METHOD OF CREATING PATTERNS FOR HYDROACOUSTICS SIGNALS

ANDRZEJ ZAK

Naval University of Gdynia
Smidowicza 69, 81-113 Gdynia, Poland
e-mail: andrzej-zak@wp.pl

The paper presents method used to creating patterns for hydroacoustics signals for necessity of sound identification or classification. First the mathematical fundamentals, with breaking to separate processed blocks, of proposed method were introduced. Next the description of realized research and discussion about some obtained results were presented. At the end of the paper the direction of development in creating patterns for hydroacoustics signals and its selectors were pointed.

1. INTRODUCTION

Hydroacoustics signals identification or classification is the process of automatically recognizing what kind of object is generating acoustics signals on the basis of individual information included in generated sounds. All signal recognition systems, at the highest level, contain two main modules (Fig. 1) feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the hydroacoustics signals that can later be used to represent each object. Feature matching involves the actual procedure to identify the unknown object by comparing extracted features from input sounds with the ones from a set of known stored in some kind of database. Therefore signal recognition systems have to serve two distinguish phases. The first one is referred to the enrollment sessions or training phase while the second one is referred to as the operation sessions or testing phase. In the training phase, each registered object has to provide samples of their sounds so that the system can build or train a reference model for that object. During the testing – operational phase, the input sound is matched with stored reference models and recognition decision is made.

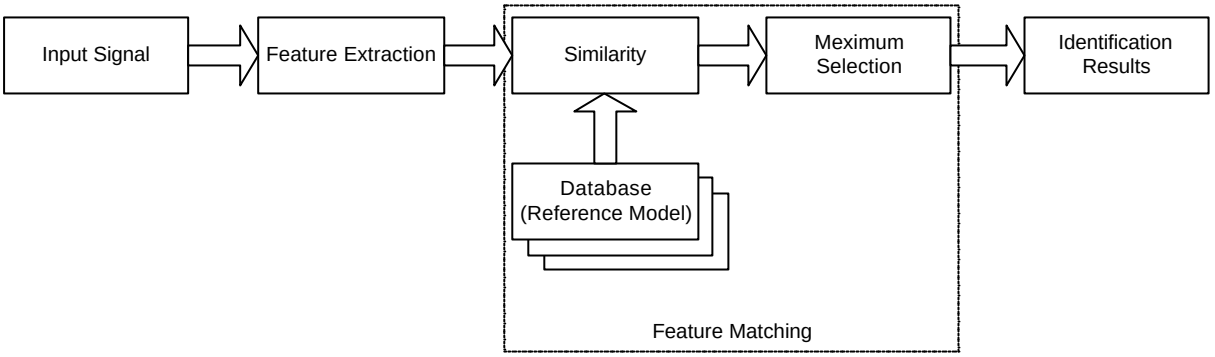

Fig.1 Signal identification

Hydroacoustics signal recognition is a difficult task and it is still an active research area. Automatic signal recognition works based on the premise that sounds emitted by object to the environment are unique for that object. However this task has been challenged by the highly variant of input signals. The principle source of variance is the object himself. Sound signals in training and testing sessions can be greatly different due to many facts such as object sounds change with time, efficiency conditions (e.g. some elements of machinery are damaged), sound rates, etc. There are also other factors, beyond object sounds variability, that present a challenge to signal recognition technology. Examples of these are acoustical noise and variations in recording environments and changes of environment itself.

As a hydroacoustics signals in this paper will be understood only human made disturbance, especially sound made by ships in motion. Researches produce that ships has characteristic for them components of spectra in frequency range from about 5 Hz to 2 kHz. Moreover hydroacoustics signals are quasi-stationary so short-time spectral analysis is one of the common way to characterize this kind of sounds.

2. SIGNAL FEATURE EXTRACTION

The purpose of signal feature extraction module is to convert the sound waveform to some type of parametric representation for further analysis and processing. This is often referred as the signal-processing front end. A wide range of possibilities exist for parametrically representing the signals for the sound recognition task, such as Linear Prediction Coding (LPC), Mel-Frequency Cepstrum Coefficients (MFCC), and others. Mel-Frequency Cepstrum Coefficients method will be discussed in this paper.

MFCC's are based on the known variation of the human ear's critical bandwidths with frequency, filters spaced linearly at low frequencies and logarithmically at high frequencies have been used to capture the phonetically important characteristics of speech. This is expressed in the mel-frequency scale, which is a linear frequency spacing below 1000 [Hz] and a logarithmic spacing above 1000 [Hz].

A block diagram of the structure of an MFCC processor is given in Fig. 2. As been discussed previously, the main purpose of the MFCC processor is to mimic the behavior of the human ears. In addition, rather than the speech waveforms themselves, MFFC's are shown to be less susceptible to mentioned variations.

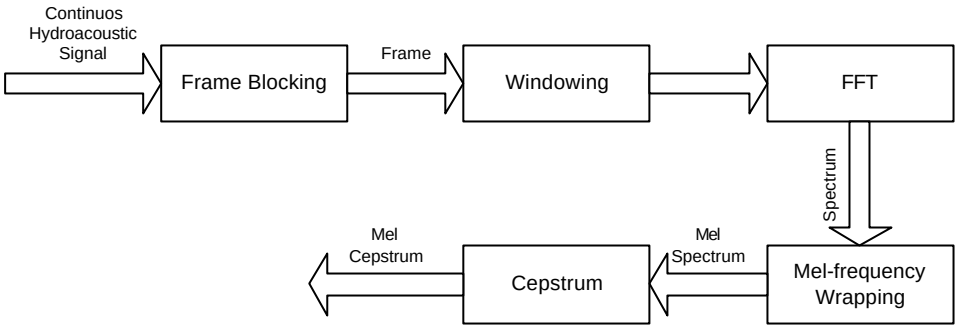

Fig.2 Block diagram of the MFCC processor

First step of MFCC processor is the frame blocking. In this step the continuous sound is blocked into frames of N samples, with adjacent frames being separated by M where $M < N$. The first frame consists of the first N samples. The second frame begins M samples after the first frame, and overlaps it by $N - M$ samples. Similarly, the next frames are created so this process continues until all the sound is accounted for within one or more frames.

The next step in the processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. The concept here is to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. If we define the window as: $w(n)$, $0 \le n \le N-1$, where N is the number of samples in each frame, then the result of windowing is the signal:

$$y(n) = x(n)w(n), \quad 0 \le n \le N-1 \tag{1}$$

Typically the Hamming window is used, which has the form:

$$w(n) = 0.54 - 0.46\cos(\frac{2pn}{N-1}), \quad 0 \le n \le N-1 \tag{2}$$

The next processing step is the Fast Fourier Transform, which converts each frame of N samples from the time domain into the frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier Transform (DFT) which is defined on the set of N samples, as follow:

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2pjkn/N}, \quad n = 0,1,2,\ldots,N-1 \tag{3}$$

Next step in MFCC processor is the Mel-frequency Wrapping. As mentioned above, psychophysical studies have shown that human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency f, measured in [Hz], a subjective pitch is measured on a scale called the 'mel' scale. The mel-frequency scale is a linear frequency spacing below 1000 [Hz] and a logarithmic spacing above 1000 [Hz]. As a reference point, the pitch of a 1 [kHz] tone, 40 [dB] above the perceptual hearing threshold, is defined as 1000 mels. Therefore we can use the following approximate formula to compute the mels for a given frequency f in [Hz]:

$$mel(f) = 2595 \cdot \log_{10}(1 + \frac{f}{700}) \tag{4}$$

One approach to simulating the subjective spectrum is to use a filter bank, spaced uniformly on the mel scale (Fig. 3). That filter bank has a triangular bandpass frequency response, and the spacing as well as the bandwidth is determined by a constant mel frequency interval.

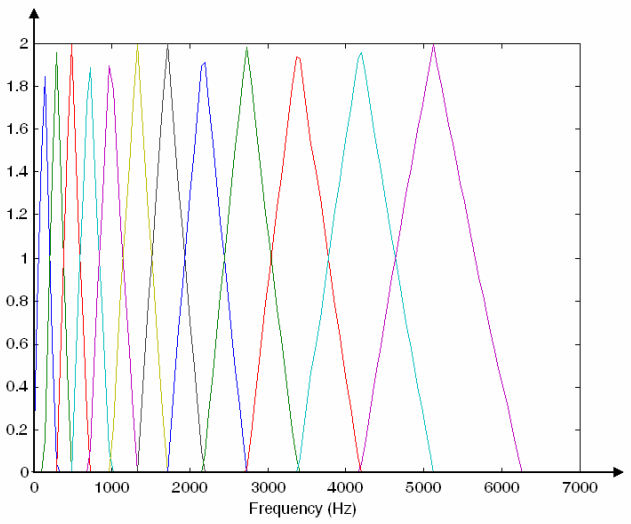

Fig.3 An example of mel-spaced filterbank

In this final step, we convert the logarithmic mel spectrum back to time. The result is called the mel frequency cepstrum coefficients (MFCC). The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the mel spectrum coefficients, and so their logarithm, are real numbers, we can convert them to the time domain using the Discrete Cosine Transform (DCT). Therefore if we denote those mel power spectrum coefficients that are the result of the last step are S_k, $k = 1,2,\ldots,K$, we can calculate the MFCC's, as

$$c_n = \sum_{k=1}^{K} \log(S_k) \cos(\frac{n(k-0.5)p}{K}), \quad n = 1,2,\ldots,K \tag{5}$$

Note that we exclude the first component, c_0 from the DCT since it represents the mean value of the input signal which carried little speaker specific information.

3. RESULTS OF RESEARCH

The problem of signal recognition belongs to a much broader topic in scientific and engineering so called pattern recognition. The goal of pattern recognition is to classify objects of interest into one of a number of categories or classes. The objects of interest are generically

called patterns and in our case are sequences of acoustic vectors that are extracted from an input sounds using the techniques described in the previous section. The classes here refer to individual objects. Since the classification procedure in this case is applied on extracted features, it can be also referred to as feature matching.

The state-of-the-art in feature matching techniques used in speaker recognition include Dynamic Time Warping (DTW), Hidden Markov Modeling (HMM), Vector Quantization (VQ) and artificial neural networks. In this project, the artificial neural network approach will be used, due to ease of implementation and high accuracy. The main task of neural network is a process of mapping vectors from a large vector space to a finite number of regions in that space.

For research the selector based on the feed forward artificial neural network was created. Topology of neural network was choosed during research and finally has input layer consist of 40 neurons, three hidden layer with 120, 240, 60 neurons in layers respectively and output layer with 8 neurons. As a teaching method the back propagation method based on conjugate gradient method was applied. The learn rate factor was calculated adaptively using directed minimization method. Input vector was created by applying the procedures described above so a set of mel-frequency cepstrum coefficients as an acoustics vectors were computed. The training vector consist of set of mel-frequency cepstrum coefficients and the number of class which this vector represents. Uses neurons has sigmoidal, bipolar activation function with bias.

Researches were divided into two phases. The first was training phase during which the teaching vectors were presented to artificial neural network, and after achievement of required level of average square error of neural network answer the testing phase was conducted. In this phase the already presented input vectors were recalculated by neural network and the errors in answer were measured. Figure 4 shows the changes of neural selector incorrect answer during training phase.

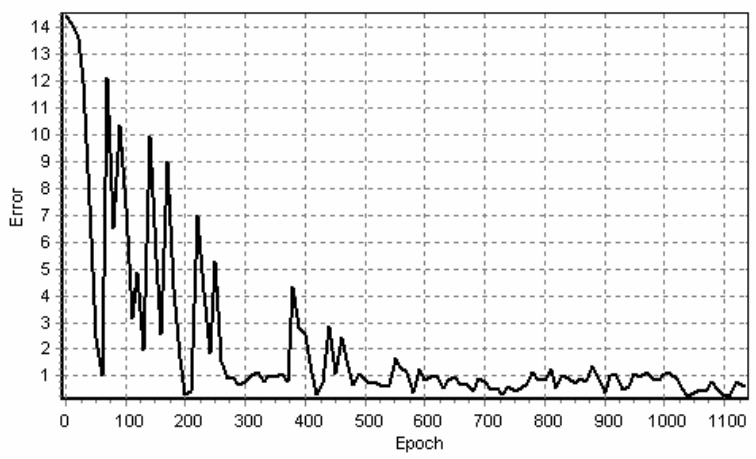

Fig.4 Graph of errors produced by artifical neural network during training phase

The next step was to check the network proper behaviour after addition to signal some noises. The last step of testing was to present to the neural selector some unknown acoustics vectors. The percent of incorrect answers of artificial neural network in every three steps of testing phase were presented in table below.

Tab.1 The error of answer of neural selector for presented acoustic vector

Step of test phase	Error [%]
Presented before vectors	0.4
Presented before vectors with added 15 % noise	2.3
Presented before vectors with added 25 % noise	3.7
Not presented before vectors	4.2

This results shows that neural selector was working correct, and what is the most important, considered method of creating patterns for hydroacoustics signals fulfill the expectation about possibility of object distinction.

4. SUMMARY

As the result of research shows the proposed method of creating pattern for hydroacoustics signals is good enough as a basic method used for necessity of signal identification or classification. To find out how precision proposed method could be the next much broader research should be done. The future research should first of all give an examination if this presented method can not only answer what kind of ship is actually measured but also what is its board number or even what is the machinery state despite of noises during measurements.

Next research will be focused first of all to increase the possibilities of selector such as it can automatically creating clusters for new objects and has possibilities to describe object much more widely what has main meaning on classification precision. Author is thinking that using self organized Kohonen networks or Grosberg neural networks with unsupervised learning method can solve some of these problems. Other task, which refer to the method of creating patterns for hydroacoustics signals, is to compare presented method with another commonly used in sound recognition. Author see some possibilities in increasing precision in description of sound source using wavelets which to this time hasn't application in this area of research.

REFERENCES

[1] L.R. Rabiner, B.H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, N.J., 1993.

[2] L.R Rabinem, R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Englewood Cliffs, N.J., 1978.

[3] Minh N. Do, An Automatic Speaker Recognition System, Swiss Federal Institute of Technology, Lausanne, 1999

[4] S. Osowski, Sieci neuronowe w ujeciu algorytmicznym, WNT, Warszawa, 1996

[5] T. Schalk, P. J. Foster, Speech Recognition: The Complete Practical Reference Guide, Telecom Library Inc, New York, ISBN O-9366648-39-2

[6] C.H. Lee, F.K. Soong, K.K. Paliwal, Automatic Speech and Speaker Recognition: Advanced Topics, Kluwer, Boston, 1996