signals, relatively larger ambient noises could affect the acoustical characteristics of the VAG signals and cause the noise to be confused with the inherent VAG signals, due to the very low amplitude of the latter. In previous studies, adaptive filtering [9] has been widely used for reducing noise component such as MCI, mixed together with inherent VAG signals. However, this method needs the supplementary recording of the vibration of the muscles which are located around auscultation area. Moreover, the features of MCI-filtered VAG signals did not increase the classification accuracy significantly compared with the classification accuracy obtained using the features of the original VAG signals [14]. Therefore, in this study, the SVD algorithm was applied to the TFD in order to separate the interesting signals from noisy VAG signals and remove noise unnecessarily mixed within the TFD, instead of adaptive filtering.

A singular value decomposition of an M x N matrix \( X \) is of the form, where \( U \) (M x M) and \( V \) (N x N) are orthogonal matrices, and \( \Sigma \) is an M x N diagonal matrix of singular values (SVs) with various components.

\[
X = U \Sigma V^T
\]

In Eq. (3), \( X \) can be divided into two subspaces, a strong interference subspace, \( X_i \), and an orthogonal alternate subspace, \( X_a \). The strong interference subspace contains the signal of interest, whereas the alternate subspace contains the ambient noise. Mathematically, the subspace separation can be expressed as follows;

\[
X = U \Sigma V^T = (U_i, U_a) \begin{bmatrix} \Sigma_i & 0 \\ 0 & \Sigma_a \end{bmatrix} (V_i^T, V_a^T)
\]

In order to divide the matrix \( X \) into the signal of interest, \( X_i \) and noise, \( X_a \), the rank of the matrix \( U \) or \( V \) has to be prescribed. In this paper, the rank of the matrix (denoted by \( r \)) was estimated from the output of the derivative of the normalized SVs with respect to the rank order [15]. Finding the \( r \) using the derivative reduces the effect of space intersections on altering the structure of important information in the signal. Given that the derivative is smaller than a predefined threshold, the SVs are considered to be members of \( \Sigma_i \). (The way to find the optimal threshold was explained in next section)
The separated interest matrix, $X_i$, can be expressed as Eq. (5), and the small amplitude fluctuations of the SVs of matrices $U_i$ and $V_i$ were reduced using an 3rd-order polynomial Savizky-Golay smoothing filter, $(U_i'; V_i')$, which frame size was 15.

$$X_i = U_i \Sigma_i V_i^T \quad (5)$$

The enhanced signal matrix, $X_e$, was obtained by multiplying together, $U_e$, $V_e$ and $\Sigma_e$, as in Eq. (6), where the matrices $U_e$, $V_e$ and $\Sigma_e$ are the normalized versions of $U_i'$, $V_i'$ and $\Sigma_i$, respectively.

$$X_e = U_e \Sigma_e V_e^T, \quad (6)$$

C. Noise Attenuation of the Enhanced TFD

It has been known that low frequencies below 500 Hz are clinically more important among all frequencies and they are clinically useful for diagnosis of pathological conditions [6]. So, it is important to minimize effectively the noise mixed within low frequencies band as keeping the interest signals. The rate of changes in SV is gradually reduced toward the larger index numbers. Also, values on lower index numbers have more information about the signal of interest. However, selecting excessively lower index could lead to the data loss, even more. Therefore, we determined the optimal threshold as 20 % of the minimum derivative of the normalized SVs (SNR20%). Fig. 2 shows a comparison of the raw TFD of the VAG signals with the enhanced TFD obtained by the SVD technique.

![Figure 2. Denoising in TFD using SVD algorithm](image)

D. Feature Extraction and Classification

The four time-frequency parameters of the enhanced matrix, $X_e$, which are derived from the TFD of the VAG signals, are the energy parameter (EP), energy spread parameter (ESP), frequency parameter (FP) and frequency spread parameter (FSP), as described by Eqs. (7) ~ (10), respectively [12], where $f_m$ is the maximum frequency present in the signal.

$$EP(t) = \frac{\sum_{f=a}^{f_m} X_e(t, f)}{f_m} \quad (7)$$

$$FP(t) = \frac{\sum_{f=a}^{f_m} |X_e(t, f)|}{\sum_{f=a}^{f_m} X_e(t, f)} \quad (8)$$

$$ESP(t) = \left[ \frac{\sum_{f=a}^{f_m} X_e(t, f) - EP(t)}{f_m} \right]^{\frac{1}{2}} \quad (9)$$

$$FSP(t) = \left[ \frac{\sum_{f=a}^{f_m} |X_e(t, f)| - FP(t)}{\sum_{f=a}^{f_m} X_e(t, f)} \right]^{\frac{1}{2}} \quad (10)$$

The four parameter arrays are dependent on the functional state of the cartilage surfaces in the knee joint, and are thought to be suitable for discriminating pathological knees from normal knees. The average (mEP, mESP, mFP, mFSP) and standard deviation values (sEP, sESP, sFP, sFSP) of each parameter, described above, were used as the features of the

Table 1 shows the comparison of the SNRs of the TFD enhanced by the SVD algorithm with those of noisy TFD, obtained from the raw VAG signals. SNRs increased gradually until the threshold of 20 % (SNR20%), but after that, they were saturated at about 52 dB. Also, the SNRs (51.8 ± 43.3 dB) of the TFD enhanced by the SVD algorithm at SNR20% were relatively 38.3% higher than those of noisy TFD (37.4 ± 29.2 dB) (p<0.01). Therefore, we determined the optimal threshold as 20 % of the minimum derivative of the normalized SVs (SNR20%).

<table>
<thead>
<tr>
<th>Noisy TFD (dB)</th>
<th>Enhanced TFD (dB)</th>
<th>(%)</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>37.4 ± 29.2</td>
<td>47.0 ± 39.5</td>
<td>125.6</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>49.9 ± 42.1</td>
<td>51.8 ± 43.3</td>
<td>133.1</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>51.8 ± 43.7</td>
<td>52.2 ± 43.8</td>
<td>138.3</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>52.3 ± 43.8</td>
<td>139.3</td>
<td>139.7</td>
<td>p&lt;0.01</td>
</tr>
</tbody>
</table>

\(\%\) : (SNRs of enhanced TFD) / (those of noisy TFD) * 100

\(p\)-value : comparison of the SNRs of enhanced TFD with those of noisy TFD

Values are mean ± standard deviation
input vector for the ANN. Each feature was normalized between -1 and 1.

The classification of the knees as normal or abnormal was evaluated using a statistical pattern classifier, back-propagation neural network (BPNN). The BPNN used in this study was a multi-layer shape, input-hidden-output layer BPNN, and the input and output layers of the BPNN consisted of 9 nodes (8 features and 1 bias) and 1 node, respectively. The VAG signals were classified into the normal group in the case of a positive output from the output node or the abnormal group in the case of a zero or negative value. The classification accuracy of the BPNN can depend on the learning rate ($\alpha$) and the number of hidden nodes ($p$). Therefore, in this paper, we determined the optimum values of $\alpha$ and $p$ which can provide the maximum accuracy. The values of $\alpha$ tested were 1, 0.9, … , 0.1, 0.05 and 0.01, and the values of $p$ were 5, 10, … , 30 and 35.

A total of 1408 VAG segments were used for training and testing the BPNN, where the numbers of VAG segments obtained from the normal and patient subjects were 1031 and 377, respectively. Among the normal segments, only 377 segments were selected after the random rearrangement because this different number can potentially skew the results. Then, the performance of our algorithm was evaluated by k-fold cross validation method as follows; after the random rearrangement of all of the VAG segments, 80 percent of them (about 603 segments) were used for training and the remaining 20 % (about 151 segments) were used for testing the BPNN. The mean and maximum values of the classification accuracy were calculated after repeating all procedures of rearranging, training and testing 20 times.

In order to determine the optimal structure of the BPNN, we compared the accuracies according to the various learning rate and the number of hidden nodes of the BPNN. As a result, the optimum learning rate and number of hidden nodes were 0.3 and 15, respectively, at the maximum and mean accuracies of 95.4 and 91.4 (S.D. ±1.7) %, respectively. Table 2 shows the maximum classification result of 151 VAG segments when applying these optimum values to the designed BPNN.

![Figure 3. Schematic diagram of the BPNN used for classification of normal and abnormal VAG signals](image)

### IV. RESULTS

#### V. DISCUSSIONS

The VAG signals were recorded by an electro-stethoscope in this study, because VAG signals are sensitive and minute. However, because unwanted noise, which could be generated in the contact area between the skin and the probe of the stethoscope, might interfere with the inherent VAG signals. So, in this study, we removed the noise mixed with the TFD using the SVD algorithm, especially spread over low frequency band, because low frequencies are clinically useful among all frequencies for diagnosis of pathological conditions as reported in [6]. Consequently, the SNRs (51.8 ± 43.3 dB) of the TFD enhanced by the SVD algorithm were relatively 38.3% higher than those (37.4 ± 29.2 dB) of the raw TFD, hence, it was noticed that clinically useful information in low frequency band was increased while the noise distributed evenly among the whole frequency band was decreased by our developed algorithm.

In preceding papers about the detection of VAG signals, most researchers prescribed the protocol only related to the range of knee movement without the velocity. However, this protocol had a serious problem in that they did not consider the physiological properties of the VAG signals due to the change of the muscle and cartilage, such as muscle fatigue during repetitive flexion and extension of the knee. Moreover, it could be difficult to separate correctly the VAG signals, generated during only one knee movement, from the whole VAG waveform, because there was no information about the knee joint angle. To minimize the possibility that the variation of the velocity of knee movement could affect the frequency characteristics of the VAG signals, we attempted to keep a constant velocity by normalizing the period of one cycle of knee movement using the DTW algorithm. Consequently, it is thought that our proposed method showed the relatively higher accuracies because the acoustical properties of normal and abnormal VAG signals such as the ‘clicking’, ‘creaking’ and ‘crepitus’ were correctly reflected to the features of the BPNN.

In the previous studies conducted since the 1990s, a classification accuracy of 75.6 % was obtained by logistics regression analysis using the cepstrum coefficients (CC-LRA), 68.9 % by TF analysis (TFD-MP), 76.4 % by linear discriminant analysis using wavelet decomposition algorithm (WD-LDA), 74.2 % by the linear strict 2-surface proximal (GA-LS2SP) classifier using a genetic algorithm and 91.0 % by non-linear S2SP (GA-NS2SP) [12]. On the other hand, our results revealed a relatively higher accuracy than these methods with a maximum and average classification accuracy of 95.4 %

<table>
<thead>
<tr>
<th>Actual group</th>
<th>Predicted group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Abnormal</td>
</tr>
<tr>
<td>Normal</td>
<td>94.3 ± 3.7 %</td>
</tr>
<tr>
<td>Abnormal</td>
<td>11.6 ± 4.4 %</td>
</tr>
<tr>
<td>Total</td>
<td>Accuracy : 91.4 ± 1.7 %</td>
</tr>
</tbody>
</table>

Table 2 The Means and Standard Deviations of the Classification Accuracies of the Normal and Abnormal VAG Segments using 8 Features per Signal
and 91.4 ± 1.7 %, respectively. Although it is difficult to compare the previous results with those obtained by our method, because of the absence of a standard protocol for recording VAG signals or authorized VAG database, we believe that our proposed method can enhance the classification accuracy and the higher accuracy resulted from our study can help to establish the standard protocol for recording the VAG signals.

In order to have patients with the same pathological condition, we selected only patients with osteoarthritis for the abnormal group. However, as shown in Table II, the classification accuracy of the abnormal group (88.4 ± 4.4%) was relatively lower than that of the normal group (94.3 ± 3.7%) (p<0.01). It is thought that that the pathological conditions of each patient, such as the sex, age, bone density, low extremity functional scale, onset and treatment duration, might be slightly different. In a future study, we plan to improve the classification accuracy by considering these conditions and obtaining VAG signals from more patients, and to develop supplementary signal processing techniques for reducing unwanted noise, generated unavoidably between the skin and stethoscope. Also, we will apply this technique to a larger number of patients with various joint disorders, such as rheumatoid arthritis, chondromalacia as well as osteoarthritis, and demonstrate its feasibility.

VI. CONCLUSION

In this study, we extracted the features of normal and abnormal VAG signals by TFD, and evaluated the classification accuracy using a BPNN. VAG signals, generated during one flexion and one extension of the knee, were segmented at 2 sec/cycle, in order to keep a constant velocity after the time axis of the VAG signals was normalized by an electro-goniometer and the DTW algorithm. Also, unwanted noise, which might interfere with the discrimination of the normal and abnormal VAG signals, was wholly eliminated by applying the SVD algorithm, in order to divide the TFD of the VAG signals into interesting and noise subspaces and, consequently, the SNR of the enhanced TFD by our algorithm was relatively 38.2 % higher than those of the raw TFD. Therefore, more enhanced TFD parameters could be selected as the features of the normal and abnormal VAG signals. Moreover, the optimum condition of the learning rate and the number of hidden nodes of the designed BPNN was determined by a repeatability test consisting of twenty cycles. The mean and standard deviation of the EP, ESP, FP and FSP obtained by TFD were used as the features of the input vector for the BPNN. In order to evaluate the performance of our devised algorithm, the VAG signals from a total of 1408 segments (normal 1031, abnormal 377) were used for training and testing the BPNN by k-fold cross validation method. As a result, the average accuracy was 91.4 (S.D. ±1.7) %. The proposed method showed good potential for the non-invasive diagnosis and monitoring of joint disorders such as osteoarthritis and chondromalacia.

REFERENCES