CLASSIFICATION OF TRANSIENT TIME-VARYING SIGNALS USING DFT AND WAVELET PACKET BASED METHODS

Christoph Delfs and Friedrich Jondral

Nachrichtentechnik, Universität Karlsruhe
D-76128 Karlsruhe, Germany
delfs@inss1.etec.uni-karlsruhe.de

ABSTRACT

The classification of transient time-varying signals is important for industrial, biomedical and military applications. The attack phase of piano sounds is used as an example for transient, time-varying signals in a real data application. Discrete Fourier transform and time-invariant wavelet packet based algorithms are used alternatively for feature extraction. The training set is used for determining an appropriate feature selection. A classifier checks whether the generated features are sufficient in order to identify the correct piano. Classification results are presented and discussed.

1. INTRODUCTION

In this paper the discrete Fourier transform (DFT) and the translation-invariant wavelet packet transform are compared with regard of their use as feature extractors for time-varying transient signals. Piano sounds are used as an example for this signal class. As complete piano sounds can easily be classified [DJ97a], this paper focuses on the attack phase of piano sounds. Translation-invariant filtering is used for the DWPT, as it improves the classification results [DJ97b].

Fig. 1 shows the flow chart of the signal processing associated with this task. After preprocessing the sound, the resulting sequence $\{s(m)\}_m$ is subject to a feature extraction by means of the discrete Fourier transform (DFT) and the dyadic orthogonal wavelet packet transform (DWPT).

The DFT provides a frequency representation whereas the DWPT calculates a translation-dependent time-frequency map of the input sequence.

Both feature sets can be described using the time-frequency representation matrix $\{F^k(n,m)\}_{n,m}$ with translation index k. After feature selection the input signal is classified.

Section 2 explains the signal preprocessing unit. Section 3 presents the feature extraction techniques used. The following section 4 gives details of the feature selection. Section 5 provides information about the classification process. Finally, classification results are shown and discussed in section 6.

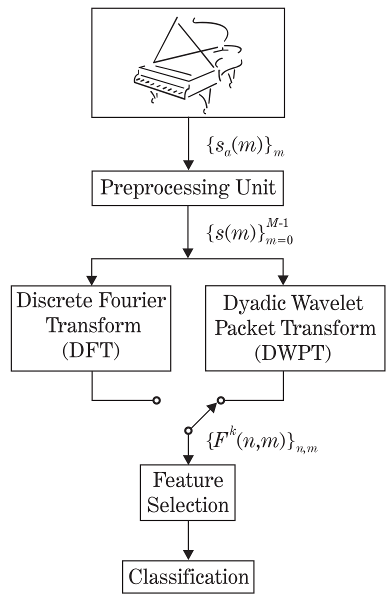

Figure 1: Flow chart of the signal processing for classification

2. PREPROCESSING UNIT

On $C = 34$ different pianos the chord $C_4 - E_4 - G_4$ is touched. The task of the preprocessing unit is to obtain the attack phase of each piano sound.

Fig. 2 shows the signal flow inside the preprocessing unit.

The piano sound $\{s_a(m)\}_m$ is sampled at a rate of $f_a = 48kHz$. The preprocessing unit detects the start sample of a piano sound attack by measuring the average power within a window of length W starting at index M:

$$P_M^W = 10\log_{10}\Big\{\frac{1}{W}\sum_{l=M}^{M+W-1}|s_a(l)|^2\Big\} \quad .$$

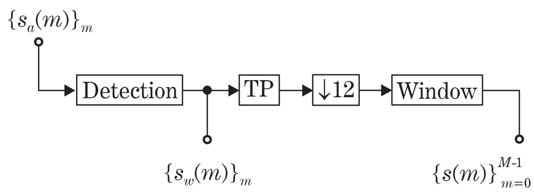

Figure 2: Signal flow inside the preprocessing unit

In order to detect the start of the piano sound precisely, the length of the window is set to $W_{start} = 80$. The start sample M_{start} is the first sample fulfilling:

$$P^{80}_{M_{start}} \geq P_{start} = -25dB.$$

The resulting sequence $\{s_w(m)\}$ is filtered by a linear phase FIR lowpass of odd length L and decimated by 12, as 99% of the signal power are within the frequency range 0 - 2kHz. The length of a piano attack is about 25 ms [Dic87]. By applying a rectangular window of length $M = 100$ on the decimated sequence and taking care of the filter delay, the attack phase is mapped to the sequence $\{s(m)\}_{m=0}^{M-1}$. The operations on the sequence $\{s_w(m)\}$ can be described by:

$$\begin{aligned} s(m) &= \sum_{i=0}^{L-1} h(i) s_w(12m + \frac{L-1}{2} - i), \\ m &= 0, 1, ..M-1 \end{aligned}$$

3. FEATURE EXTRACTION

In the following, details of the extraction methods are explained. They map the shifted input sequence $\{s(m-k)\}_{m=k}^{M-1+k}$ onto the sequence $\{F^k(n,m)\}_{n,m}$. The set $\mathcal{K}$ contains the set of translations which is used by each feature extraction technique.

3.1. The Discrete Fourier Transform (DFT)

In the DFT, the continous Fourier transform of the input sequence is sampled by performing:

$$DFT(n) = \sum_{m=0}^{M-1} s(m) e^{-j2\pi mn/N} \quad n = 0, 1, 2, ..N-1.$$

The sampling parameter $N \geq M$ is chosen to 128, in order to be able to use the fast Fourier transform. As feature set the squared modulus of the DFT (periodogram) is used:

$$\begin{aligned} F^0(n,1) &= |DFT(n)|^2 \qquad n \in \mathcal{G}_{DFT}, && (1) \\ \mathcal{K} &= \{0\}, \quad \mathcal{G}_{DFT} = \{0, 1, 2, ..N/2\}. && (2) \end{aligned}$$

3.2. The Dyadic Wavelet Packet Transform (DWPT)

By applying the DWPT on the sequence $\{s(m)\}_{m=0}^{M-1}$ the expansion of the sequence with respect to a set of orthonormal basis functions is calculated [CW92].
The transform of the not shifted input sequence can be described iteratively as follows given a lowpass filter $\{h(l)\}_{l=0}^{L_2}$, a high pass filter $\{g(l)\}_{l=0}^{L_2}$ and a decomposition depth K [DJ97a]:

$$\begin{aligned} W_0^0(m) &= s(m), \\ W_{2l}^{k+1}(m) &= \sum_{i=0}^{L_2-1} h(i) W_l^k(2m-i), \\ W_{2l+1}^{k+1}(m) &= \sum_{i=0}^{L_2-1} g(i) W_l^k(2m-i), \\ k &= 0, 1, 2 \ldots K-1 \quad , \quad m \geq 0, \\ l &= 0, 1, 2 \ldots 2^k - 1 \quad . \end{aligned}$$

By using $n = 2^k + l$ the sequences $\{W_l^k(m)\}_m$ can be mapped onto the matrix $\{F^0(n,m)\}_{n,m}$:

$$F^0(2^k + l, m) = W_l^k(m). \qquad (3)$$

A basis, i.e. a complete representation of the input sequence, can be described by an appropriate set of indices $n \in \mathcal{G}$. The basis selection is part of the feature selection.
The transform of $\{s(m-k)\}_{m=k}^{M-1+k}, k \in \mathcal{K}$ requires translation-invariant filtering. The algorithms needed are described in [PKC96, DJ97b] and lead to translation-dependent representations $\{F^k(n,m)\}_{n,m}, k \in \mathcal{K}$. The set $\mathcal{K}$ is chosen to

$$\mathcal{K} = \{-R, -(R-1), \ldots, 0, \ldots, R-1, R\}, \quad R \in \mathbf{Z}. \quad (4)$$

Here, the DWPT is performed using Daubechies filters with length $L_2 = 20$, decomposition depth $K = 1, 2$ and $R = 8$.

4. FEATURE SELECTION

The task of the feature selection is to obtain the features which are essential for class separation. In the case of the DWPT the feature selection determines the optimum orthogonal transform additionally.

The basic idea of the feature selection can be described as follows for the DWPT [Sai94]: Given a sequence $\{x(m)\}_m$ and its wavelet packet transform $\{F(n,m)\}_{n,m}$, the best basis algorithm [CW92] gives an optimum basis by determining a set $\mathcal{G}'$ which allows to represent

$\{x(m)\}_m$ with a minimum number of expansion coefficients $\{F(n,m)\}_{n\in\mathcal{G}',m}$ at a given reconstruction error. In the same way, the difference between two signals $\{x_1(m)\}_m$ and $\{x_2(m)\}_m$ can be efficiently expressed by calculating the best basis for the difference signal $\{\Delta x(m) = x_1(m) - x_2(m)\}_m$.

In this paper this approach is used to obtain a class specific feature selection for C classes. The structure given in figure 1 is expanded for feature selection and the classification process in figure 3.

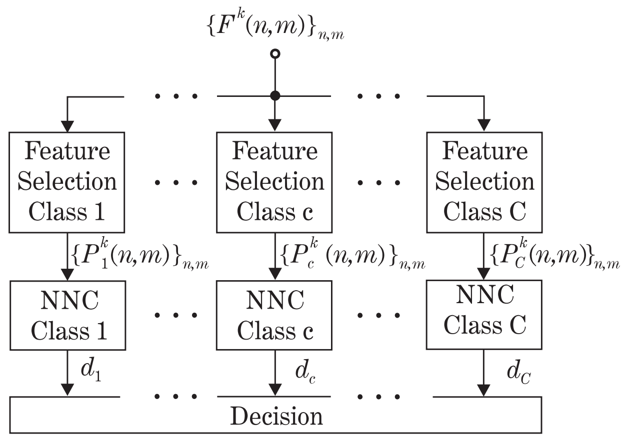

Figure 3: Class specific feature selection and classification (NNC $\hat{=}$ nearest neighbor classifier)

4.1. Adapting the feature selection

During adaptation for each class c a selection template $\{S_c(n,m)\}_{n,m}$ is calculated:

The training data set is divided into two classes. Class c contains N_c normalized signals of class c whereas class $\bar{c}$ contains $N_{\bar{c}}$ normalized signals of all the other classes.

The class signals and its translation-dependent representations are given by:

$$\{s_{c,\alpha}\}_{m=0}^{M-1} \circ\!\!-\!\!\bullet \{F_{c,\alpha}^{k_\alpha}(n,m)\}_{n,m} \qquad \alpha = 1,2,..,N_c,$$
$$\{s_{\bar{c},\beta}\}_{m=0}^{M-1} \circ\!\!-\!\!\bullet \{F_{\bar{c},\beta}^{k_\beta}(n,m)\}_{n,m} \qquad \beta = 1,2,..,N_{\bar{c}}$$

with given signal specific translations $\{k_\alpha\}$ and $\{k_\beta\}$. For the DWPT, two different strategies to choose the translations $\{k_\alpha\}$ and $\{k_\beta\}$ are compared as with regard to the classification results:

A: The translations are set to zero:

$$\begin{aligned} k_\alpha &= 0 \qquad \alpha = 1,2,\ldots,N_c \\ k_\beta &= 0 \qquad \beta = 1,2,\ldots,N_{\bar{c}} \end{aligned}$$

B: The translations $\{k_\alpha\}$ are chosen to minimize the summed squared euclidian distances E_1 between the signals of class c:

$$E_1 = min_{\{k_i\}} \sum_{i=1}^{N_c-1} \sum_{j=i}^{N_c} e^2(\{s_{c,i}(m-k_i) - s_{c,j}(m-k_j)\})$$

using

$$e(\{x(m)\}) = \sqrt{\sum_{m=-\infty}^{\infty} |x(m)|^2}.$$

Given the translations $\{k_\alpha\}$, the translations $\{k_\beta\}$ are chosen to minimize the summed squared euclidian distances E_2 between the signals of class c and the signals of class $\bar{c}$:

$$E_2 = min_{\{k_\beta\}} \sum_{\alpha=1}^{N_c} \sum_{\beta=1}^{N_{\bar{c}}} e^2(\{s_{c,\alpha}(m-k_\alpha) - s_{\bar{c},\beta}(m-k_\beta)\}).$$

The minima E_1 and E_2 are approximated using genetic algorithms [Mic92] for translations from the set $\mathcal{K}$ (4).

The matrix $\{\Delta F_c(n,m)\}_{n,m}$

$$\Delta F_c(n,m) = \frac{1}{N_c N_{\bar{c}}} \sum_{\alpha=1}^{N_c} \sum_{\beta=1}^{N_{\bar{c}}} |F_{c,\alpha}^{k_\alpha}(n,m) - F_{\bar{c},\beta}^{k_\beta}(n,m)|^2 \tag{5}$$

indicates the differences between both classes by averaging the representations of the class signals. For the DWPT, (5) is an additive selection criterion which can be used for determining the set $\mathcal{G}_c$ via the best basis algorithm, i.e. the set $\mathcal{G}_c$ describes the transform which optimally separates both classes given the criterion (5). For the DFT, the set $\mathcal{G}_c$ is set to $\mathcal{G}_c = \mathcal{G}_{DFT}$ (2).

Finally, the selection template $\{S_c(n,m)\}_{n,m}$ is obtained by defining a threshold T_c which controls the number of features to be included in the pattern:

$$S_c(n,m) = \begin{cases} 1 & \text{if} \quad \Delta F_c(n,m) > T_c, \quad n \in \mathcal{G}_c \\ 0 & \text{otherwise.} \end{cases}$$

In simulations, the threshold T_c is chosen so as to obtain a fixed number N_F of features for each class and extraction technique.

4.2. Feature selection during classification

During classification for each translation $k \in \mathcal{K}$ a set of class specific normalized patterns $\{P_c^k(n,m)\}_{n,m}$ is calculated by

$$\begin{aligned} \{Q_c^k(n,m)\}_{n,m} &= \{F^k(n,m) S_c(n,m)\}_{n,m}, \\ \{P_c^k(n,m)\}_{n,m} &= \{Q_c^k(n,m)/d(\{Q_c^k(n,m)\})\}_{n,m} \end{aligned} \tag{6}$$

and

$$d(\{X(n,m)\}) = \sqrt{\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} |X(n,m)|^2}$$

and is fed into the classification process (figure 3).

5. CLASSIFICATION

5.1. Adapting the classifier

After calculation of the selection templates $\{S_c(n,m)\}_{n,m}$ the normalized training patterns are calculated for each class c by:

$$\begin{aligned} \{Q^{REF}_{c,\alpha}(n,m)\}_{n,m} &= \{F^{k_\alpha}_{c,\alpha}(n,m)S_c(n,m)\}_{n,m}, \\ \{P^{REF}_{c,\alpha}(n,m)\}_{n,m} &= \\ &\{Q^{REF}_{c,\alpha}(n,m)/d(\{Q^{REF}_{c,\alpha}(n,m)\})\}_{n,m} \qquad (7) \\ \alpha &= 1,2,\ldots,N_c. \end{aligned}$$

5.2. Classification process

The classification process consists of two steps (figure 3).

First, for each class c a nearest neighbor classifier (NNC) calculates the minimum distance d_c between the set of training patterns (7) and the set of incoming patterns (6):

$$d_c = min_{\alpha,k}(D_{ext}\ d(\{P^k_c(n,m) - P^{REF}_{c,\alpha}(n,m)\}))$$

The constant D_{ext} depends on the extraction technique: $D_{DFT} = 1/\sqrt{2}$ and $D_{DWPT} = 1/2$. The NN classifier c tests the hypothesis that the topical piano attack belongs to class c. The value $1 - d_c$ can be interpreted as probability that the sound belongs to class c.

Therefore, in the second step (decision) the minimum d_{c^*} of the set $\{d_c\}^C_{c=1}$ is calculated:

$$d_{c^*} = min_c d_c$$

and c^* is the class the classificator assigns the sound to.

6. RESULTS AND DISCUSSION

The data set consists of 18 grand pianos, 15 pianos and one keyboard, each with 20 attacks. The classificator was adapted using the technique leaving-5-out, i.e. the data set for each piano was divided into four parts with 5 sounds. For each of the 4 runs three parts were used for adaptation and one for classification. In figure 4 the classification results, averaged with regard to the runs, are shown.

The DFT based features perform best for a small number of features.

The results obtained with DWPT based features reach better results for $N_F \in \{50, 60\}$, $K = 2$ and strategy A. Strategy A seems to create better templates than B, as B performs better than A for $N_F = 10$ only. The influence of the decomposition depth K on the classification rates is small.

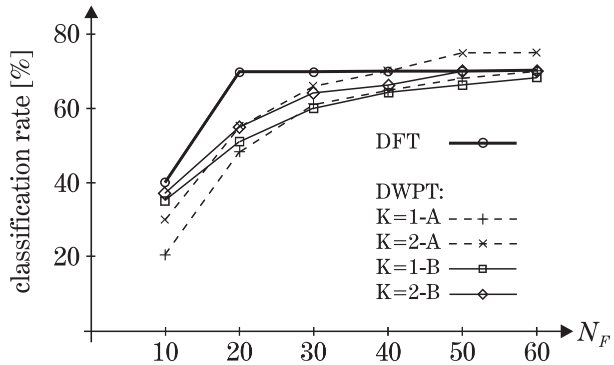

Figure 4: Classification results based on DFT and DWPT features (DWPT: strategies A,B and $K = 1, 2$)

Though sophisticated methods for the feature selection of transient time-varying signals are used, wavelet packet based methods do not seem to offer any advantages in comparison with simple DFT features.

7. REFERENCES

[CW92] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inf. Theory*, vol.38(2):713–718, 1992.

[Dic87] M. Dickreiter. *Musikinstrumente: Moderne Instrumente, historische Instrumente, Klangakustik*. Kassel, Bärenreiter, 1987.

[DJ97a] C. Delfs and F. Jondral. Classification of piano sounds using time-frequency signal analysis. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 2093–2096, 1997.

[DJ97b] C. Delfs and F. Jondral. New aspects on classification using translation-invariant wavelet packet transforms. In *IEEE UK Symposium on Applications of Time-Frequency and Time-Scale Methods*, pages 73–76, 1997.

[Mic92] Z. Michalwicz. *Genetic Algorithms + Data Structures = Evolution Programs*. New York: Springer, 1992.

[PKC96] J.C. Pesquet, H. Krim, and H. Carfantan. Time-invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, vol. 44(no. 8):pp. 1964– 1970, 1996.

[Sai94] N. Saito. *Local Feature Extraction and Its Applications Using a Library of Bases*. PhD thesis, Yale University, December 1994.