On the construction of wavelet network

Chen, Fei

2006

Chen, F. (2006). On the construction of wavelet network. Master' s thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/47480>

<https://doi.org/10.32657/10356/47480>

Nanyang Technological University

Downloaded on 26 Apr 2023 23:30:38 SGT

9416454

On the Construction of Wavelet Network

Chen Fei

School of Computer Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Master of Engineering

2006

QA
76.87
C518
2006

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Dr. Daming Shi, for his instruction and guidance during my research. Without his enthusiasm and encouragement, this work could not have been done.

Thanks also go to Josephine, Celine and Dr. Heng, who involve me into their pharmaceutical research group in National University of Singapore. It is indeed a great chance for me to solve practical problems with my methodologies.

Last but not least, I would like to give special thanks to the people at the Center for Computational Intelligence for giving me the most enjoyable and inspiring working environment.

TABLE OF CONTENT

ACKNOWLEDGEMENTS.....i

TABLE OF CONTENT iii

LIST OF FIGURESv

LIST OF TABLES vii

SUMMARYix

1 INTRODUCTION 1

 1.1 Motivation..... 1

 1.2 Literature Review2

 1.3 Aims of the Research..... 11

 1.4 Organization of the Thesis..... 11

2 OPTIMIZATION OF WAVELET NETWORK 13

 2.1 Introduction..... 13

 2.2 Wavelet Transform 15

 2.3 Wavelet Network Based on Wavelet Theory 19

 2.4 Wavelet Network Pruning 21

 2.5 Experiments on Approximation..... 27

 2.6 Summary 34

3 APPLICATION I:MICROCRYSTALLINE CELLULOSE GROUPING 35

 3.1 Introduction..... 35

 3.2 Wavelet Network Modeling..... 37

 3.3 Discrete Incremental Clustering 39

 3.4 Experimental Results 41

 3.5 Improvement and Validation 44

 3.6 Summary 50

4 PPLICATION II: TALKING HEAD53

4.1 Introduction.....53

4.2 Lip Modeling56

4.3 Speech-Driven with Wavelet Network57

4.4 Hybrid Driving Scheme59

4.5 Experiments and Results.....64

4.6 Summary65

5 CONCLUSIONS AND FUTURE WORK.....67

5.1 Conclusions.....67

5.2 Future Work68

APPENDIX A: GLOSSARY69

REFERENCES71

LIST OF FIGURES

Figure 1. 1. Architectural graph of a multilayer perceptron with two hidden layers..... 4

Figure 1. 2. The Architecture of RBF network 5

Figure 2. 1 The wavelet coefficients between $f(x)$ and scaled and translated wavelets.. 17

Figure 2. 2. Filter scheme of DWT 18

Figure 2. 3. Multiresolution structure of DWT 19

Figure 2. 4. The pruning algorithm of wavelet network 25

Figure 2. 5. The diagram of the stochastic gradient algorithm 26

Figure 2.6. Mother Wavelet (Mexican-hat wavelet)..... 27

Figure 2. 7. The selection of the optimal set of wavelets 29

Figure 2. 8. MSEs during the different pruning procedures 30

Figure 2. 9. Speech signal approximation by wavelet network 32

Figure 2. 10. Predicted signal by wavelet network..... 33

Figure 3. 1. Parameters in DIC 40

Figure 3. 2. Validation of measured torque with water: MCC ration..... 42

Figure 3. 3. Visualization of MDS algorithm 49

Figure 4. 1. Seven control points and one standard points 56

Figure 4. 2. User interface for region definition 57

Figure 4. 3. Audio signal analysis on “sha” 63

Figure 4. 4. The synthetic talking head system..... 64

LIST OF TABLES

Table 2.1. Approximation results of $f(x) = (x_1^2 - x_2^2)\sin(0.5x_1)$ 28

Table 2. 2. MSE and FPE of wavelet networks with different pruning algorithms.....30

Table 2. 3. Performance comparison between wavelet network and other models.....33

Table 3. 1. Calculated MSE by using wavelet network I43

Table 3. 2. Grouping of MCCs I.....43

Table 3. 3. Calculated MSE by using wavelet network II47

Table 3. 4. Grouping of MCCs II.....48

Table 4. 1. Chinese phoneme combination scheme.....60

Table 4. 2. Phoneme to viseme conversion of basic phonemes.....61

Table 4. 3. Phoneme to viseme conversion of multi-vowel phonemes61

SUMMARY

By taking advantage of both the scaling properties of wavelets and the high learning ability of neural networks, a wavelet network exhibits a high approximation and prediction capability. It can well approximate data, and thus be a good mathematic tool to find the hidden nature of a sequence of data, and to visualize the behavior of the data. According to Ocam's Razor hypothesis, compact models are believed to have the best generalization ability, since they represent data with the lowest structural complexity. Therefore, this research aims to construct a compact wavelet network, which can achieve an ideal approximation performance with the least number of wavelets. There are three steps involved to construct a wavelet network. First, a wavelet candidate pool is built up with a pyramid type initialization method. After that, the orthogonal least square (OLS) algorithm is employed to find those most critical wavelets. Finally, a stochastic gradient algorithm is utilized to recursively adjust the parameters.

In this report, a wavelet network has been successfully applied to solve some practical problems, especially microcrystalline celluloses (MCC) grouping, which is especially advantageous for pharmaceutical formulation scientists during the preformulation and optimization phases of product development. The interaction of water and microcrystalline celluloses is modeled by wavelet networks. Then, the mean square errors (MSE) from wavelet network models are used to calculate the membership function value in the discrete incremental clustering (DIC) algorithm. At last, the multidimensional scaling (MDS) algorithm is integrated to provide users a visual representation of the MCC grouping result.

CHAPTER

1

INTRODUCTION

1.1 Motivation

Nowadays, statistical learning plays an important role in many fields, such as function approximation, audio signal processing, video surveillance, finance forecasting, etc. There are two main goals of statistical learning, namely: finding a mathematical representation, and forecasting future values of the time series variable. Both of these goals require that the pattern of observed data is identified and more or less formally described. Once the pattern is established, we can interpret and integrate it with other data (i.e., use it in our theory of the investigated phenomenon). Regardless of the depth of our understanding and the validity of our interpretation (theory) of the phenomenon, we can extrapolate the identified pattern to predict future events.

In addition, mathematical models with low structural complexity are more and more popular for their better generalization performance and lower computational load. Therefore, to find a robust model, and then to employ the idea from compact models to optimize it, we can construct a useful mathematical model for various applications.

Introduction

On the other hand, in pharmaceutical area, researchers often encounter an awkward situation where they can collect many useful observation data, yet hardly uncover the rules within these data. Thus our compact mathematical model would be an ideal method to lighten the burden in these cases. As might be expected, there are many areas to which our model can also contribute, such as media mapping.

1.2 Literature Review

In this section, we will first briefly review some successful mathematical models in machine learning area. Following that, we will have a look at machine learning techniques utilized in the pharmaceutical area. At last, the state of the art in talking head field will be introduced.

1.2.1 Neural Networks

A neural network is an information processing system that has certain performance characteristics in common with biological neuron networks. Neural networks have been developed as generalizations of mathematical models of human neural biology, based on the assumptions that: information processing occurs at many simple elements called neurons, signals are passed between neurons over connection links, each connection link has an associated weight, which multiplies the signal transmitted, and each neuron applies an activation function to the input to determine the output signal. For different purposes, many neural networks with different topologies have been developed.

1.2.1.1 Multilayer Perceptrons

Typically, the multilayer perceptrons (MLP) network consists of a set of neurons that constitute the input layer, one or more hidden layer of computation nodes, and an output

Introduction

layer of computation nodes (Haykin, 1999). The function of the hidden layer neurons is to intervene between the external input and network output. By adding one or more hidden layer, the network is able to approximate severe non-linearity. In the architecture, the hidden layer neuron outputs are used as inputs to the third layer, and so on for the rest of the network, as shown in Figure 1.1. However, it is not necessary to use all external input signals or hidden layer neuron outputs to the input of neurons at the next layer.

The model of each neuron in the network includes a nonlinear activation function. The important point to emphasize here is that the non-linearity is smooth (i.e. differentiable everywhere). A common used form of non-linearity that satisfies this requirement is a sigmoidal non-linearity defined by the logistic function:

$$y_i = \frac{1}{1 + \exp(-v_i)} \quad (1.1)$$

where v_i is the weighted sum of all synaptic inputs plus the bias of neuron i , and y_i is the output of the neuron. Moreover, the use of the logistic function is biologically motivated, since it attempts to account for the refractory phase of real neurons.

MLPs have been applied successfully to solve some difficult and diverse problems by training them in a supervised manner with a highly popular algorithm known as the error back-propagation algorithm.

Introduction

Figure 1. 1. Architectural graph of a multilayer perceptron with two hidden layers

Several results have been published, all showing that multi-layer perceptrons (MLP) of different forms and complexity can approximate arbitrarily well a continuous function, provided that an arbitrarily large number of units is available (Cybenko, 1989; Funahashi, 1989; Stinchcombe and White, 1989; Carrol and Dickinson, 1989).

1.2.1.2 Radial Basis Function Network

A RBF classifier is a three-layer neural network model, in which an N -dimensional input vector $\mathbf{x} = (x_1, x_2, \dots, x_N)$ is broadcast to each of K neurons in the hidden layer. The value of each hidden neuron is obtained from the neuron's activation function, typically a Gaussian kernel:

$$h_i = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{2\sigma_i^2}\right), \quad i = 1, 2, \dots, K, \quad (1.2)$$

Introduction

where $\mathbf{c}_i$ and σ_i^2 are the center and width of the Gaussian basis function of the i th hidden unit, respectively. The units in the output layer have interconnections with all the hidden units.

A single-output RBF neural network with K hidden neurons has the form:

$$f(\mathbf{x}) = \mathbf{w}\mathbf{h} = \sum_{i=1}^K w_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{2\sigma_i^2}\right), \quad (1.3)$$

where $\mathbf{h} = (h_1, h_2, \dots, h_K)$ is the input vector from the hidden layer, and the w_{ij} is the interconnection weight between the j th output neuron and the i th hidden neuron. The architecture of RBF network is shown in Figure 1.2.

Figure 1. 2. The Architecture of RBF network

Introduction

RBF networks have shown considerably better scaling properties, when increasing the number of hidden neurons, than networks with the sigmoid activation function (Robel, 1996).

1.2.1.3 Wavelet Network

Wavelet theory is one of the hottest topics since last decade for its ability of analyzing signals in multi-resolution. Based on the theory, wavelet analysis processes signal at different scales or resolutions. It provides a Time-Scale presentation of a sequence of input signal, so that we can notice the gross features at large scales, and detailed changes at small scales. Thus, this attractive theory has offered very efficient algorithms for analyzing, approximating, estimating, and compressing functions or signals. On the other hand, a neural network can also be considered as a function represented by the composition of many basic functions.

Therefore, based upon this similarity, the wavelet network was first proposed by Zhang and Benveniste in 1992 (Zhang and Benveniste, 1992). It is a special feed-forward network supported by wavelet theory. By taking advantage of both the scaling properties of wavelets and the high learning ability of neural network, the wavelet network becomes a powerful tool for non-parametric approximation. The advantages of wavelet network include:

- 1) In contrast to the random initialization commonly used in neural networks, wavelet networks can be initialized with regular pyramid type wavelet lattice, which is derived from the multi-resolution representation of wavelet analysis (Zhang and Benveniste, 1992). This method is intuitively more reasonable than random initialization.

Introduction

- 2) Taking the observation data into account, the pyramid type initialization method of wavelet network can be even more efficient.
- 3) With the support from wavelet theory, the parameters in the wavelet network have an explicit meaning.

Attracted by these advantages, more and more researchers have started investigating the potentials of wavelet networks.

Jiao *et al.* have employed multiwavelets to replace wavelets in wavelet networks (Jiao *et al.*, 2001). As an extension of wavelets, a multiwavelet can preserve all the advantages the wavelet has. Furthermore, it can simultaneously have several properties very useful in practical applications such as orthogonality, regularity, symmetry, and compact support, which is impossible for a scalar wavelet. In addition, the smoothness of a multiwavelet can be improved by means of two-scale transforms. Their experimental results show that the multiwavelet neural network has some advantages over the wavelet network, especially in the approximation of smooth functions.

Inspired by the theory of multiresolution analysis (MRA) of wavelet transforms and fuzzy concepts, Ho *et al.* have proposed a fuzzy wavelet network (FWN) for approximating arbitrary nonlinear functions (Ho *et al.*, 2001). The FWN consists of a set of fuzzy rules. Each rule corresponding to a sub-wavelet neural network consists of wavelets with a specified dilation value (i.e., resolution). Thus the sub-WNNs at different resolution levels are used to capture different behaviors (global or local) of the approximated function. Through efficient bases selection, the dimension of the approximated function does not cause the bottleneck for constructing FWN. Especially, by learning the translation parameters of the wavelets and adjusting the shape of

Introduction

membership functions, the model accuracy and the generalization capability of the FWN can be remarkably improved. The experimental results presented illustrate the effectiveness of FWN.

In 2005, Billings and Wei have developed a new class of wavelet networks for nonlinear system identification (Billings and Wei, 2005). In the new networks, the model structure for a high-dimensional system is chosen to be a superimposition of a number of functions with fewer variables. By expanding each function using truncated wavelet decompositions, the multivariate nonlinear networks can be converted into linear-in-the-parameter regressions. Then, a forward orthogonal least squares (OLS) algorithm and the error reduction ratio (ERR) is incorporated to solve the linear-in-the-parameters problem.

Encouraged by the powerful approximation ability of wavelet networks for nonlinear systems, wavelet networks also have been widely applied in various engineering fields, such as digital telecommunication system monitoring (Daponte, *et al.*, 2001), adaptive control for mobile robot (de Sousa, *et al.*, 2002), power transformer condition monitoring (Huang, *et al.*, 2002), dissolved gas analysis of oil-insulated power apparatus (Huang, 2003), signal clustering (Galvão, *et al.*, 2004), control method for disk drives (Chang and Liu, 2006). In the foreseeable future, wavelet networks will keep attracting attentions in both academic and engineering fields.

1.2.2 Pharmaceutical Research

Neural networks represent a promising modeling technique especially for data sets having the kind of non-linear relationships, which are frequently encountered in pharmaceutical processes (Agatonovic-Kustrin and Beresford, 2000). Neural networks are able to detect complex non-linear relationships between dependent and independent variables and all

Introduction

possible interactions without complicated equations, and can use multiple training algorithms. In terms of model specification, neural networks require no knowledge of the data source but, since they often contain many weights that are estimated, they require large training sets. In addition, neural networks can combine and incorporate both literature-based and experimental data to solve problems. The use of neural networks is a new but expanding area in the field of pharmaceutical research (Hussain, *et al.*, 1991; Murtoniemi, *et al.*, 1994; Gasperlin, *et al.*, 1998; Takayama, *et al.*, 1999; Achanta, *et al.*, 1995).

The potential applications of neural network methodology in the pharmaceutical science are broad, based on these abilities. They range from interpretation of analytical data (modeling the pharmaceutical analysis in quality control), drug design (quantitative structure-activity relationships and molecular modeling) and dosage from design (optimization of manufacturing processes) to clinical pharmacy through biopharmacy (pharmacokinetic and pharmacodynamic modeling, in vitro/in vivo correlation). A number of commercial software products for pharmaceutical research exists.

1.2.3 Talking Head

Lifelike talking faces for interactive services are an exciting new modality for human-machine interactions. Recent developments in speech and language analysis, and computer graphics enable the real-time synthesis of faces that look and behave like real people, opening opportunities to make interactions with computers more like face-to-face conversations (Cosatto, *et al.*, 2003). With advances in speech recognition, natural language interpretation and speech synthesis, conversational interfaces are finding wider acceptance. At the same time, animation techniques can synthesize human faces

Introduction

providing visual output by a machine. All these techniques combined give human-machine interfaces the look and feel of human-human interaction, which is always the most preferred way for human communication. They definitely have a wide range of applications, such as remote education, customer care, information center, and so on (Gratch, *et al.*, 2002).

From the point of view of creating talking heads, there are two main approaches: sample-based animations and model-based animations. The sample-based approach selects and concatenates segments of recorded videos (Bregler, *et al.*, 1997; Cosatto, *et al.*, 1998; Cosatto, *et al.*, 2000). A few applications also employ some morphing methods to smooth boundaries (Ezzat, *et al.*, 1998; Ezzat, *et al.*, 2002;). On the other hand, the model-based approach uses three-dimensional wire-frame models, which can be animated from high-level parameters such as muscle actions, lip postures, and even facial expressions (Lee, *et al.*, 1995; Gutierrez-Osuna, *et al.*, 2005).

To be considered natural, a face has to be not just photo-realistic in appearance, but must also exhibit proper postures of the lips, synchronized perfectly with the speech. Therefore, besides the animation techniques, the most important issue in giving natural talking heads is to synchronize animations with other media, mainly speech signals. It is also mentioned as the “driven” technique of talking head sometimes. The fundamental idea of synchronization is to find mappings from some inputs to visemes (A viseme is an experimental concept (Pelachaud, 2002), and commonly defined as lip shapes). There are many different mappings of different media to visemes exist.

Excluding some intermediate representations, both text and speech signals can be used as inputs to drive animations. Indisputably, speech signals are the most dominant

Introduction

inputs for the talking head synchronization. Thus, speech-driven techniques are the mainstream of current talking head systems. However, a simple mapping between speech signals and lip shapes (visemes) does not exist. Therefore, many researchers investigate how to animate a talking face from the human voice. In order to derive lip shapes directly from audio, neural networks (Hong, *et al.*, 2002), hidden Markov models (HMMs) (Brand, 1999; Nakamura, 2000; Fu, *et al.*, 2005), k nearest-neighbor procedure (KNN) (Gutierrez-Osuna, *et al.*, 2005), linear prediction analysis (Lewis, 1987), and sound segmentation (Nahas, 1988) were developed. These systems are successful to an extent, and real-time capable.

1.3 Aims of the Research

In this chapter, some mathematical models in machine learning have been introduced. The wavelet network shows a higher approximation ability, as well as a robust and flexible structure.

This research is aimed at constructing a compact wavelet network, which can achieve an ideal generalization performance with as few as possible wavelets. The optimized wavelet network will be compared to wavelet networks with other architectures from the point of view of approximation capability. Thereafter, the optimized wavelet network will be applied to solve some practical problems in pharmaceutical research, as well as audio-visual synthesis.

1.4 Organization of the Thesis

The rest of this thesis is organized as follows. In Chapter 2, the optimization method of the wavelet network is discussed and analyzed. Then, the optimized wavelet network

Introduction

applied to pharmaceutical field is reported in Chapter 3. In Chapter 4, another useful application, audio-visual synthesis, is introduced. Finally, the conclusions are offered in Chapter 5.

CHAPTER

2

OPTIMIZATION OF WAVELET NETWORK

2.1 Introduction

The wavelet theory is a rapidly developing branch of mathematics, which has many applications, for instance, numerical analysis and signal processing (Chui, 1992; Ruskai, *et al.*, 1992). Though this attractive theory has offered very efficient algorithms for analyzing, approximating, estimating and compressing functions of signals, its applications are usually limited to problems of small input dimension d , typically with $d=1$ or 2, because constructing and storing wavelet basis of large input dimension are of prohibitive cost (Benveniste, *et al.*, 1994).

On the other hand, the neural networks are a class of computational architectures that are composed of inter connected nodes (neurons). Although some theoretical results have shown the potential of using feedforward neural networks as universal approximators (Cybenko, 1989; Barron, 1993), the implementation of such networks suffers from the lack of efficient constructive method for the specification of parameters and network architecture. For MLP, due to its multilayered structure and the greedy nature of the BP algorithm, the training processes often settle in undesirable local minima of the error surface. Furthermore, without any localizability, the MLP converges slowly.

Optimization of Wavelet Network

As an alternative to the MLP, the radial basis function (RBF) network introduces two parameters in its node functions: the center and the variance. RBFs thus perform better in learning functions with local variations and discontinuities (Poggio and Girosi, 1990; Moody and Darken, 1989; Hartman, *et al.*, 1990). However, the two parameters and the number of the hidden nodes have to be determined by training because there is not any theoretical guidance in choosing them, which is bound to increase the complexity of the algorithm or result in a large error, as is obviously seen in (Zhang, *et al.*, 1995; Jiao, *et al.*, 2001).

Due to the similarity between discrete inverse wavelet transform and Radial Basis Function (RBF) network, the idea of combining both wavelets and neural networks has been proposed by Zhang and Benveniste in 1992 and further developed in subsequent works. In the paper (Zhang and Benveniste, 1992), the wavelet network was introduced that is a class of feedforward networks composed of wavelets. There are several advantages to incorporate wavelets with neural networks, for instance, wavelet theories provide approximation abilities for feedforward neural networks, and on the other hand, neural networks provide effective ways to determine wavelet weights.

Several researchers have independently studied on the incorporation between the wavelet theory and neural networks. In 1993, the discrete wavelet transform was used by Pati and Krishnaprasad for analyzing and synthesizing feedforward neural networks; in 1992, Hong used orthonormal wavelet bases for constructing wavelet based neural network; and similar works have been reported by Bakshi and Stephanopoulos in 1993.

2.2 Wavelet Transform

The past decade has witnessed the development of wavelet analysis, a new tool which emerged from mathematics and was quickly adopted by diverse fields of science and engineering. Wavelet analysis has begun to play a serious role in a broad range of applications, including signal processing, data and image compression, modeling multiscale phenomena, and statistics. It provides a systematic new way to represent and analyze multiscale structures. Wavelet analysis is also a far-reaching generalization of orthogonal bases of functions whose particular new contribution is a systematic way to present functions on unbounded domains by linear combinations of orthogonal basis functions that are compactly supported and overlapped (Young, 1992; Resnikoff and Wells, 1998; Burrus, *et al*, 1998).

In contrast to Fourier analysis, Wavelet analysis processes signal at different scales or resolutions. It provides a Time-Scale presentation of a sequence of input signal, so that we can notice the gross features at large scales, and detailed changes at small scales. Also since the wavelet has compact support in both time and frequency domains, it is especially good at approximating any signals in finite domains. As an advanced alternative over classic Fourier analysis, wavelets have been successfully used in many aspects of signal processing, such as de-noising, compressing, edge detection, and so on.

In other words, wavelet analysis breaks a signal or a function into a set of wavelets. This breaking down process is termed as wavelet decomposition or wavelet transform. Similarly, the process of putting the wavelet pieces back together to retrieve the original signal is called wavelet reconstruction or inverse wavelet transform.

Optimization of Wavelet Network

The mother wavelet, $\psi(x)$, which defines the characteristics of the wavelet transform and coefficients, is the kernel of the wavelet transform. Performing affine operations on the mother wavelet creates a set of scaled and translated versions of this mother wavelet function, $\left\{ \psi\left(\frac{x-b}{a}\right) \right\}$. The continuous wavelet transform (CWT) is given by

$$W(a,b) = |a|^{-\frac{1}{2}} \int f(x) \psi\left(\frac{x-b}{a}\right) dx \quad (2.1)$$

where a is the scale parameter,

b is the translation parameter, and

$f(x)$ is the original signal.

A wavelet coefficient, $W(a,b)$, at a particular scale and translation represents how well the signal, $f(x)$, and the scaled and translated wavelet match; if the signal is similar to the scaled and translated wavelet, then the coefficient will have a big magnitude. In another word, the coefficient represents the “degree of correlation” between the two functions, $f(x)$ and scaled and translated wavelet, as shown in Figure 2.1. Therefore, the set of all wavelet coefficients, $W(a,b)$, are the wavelet domain representation of the function $f(x)$ with respect to wavelet $\psi(x)$. This two dimensional surface is the multiresolution representation of the specific $f(x)$.

Figure 2. 1 The wavelet coefficients between $f(x)$ and scaled and translated wavelets of different translation parameters

If these scaled and translated wavelets are considered as a window, then this window size will change for each different scale. The changing window size causes the number of cycles or oscillations within a fixed period to change too. If the scale is less than one, then the mother wavelet or window is compressed and more cycles of the mother wavelet will be included in a shorter time interval. More cycles in a shorter time interval implies higher frequencies. That is how wavelet analysis handles high frequency signals with small scale and low frequency signals with large scale, instead of a fixed window in Short Time Fourier Transform (STFT).

Practically, all the digital signals are discrete signals. The finite scale multiresolution representation of a sampled or discrete function is referred to as the

Discrete Wavelet Transform (DWT). The DWT basic computational step is a convolution followed by decimation. The filter scheme of DWT is presented in Figure 2.2.

Figure 2. 2. Filter scheme of DWT

The filters includes a low pass filter, $G(z)$, and a high pass filter, $H(z)$. These filters are intimately related to the mother wavelet of the wavelet transforms. The high pass filter, $H(z)$, is usually considered as the “mother wavelet” and the output of the high pass filter are thus the wavelet coefficients.

By using a hierarchically decomposing structure, a function can be described in terms of a coarse overall shape, plus details that rang from broad to narrow (shown in Figure 2.3). Signal characteristics can be efficiently located in the space and frequency domains. Therefore, DWT is adequate for the study of non-stationary signals with both low frequency components and sharp transitions.

Optimization of Wavelet Network

Figure 2. 3. Multiresolution structure of DWT

As stated above, wavelet theory is a general mathematical tool primarily used for representing functions more efficiently. It can be utilized in signal processing field, such as de-noising, manipulating signals to extract or / and encode desired information, and so on. Since most of signals are wideband signals, the multi-resolution representation capability of wavelet analysis makes it an appropriate and efficient tool for signal processing.

2.3 Wavelet Network Based on Wavelet Theory

Wavelet decomposition reconstructs a signal by a series of scaled and translated wavelets. A function $f \in L_2(R)$ can be represented as an infinite sum of weighted wavelet expansions:

$$f = \sum_{j,k \in \mathbf{Z}} \beta_{jk} \psi_{jk}, \quad (2.2)$$

where a typical $\psi_{j,k}(x) = \left\{ 2^{\frac{j}{2}} \psi(2^j x - k) : j, k \in \mathbf{Z} \right\}$, j is the scale index, and k is

the translation index. We can see that the elements of wavelet basis are translated and dilated versions of the mother wavelet. However, to satisfy the constraints of basis, the

Optimization of Wavelet Network

wavelet set, $\{\psi_{jk}\}$, is very rigid, and irrelevant to the input data. Thus, it can be expected that the wavelet reconstructor will be more efficient if the wavelet “basis” is constructed with respect to the input data.

Therefore, based on the theory of discrete wavelet transform, Zhang and Benveniste (Zhang and Benveniste, 1992) tried to link the network coefficients with this appropriate transform. In 1992, they proposed wavelet network for approximation, which had an explicit link between the network coefficients and the discrete inverse wavelet transform. In this way, they took advantage of the concept of discretization of inverse wavelet transform to initialize the coefficients. The wavelet network structure is of the following form:

$$f(\mathbf{x}) = \sum_{i=1}^N w_i \psi(\text{diag}(\mathbf{s}_i)(\mathbf{x} - \mathbf{t}_i)) + \mathbf{c}^T \mathbf{x} + \bar{g} \quad (2.3)$$

where N is the number of wavelets,

w_i is the weight of the i th wavelet,

$\psi(\cdot)$ is the mother wavelet,

$\mathbf{t}_i$ is the translation factor of the i th wavelet,

$\mathbf{s}_i$ is the dilation factor of the i th wavelet,

diag denotes diagonal matrix,

$\mathbf{c}$ is a weight matrix from the input to the output,

Optimization of Wavelet Network

$\bar{g}$ is a bias to the output.

From Equation (2.3), one can see that the architecture of a wavelet network is exactly specified by the number of wavelets required for a given classification or regression application. The wavelet network with optimized structure will achieve better approximation and prediction capability than wavelet networks with other structures, and than some conventional signal processing approaches as well (Experimental results are given in section 2.5.). In this research, a pool of wavelets will be selected to be the candidate hidden neurons (wavelets), and then the optimized architecture is constructed by pruning these hidden neurons. In other words, with the optimization scheme from machine learning technique we can build up a wavelet reconstructor, which can achieve the same accuracy with less wavelets, i.e. with lower structural complexity.

2.4 Wavelet Network Pruning

As mentioned above, the optimized wavelet network can be constructed in this way: First, select some scaled and translated wavelets (regressors), and then, some network pruning algorithms were employed to select the significant regressors (Zhang, 1997).

2.4.1 Wavelet Pool Selection

A wavelet pool consists of discretely scaled and translated versions of a given mother wavelet ψ . Thus, our job is to determine the scaling and translation values of the mother wavelet. First, we should specify the discretization step sizes for both factors, s_0 for scaling discretization, t_0 for translation discretization. Typically, a dyadic lattice is employed. After that, we must decide the maximum scale level, j , of the wavelet network. Then, for each scale level (from 1 to j), we will calculate all possible

Optimization of Wavelet Network

translation values regarding the input data, $\mathbf{x}$. In this way, we finally construct the wavelet pool as follows:

$$\{\psi(\text{diag}(s_0^j)(\mathbf{x} - \mathbf{k}t_0)) : j \in \mathbf{Z}, \mathbf{k} \in \mathbf{Z}^d\} \quad (2.4)$$

where d is the input dimension. These initial values of scaling and translation factors will be updated in the training procedure.

2.4.2 Network Pruning by Orthogonal Least Square Algorithm

The number of wavelets, N_s , may be decided manually or automatically. Therefore, in an N -candidate wavelet pool, the goal is to find the optimal N_s wavelet regressors, which minimize the error between the real output and the expected output. There are two different ways to specify the number of hidden neurons. One way is to cluster the training examples and then assign a neuron to each cluster. The other way is to create a neuron for each training example and then prune the hidden neurons by example selection.

Orthogonal least square (OLS) algorithm (Chen, *et al.*, 1991) is to find the regressors, which provide the most significant contribution to approximation error reduction. It has been widely accepted as an effective method for regressor selection in RBF network (Chen, *et al.*, 1996; Chen, *et al.*, 1999; Gomm and Yu, 2000). The advantage of employing OLS is that the responses of the hidden layer neurons are decorrelated so that the contribution of individual candidate neurons to the approximation error reduction can be evaluated independently.

Let $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_L)^T$ be the output matrix corresponding to all the number L of training examples. $\mathbf{y}_i (i = 1, 2, \dots, L)$ is an M -dimensional vector, denoting number M of

Optimization of Wavelet Network

output units. We have $\mathbf{Y} = \mathbf{H}\mathbf{W} = (\mathbf{Q}\mathbf{A})\mathbf{W}$, where $\mathbf{Y}$, $\mathbf{H}$, $\mathbf{W}$ are $L \times M$, $L \times N$, $N \times M$ matrices, respectively. The selection of wavelets is equivalent to the selection of the most significant columns of $\mathbf{H}$. The matrix $\mathbf{H}$ can be decomposed into $\mathbf{Q}\mathbf{A}$. The $\mathbf{Q}$ is an $L \times N$ matrix with columns $[\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N]$, and $\mathbf{A}$ is an $N \times N$ upper triangular matrix which is as follows:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & & \vdots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ h_{L1} & \cdots & \cdots & h_{LN} \end{bmatrix} \quad (2.5)$$

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & \cdots & \cdots & a_{1N} \\ 0 & 1 & a_{23} & & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & 0 & 1 & a_{(N-1)N} \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad (2.6)$$

Only one column of $\mathbf{H}$ is orthogonalized at a time. At the K th iteration, the K th column is made orthogonal to each of the $K-1$ previously orthogonalized columns and the operation for $K=2, \dots, N$ is repeated. The computational procedure is as follows:

$$\begin{cases} \mathbf{q}_1 = \mathbf{h}_1, \\ a_{iK} = \frac{\mathbf{q}_i^T \mathbf{h}_K}{\|\mathbf{q}_i\|}, \quad 1 \leq i \leq K, \\ \mathbf{q}_K = \mathbf{h}_K - \sum_{i=1}^{K-1} a_{iK} \mathbf{q}_i, \end{cases} \quad (2.7)$$

Optimization of Wavelet Network

Here, $\|\mathbf{q}_i\|$ is the Euclidean norm. As we know, the significant wavelet selection is an iterative procedure. The iteration can be stopped by introducing a criterion -- Akaike's final prediction error (FPE) (Akaike, 1969). It is actually the mean square error (MSE) of the function $f(x)$, multiplied by a factor of penalty in terms of the number of parameters. The definition of FPE is:

$$FPE(f) = \frac{1 + n_p/N_t}{1 - n_p/N_t} \frac{1}{2N_t} \sum_{n=1}^{N_t} (f(x_n) - y_n)^2 \quad (2.8)$$

where n_p is the number of regressors in the estimator, y_n is the expected output of the training data x_n , and N_t is the length of training data. If the FPE value begins to increase monotonically, and the change of MSE between two consecutive iterations is less than a threshold value, ε , the pruning procedure will be stopped. The stopping criteria are given by

$$\begin{cases} FPE^{(k)} - FPE^{(k-1)} \geq 0, & \text{and} \\ \Delta MSE < \varepsilon \end{cases} \quad (2.9)$$

A formal statement of the network pruning algorithm for the selection of wavelets is given as follows:

Optimization of Wavelet Network

STEP 1 Initialization. Form the matrix $\mathbf{H}$ in equation (3) by the wavelet function responses of all the training samples.

STEP 2 First wavelet selection. Set $I=\{1,2, \dots, N\}$, then for each $i \in I$, calculate the approximation error with the $\mathbf{h}_i$. The column that provides the maximum error is selected as the first column of $\mathbf{Q}^{(1)}$.

STEP 3 Orthogonalization and wavelet selection. Let $K=2$.

STEP 4 Orthogonalize all remaining columns of $\mathbf{H}$ with all the columns of $\mathbf{Q}^{(k-1)}$ by equation (4).

STEP 5 Calculate the FPE and MSE values at each K th iteration. Go to STEP 7 if the stopping criteria are satisfied.

STEP 6 $K:=K+1$, go to STEP 4.

STEP 7 End.

Figure 2. 4. The pruning algorithm of wavelet network

The wavelets corresponding to the first K columns in $\mathbf{Q}^{(k-1)}$ will be selected to build the wavelet network.

2.4.3 Network Training

After the construction of the network, the training procedure is carried out based on a stochastic gradient algorithm to recursively minimize an objective error function using input / output pairs (Zhang and Benveniste, 1992).

Let $\Theta = (g, \mathbf{c}, w_i, \mathbf{t}_i, \mathbf{s}_i)$, then the objective is to minimize the following objective function,

$$E(\Theta, x_n, y_n) = \frac{1}{2} (f_{\Theta}(x_n) - y_n)^2. \quad (2.10)$$

Therefore, the partial derivatives of the above function with respect to each component of the parameter vector Θ are calculated. After that, the gradient is used to recursively adjust these parameters, as shown in the figure below.

Figure 2. 5. The diagram of the stochastic gradient algorithm

2.5 Experiments on Approximation

In this section, we will show the performance of the optimized wavelet network on a segment of speech signal. The sampling rate of the speech signal is 22.50 kHz, and the bits per sample are 16. The test segment has 1000 samples. The mother wavelet function is Mexican Hat wavelet, defined by

$$\psi(x) = \left(d - \|x\|^2\right) e^{-\frac{\|x\|^2}{2}}, \quad (2.11)$$

where $\|x\|$ is the Euclidean norm of x , and d is the dimension of input data.

Figure 2.6. Mother Wavelet (Mexican-hat wavelet)

The advantage of choosing this wavelet is that it is radial, smooth, and well concentrated in both time and frequency domains. The Figure 2.6 shows the form of this wavelet for the case $d = 1$.

2.5.1 Wavelet Network versus Other Models

In the previous work (Zhang and Benveniste, 1992), a performance comparison among the wavelet network, the neural network and the wavelet decomposition (i.e., with fixed dilations and the translations) has been conducted. The tested neural network follows Cybenko’s structure with the sigmoidal function

$$\sigma(x) = \frac{1 - e^{-x}}{1 + e^{-x}}.$$
 (2. 12)

The experimental results are as follows,

Table 2.1. Approximation results of $f(x) = (x_1^2 - x_2^2)\sin(0.5x_1)$

Method	Number of units	Number of parameters	Number of iterations	NSRMSE
Wavelet Network	49	442	40000	0.03395
Neural Network	225	1126	40000	0.29381
Wavelet Decomposition	961	962	40000	0.14289

In this table, NSRMSE represents the normalized square root of the mean square errors. Its definition will be given in the following section. The approximation results have shown the high approximation ability of wavelet network on some time-series data. In the following parts, the wavelet network will be utilized on some speech signals, and also compared to some traditional speech synthesis approaches.

2.5.2 Comparison among network pruning algorithms

To show the effectiveness of our OLS-based pruning, we compare its performance with those of the three different network pruning algorithms used in (Zhang, 1997).

Algorithm 1: Residual based selection. The idea of this method is to select, for the first stage, the wavelet in W that best fits the training data, and then repeatedly select the wavelet that best fits the residual of the fitting of previous stage.

Algorithm 2: Stepwise selection by orthogonalization. It first selects the wavelet in the W that best fits the training data, and then repeatedly selects the wavelet that best fit the training data while working together with the previously selected wavelets.

Algorithm 3: Backward elimination. This method first builds the regression with all the wavelets in W , and then eliminates one wavelet in each step while trying to increase the residual as less as possible.

Figure 2. 7. The selection of the optimal set of wavelets

Optimization of Wavelet Network

The OLS-based algorithm has a lower computational complexity, thanks to its normalized diagonal and Gram-Schmidt method to compute columns. In addition, by considering MSE and FPE in the selection procedure, we can get a tradeoff between the minimum set of wavelets and the minimum error caused. Figure 2.7 shows the selection procedure.

Table 2. 2. MSE and FPE of wavelet networks with different pruning algorithms

Scale Level	Algorithm 1		Algorithm 2		Algorithm 3		OLS	
	MSE	FPE	MSE	FPE	MSE	FPE	MSE	FPE
4	0.11059	0.05551	0.09383	0.04767	0.09383	0.04767	0.09083	0.04624
5	0.07510	0.03808	0.04639	0.02419	0.05195	0.02704	0.04627	0.02408
6	0.02446	0.01273	0.01153	0.00646	0.01260	0.00703	0.01197	0.00668
7	0.00558	0.00300	0.00280	0.00177	0.00282	0.00176	0.00282	0.00179
8	0.00249	0.00143	0.00070	0.00050	0.00100	0.00068	0.00073	0.00052
9	0.00158	0.00095	0.00011	0.00010	0.00013	0.00011	0.00005	0.00005

Figure 2. 8. MSEs during the different pruning procedures

Optimization of Wavelet Network

Table 2.2 and Figure 2.8 show the performance of the wavelet networks pruned by the three algorithms in (Zhang, 1997) and our OLS-based algorithm. Since the construction of wavelet pool is based on a dyadic lattice (i.e. the number of candidate wavelet is equal to $2^n - 1$, where n is the maximum scale level of wavelet network), to the 1000-sample speech signal, we selected the levels from level 4 to level 9. Thus, the number of candidate wavelets was from $(2^4 - 1)$ to $(2^9 - 1)$.

It can be seen that the OLS performs better than other network pruning algorithms. Based on the MSE in Table 2.2, the OLS has achieved an average 54% improvement over Algorithm 1, 8% over Algorithm 2, and 18% over Algorithm 3 in precision. Those methods, directly adding or eliminating regressors, do not take account of the correlation between regressors. In contrast, the OLS algorithm decouples the correlations among the responses of candidate wavelets. With correlation analysis, the individual contribution of each wavelet to the approximation error reduction can be calculated. That is why the OLS algorithm can achieve better results.

2.5.3 Performance Comparison

The approximation result of wavelet network on a segment of speech signals is shown in Figure 2.9. The mean square error after initialization is 0.01511 (Figure 2.9 (a)), and is reduced to 0.00349 after 20 iterations training (Figure 2.9(b)).

We can see, with the help of wavelet decomposition theory, the information of input data is effectively used to achieve a good initial state. Also due to the effective initialization, the training iteration is greatly reduced compared to common neural network models. The wavelet network reaches a relatively stable state rapidly.

Optimization of Wavelet Network

(a)

(b)

Figure 2. 9. Speech signal approximation by wavelet network. (a) Approximated signal after network pruning. (b) Approximated signal after training.

To investigate the prediction capability of wavelet network, we further employed the previous n samples as an input vector to predict the $n + 1$ sample. In the experiment, we selected $n = 140$, and used the first 400 samples as the training data. Therefore, the

Optimization of Wavelet Network

input data is a 140×400 matrix. After training, the predicted signal is shown in Figure 2.10, which quite matches the original signal.

Figure 2. 10. Predicted signal by wavelet network.

We are now in a position to compare the approximation and prediction capabilities amongst our proposed wavelet network and some other representative methods on speech signal processing. The performance comparison among wavelet network, PSOLA and the sinusoidal model is shown in Table 2.3. Here, the MSE is chosen to assess the approximation capabilities, whereas FPE is used to evaluate the prediction capabilities.

Table 2. 3. Performance comparison between wavelet network and other existing speech models

	MSE	FPE
Wavelet Network	0.00349	0.00257
PSOLA	0.00728	0.00369
Sinusoidal Model	0.01375	0.00817

Optimization of Wavelet Network

From Table 2.3, we can see that wavelet network outperforms both PSOLA and the sinusoidal model. The advantages of wavelet network include signal analysis at different scales, as well as high learning ability. Furthermore, with its flexible structure, it can achieve different accurate results by simply adjusting the number of wavelets. Since the sinusoidal model extracts tracks in frequency domain, its waveform approximation result in time domain is lower than both PSOLA and wavelet network. It must be noted that, only the optimized compact architecture is our concern in this research, and wavelet networks are time-consuming in training. So currently, wavelet networks are suitable for off-line signal processing, such as time series data forecasting, etc.

2.6 Summary

In this chapter, we have analyzed our optimized wavelet network in detail. Summaries and contributions are as follows: (1) From the point of view of machine learning, we analyze and construct wavelet network to achieve the high generalization capability with the minimization of the number of wavelets. (2) A new algorithm of constructing wavelet network is proposed with the OLS being employed to prune the hidden neurons. (3) Our proposed architecture is applied to a typical case of time series data -- speech signal in our initial experiments. The results show that the wavelet network pruned by OLS achieves the best approximation and prediction performance among the representative speech processing techniques.

CHAPTER

3

APPLICATION I: MICROCRYSTALLINE CELLULOSE GROUPING

3.1 Introduction

Neural networks collate knowledge by recognizing patterns and relationships in data, and learn an approximate nonlinear relationship through a continuous training process. During the training process, the inter-unit connections are optimized until the error in predictions is minimized and a certain level of accuracy is attained. Once the network is trained (i.e. the pattern is established), it can be fed with new input information to predict the output. Due to their powerful ability to unravel hidden mappings among data, neural networks have successfully been utilized in many applications in pharmaceutical research, ranging from interpretation of analytical data (data modeling), drug design (molecular modeling), dosage form design (optimization), to pharmacokinetic / dynamic modeling among others (Achanta, *et al.*, 1995; Hussain, *et al.*, 1991; Murtoniemi, *et al.*, 1994; Gasperlin, *et al.*, 1998; Takayama, *et al.*, 1999).

3.1.1 Importance of Microcrystalline Cellulose Grouping

Microcrystalline cellulose (MCC) is often regarded as the excipient of choice in the production of spheroids by extrusion-spheronisation. Its ability to control the amount of water required to achieve the desired plasticity in the wetted mass has been attributed to its unique interaction with water. However, the vast number of MCC grades available commercially has brought about variability in their performance during extrusion-spheronisation. This variability was brought about by the differences in their source materials and manufacturing process.

Grouping of MCCs is very useful as it allows the interchangeability of MCC grades without compromising their product quality. This grouping system will be based on the application of neural network and data clustering on the rheological properties of wetted masses containing MCC. As the rheological properties are inevitably related to the interaction of MCC with water, it can be used as an accurate predictor for the performance of MCC in extrusion-spheronisation.

3.1.2 Wavelet Network and Discrete Incremental Clustering

Neural networks provide linear algorithms capable of representing complex nonlinear mapping, and they can approximate any regular input sequence. Their learning ability from training data makes them an important tool for data analysis. On the other hand, there are different topologies of neural networks that may be used for data modeling. Among them, wavelet networks appear to be advantageous compared to conventional neural networks (Zhang, *et al.*, 1995; Jiao, *et al.*, 2001). However, in most cases many of the wavelets are redundant and only a small number of significant terms are necessary to describe a given nonlinear system with a given accuracy (Billings and Wei, 2005).

Application I: Microcrystalline Cellulose Grouping

Therefore, the optimized wavelet network with OLS pruning scheme can be more efficient for approximation tasks due to the less computational load.

However, though the wavelet network is efficient in modeling and understanding complicated relationships within a given data set, it is inadequate in situations where there is a requisite to achieve some form of objective clustering. The information derived from wavelet network alone would not be sufficient to gain an in-depth knowledge of the relationship between different MCCs. To better understand water-MCC interaction, an additional step of data clustering needs to be carried out. Data clustering involves the division of a given data set into smaller groups or clusters, each possessing certain similar characteristics. By clustering, hidden patterns or relationships may be uncovered and is immensely useful in many information retrieval areas including web mining, medical diagnostics, and marketing analysis.

A novel clustering technique known as discrete incremental clustering (DIC) is also introduced. DIC can determine the optimal number of clusters for a given data set (Tung, *et al.*, 2002) without the need to predefine the number of clusters. This is a distinct advantage as the total number of clusters for a given data set is usually unknown.

3.2 Wavelet Network Modeling

Wavelet network is a special feed-forward neural network supported by the wavelet theory (Zhang, 1997). By taking advantages of both the scaling property of wavelets and the effective learning mechanism of neural networks, it has emerged as a powerful regression estimator for nonparametric approximation. In addition, with the help of OLS pruning, the optimized wavelet network is proven to have the excellent approximation

Application I: Microcrystalline Cellulose Grouping

ability (Chen, *et al.*, 2004). Thus it is an ideal method for modeling our pharmaceutical observation data. The wavelet network structure is of the following form:

$$f(\mathbf{x}) = \sum_{i=1}^N w_i \psi(\text{diag}(\mathbf{s}_i)(\mathbf{x} - \mathbf{t}_i)) + \mathbf{c}^T \mathbf{x} + \bar{g} \quad (3.1)$$

where N is the number of wavelets,

w_i is the weight of the i th wavelet,

$\psi(\cdot)$ is the mother wavelet,

$\mathbf{t}_i$ is the translation factor of the i th wavelet,

$\mathbf{s}_i$ is the dilation factor of the i th wavelet,

diag denotes diagonal matrix,

$\mathbf{c}$ is a weight matrix from the input to the output,

$\bar{g}$ is a bias to the output.

Then, the structure of wavelet networks will be optimized by OLS pruning algorithm, which selects significant regressors from the wavelet candidate pool. For detailed experimental results please refer to section 2.5.

Training of the wavelet network was conducted on the torque data measured at different mixing times and water:MCC ratios for the 11 MCCs. A wavelet network with 20 hidden neurons was used to model the MCCs where the input vectors were mixing time and water:MCC ratio, and the output vector was torque measurement. Sampling data

Application I: Microcrystalline Cellulose Grouping

of a specific MCC was used in training the wavelet network where weights of the network were updated iteratively by the error back propagation algorithm. After the network was trained for a specific MCC, sampling data of other MCCs were entered to test the mean square errors (MSEs). The results from wavelet models were used to calculate the membership function values of DIC clusters.

3.3 Discrete Incremental Clustering

In DIC, a cluster grows by the expansion of its kernel, which is controlled by the plasticity parameter - β . The β value governs the extent to which a cluster expands its kernel in order to include a new data point that has been introduced. The β value of a cluster decreases as the cluster expands its kernel. The change of β in a cluster is modeled by the first quadrant of a cosine waveform. Figure 3.1. (a) shows the decrease of β from an initial value of 0.5.

The tendency parameter - TD is analogous to a cluster's willingness to grow. It maintains the relevance of a cluster and prevents it from incorporating too many data points that has low fitness (membership function value) to the cluster. TD is calculated according to the following equation:

$$TD_{i,j}^{new} = TD_{i,j}^{old} + (A - TD_{i,j}^{old}) \times (1 - \mu_{i,j}(x_i))^2, \quad (3.2)$$

where $\mu_{i,j}$ denotes the membership function of the nodes i, j , and $A = -0.5$. When TD is less than or equal to zero, the cluster stops expanding and sets its plasticity parameter β to zero (shown in Figure 3.1. (b)).

Application I: Microcrystalline Cellulose Grouping

The input threshold - IT , specifies the minimum membership value an input vector must have before it is considered as relevant or similar to any existing clusters. When the vectors are fed in, the first vector will be considered as the cluster center. With this cluster, a new input vector's membership function value will be calculated. If it is below the input threshold, the new data will be classified into the existing cluster. Then the winner cluster will grow a little according to the predefined β value. In addition, TD is calculated to represent the cluster's willingness to "grow". If the new data match the cluster exactly, the cluster will keep its "willingness" to grow. Otherwise, it will cease to grow further. If the membership function value is larger than IT , the new input vector becomes a new cluster.

For our purpose, a modified version of DIC was employed. The basis parameter set is defined as $\beta = 0.5$, $IT = 0.9$, and the membership function values are based on the results from wavelet networks.

Figure 3. 1. Parameters in DIC. (a) Modeling of plasticity parameter β and (b) Dynamics of tendency parameter TD

Application I: Microcrystalline Cellulose Grouping

3.4 Experimental Results

Mixer torque rheometry (MTR) is a reliable and direct method for assessing mixing resistance of moistened powder masses in granulation studies. Fifteen grams of MCC powder were added into the mixer bowl of the MTR and mixed for 30 seconds. The mean torque generated by the dry powder was recorded. Distilled water, corresponding to added water addition of 15 % w/w was added to the dry powder at intervals of 30 s for the first 10 additions and at intervals of 60 s for the remaining 9 additions. The entire mixing process was completed in 20 minutes and the mean torques generated for 19 additions were recorded. The test was performed in triplicates for each MCC.

The amount of water used ranged from 75% to 220% w/w of MCC powder. All the water was evenly distributed throughout the dry powder while the blades turned. The total mixing time was 46 min for all the MCCs and respective torque values were logged at regular intervals. Water addition amounting to 200% w/w was too wet to produce any appreciable torque reading for Avicel PH 301 and PH 302; hence, moisture contents equal to and greater than 200% w/w were not investigated for these 2 MCCs. Triplicates were performed for each MCC at all the mixing time and water additions tested.

Rheological profiles of the MCCs across the entire range of water additions are presented in Figure 3.2.

Application I: Microcrystalline Cellulose Grouping

Figure 3. 2. Validation of measured torque with water: MCC ration for (♦) Avicel PH 101, (■) Avicel PH 102, (▲) Avicel PH 301, (×) Avicel PH 302, (*) Ceolus KG 801, (°) Celex 101, (+) Emcocel 50 M, (◇) Prosolv 50 M, (Δ) Viva Pur 101, (◆) Pharmacel 101, and (■) Pharmacel 102

MSE values obtained from the optimized wavelet network (Table 3.1) were subsequently used in the calculation of membership function values required by the DIC algorithm for grouping the MCCs. Two sets of stringency were defined for the DIC parameters, β and IT . They are listed Levels One and Levels Two in Table 4.2 with the latter level having a more lenient criterion for the IT value.

Application I: Microcrystalline Cellulose Grouping

Table 3. 1. Calculated MSE by modeling the torque measurements of MCCs using wavelet network I

Testing MCC	Training MCC										
	Avicel	Avicel	Avicel	Avicel	Ceolus	Celex	Emcocel	Prosolv	Viva	Pharmacel	Pharmacel
	PH 101	PH 102	PH 301	PH 302	KG 801	101	50 M	50 M	Pur 101	101	102
Avicel PH 101	0	0.123	0.676	0.646	0.126	0.102	0.060	0.076	0.196	0.192	0.190
Avicel PH 102	0.123	0	0.574	0.502	0.019	0.046	0.044	0.047	0.039	0.037	0.035
Avicel PH 301	0.919	0.815	0	0.143	0.874	0.965	0.752	0.802	0.739	0.778	0.834
Avicel PH 302	0.887	0.716	0.145	0	0.751	0.874	0.707	0.765	0.629	0.657	0.694
Ceolus KG 801	0.129	0.016	0.614	0.525	0	0.030	0.052	0.064	0.050	0.043	0.032
Celex 101	0.095	0.047	0.709	0.633	0.033	0	0.080	0.089	0.101	0.080	0.065
Emcocel 50 M	0.060	0.045	0.535	0.500	0.054	0.082	0	0.024	0.089	0.087	0.098
Prosolv 50 M	0.074	0.047	0.555	0.525	0.064	0.086	0.024	0	0.077	0.080	0.090
Viva Pur 101	0.196	0.038	0.512	0.433	0.048	0.094	0.089	0.077	0	0.011	0.016
Pharmacel 101	0.187	0.031	0.542	0.452	0.044	0.078	0.083	0.084	0.014	0	0.012
Pharmacel 102	0.182	0.029	0.579	0.478	0.035	0.063	0.094	0.097	0.021	0.014	0

Table 3. 2. Grouping of MCCs I

Group	Level one	Level two
	$\beta = 0.5$ $IT = 0.9$	$\beta = 0.5$ $IT = 0.8$
1	Celex 101, Ceolus KG 801	Celex 101, Ceolus KG 801, Avicel PH 102
2	Emcocel 50 M, Prosolv 50 M	Emcocel 50 M, Prosolv 50 M
3	Pharmacel 101, Pharmacel 102, Viva Pur 1010	Pharmacel 101, Pharmacel 102, Viva Pur 1010
—	Avicel PH 101	Avicel PH 101
—	Avicel PH 102	Avicel PH 301
—	Avicel PH 301	Avicel PH 302
—	Avicel PH 302	—

Application I: Microcrystalline Cellulose Grouping

MCCs were classified into 6 or 7 groups depending on the leniency of the predefined threshold parameters. Classification results were also able to identify bulk and tapped densities as major factors governing water-MCC interaction. Finally, an objective grouping of MCCs was achieved.

Using wavelet network and the novel clustering technique, a pre-formulation tool for the grouping of MCCs was developed. A preliminary assessment of new or unknown MCCs can be readily attained by training the wavelet network with their torque measurements to group them. Other key properties that can dictate water-MCC interaction could be discovered and recognized with a high level of assurance. This knowledge can be extended to explore the relationship between the quality of formed spheroids and the rheological profiles. From there, the classification results obtained can be extrapolated to predict the performance of these MCCs in spheroid production. In doing so, a considerable amount of time and effort can be saved during the pre-formulation and the later optimization phases.

3.5 Improvement and Validation

To testify and refine our clustering system, more MCCs should be added to the list of existing ones. This will not only improve and validate the precision and accuracy of the classification but also help to discover other important physical attributes involved in the interaction between water and MCC. These 3 additional MCCs are Comprecel M101, Comprecel M102 and Ceolus KG 802

Additionally, a visualization algorithm, multidimensional scaling (MDS), will be employed to provide users a direct visual result of MCC clustering.

Application I: Microcrystalline Cellulose Grouping

3.5.1 Multidimensional scaling

Multidimensional scaling is a technique that involves the analysis of similarity or dissimilarity between a given set of factors or objects (Borg, *et al.*, 1997). Every factor will be represented as a point in the geometric space and the distance between points are indicative of the differences between them. In short, it provides a visual representation of the distribution of various factors or parameters in a two-dimensional space. Although MDS was initially developed for the field of computational and statistical studies, it was subsequently employed in diverse disciplines including socio-epidemiology, psychotherapy and biological sciences to analyze patterns and trends that would otherwise be too time-consuming or difficult to perform using existing methods.

MDS begins with a distance matrix $\mathbf{R}$ of K data vectors $\{\mathbf{X}_1, \dots, \mathbf{X}_K\}$ and generates a set of corresponding vectors $\{\mathbf{Y}_1, \dots, \mathbf{Y}_K\}$ in a 2D or 3D space (Meulman, 1986; Cox, 1994). In the projection procedure, the difference between the two distance matrices is measured by a distortion function $E(\mathbf{r})$. The relative positions among vectors will be preserved by minimizing the error function. MDS algorithm can be implemented in the following steps:

Step 1: Distance of data vectors in the feature space is calculated using the following equation.

$$R_{ij} = \|\mathbf{X}_i - \mathbf{X}_j\|, \quad (3.3)$$

Step 2: A dissimilarity (similarity) matrix is constructed as follows using the MSE values from the wavelet network.

Application I: Microcrystalline Cellulose Grouping

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ R_{21} & \ddots & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ R_{K1} & \dots & R_{K(K-1)} & 0 \end{bmatrix} \quad (3.4)$$

Step 3: A non-negative function known as distortion function, $E(\mathbf{r})$, is defined and minimized according to Eqn. 3.

$$E(\mathbf{r}) = \sum_{i>j}^K (R_{ij} - r_{ij})^2 \quad (3.5)$$

where $r_{ij} = \|\mathbf{Y}_i - \mathbf{Y}_j\|$ is the distance in the target space, $\mathbf{Y} \in \mathbf{R}^2$ or $\mathbf{R}^3$ and $\mathbf{R}$ is a real number.

3.5.2 Experimental Results

Three additional MCC grades were assigned to two of the existing groups after their rheological data were added to the previous data set used in the training of the neural network and data clustering (Table 3.4). The sensitivity of this system was demonstrated by the different grouping result for Ceolus KG 802 when the levels of stringency for clustering were changed slightly even though their differences were believed to be minimal.

Application I: Microcrystalline Cellulose Grouping

Table 3. 3. Calculated MSE by modeling the torque measurements of MCCs using wavelet network II

Testing MCC	Training MCC													
	Avicel PH 101	Avicel PH 102	Avicel PH 301	Avicel PH 302	Ceolus KG 801	Ceolus KG 802	Celex 101	Emcocel 50 M	Prosolv 50 M	Viva Pur 101	Pharmacel 101	Pharmacel 102	Comprecel M 101	Comprecel M102
Avicel PH 101	0	0.123	0.908	0.875	0.127	0.065	0.096	0.061	0.076	0.196	0.188	0.180	0.188	0.216
Avicel PH 102	0.130	0	0.809	0.711	0.014	0.068	0.046	0.048	0.040	0.032	0.033	0.033	0.037	0.045
Avicel PH 301	0.676	0.576	0	0.147	0.615	0.656	0.709	0.536	0.556	0.514	0.544	0.581	0.565	0.518
Avicel PH 302	0.663	0.510	0.142	0	0.534	0.622	0.648	0.509	0.531	0.441	0.460	0.488	0.466	0.414
Ceolus KG 801	0.128	0.017	0.873	0.751	0	0.049	0.035	0.053	0.064	0.049	0.043	0.034	0.037	0.057
Ceolus KG 802	0.065	0.071	0.877	0.811	0.048	0	0.034	0.053	0.082	0.128	0.116	0.107	0.111	0.150
Celex 101	0.098	0.046	0.963	0.870	0.033	0.033	0	0.080	0.086	0.097	0.078	0.064	0.084	0.120
Emcocel 50 M	0.061	0.044	0.746	0.698	0.052	0.054	0.081	0	0.025	0.090	0.084	0.095	0.092	0.095
Prosolv 50 M	0.075	0.046	0.803	0.764	0.064	0.083	0.088	0.024	0	0.076	0.083	0.096	0.083	0.091
Viva Pur 101	0.197	0.039	0.739	0.629	0.051	0.132	0.101	0.089	0.077	0	0.014	0.021	0.020	0.015
Pharmacel 101	0.192	0.037	0.782	0.661	0.043	0.117	0.080	0.087	0.080	0.011	0	0.014	0.018	0.017
Pharmacel 102	0.189	0.034	0.836	0.696	0.031	0.104	0.065	0.097	0.090	0.015	0.012	0	0.015	0.025
Comprecel M101	0.180	0.043	0.809	0.658	0.037	0.115	0.084	0.088	0.083	0.018	0.020	0.019	0	0.013
Comprecel M102	0.217	0.052	0.743	0.586	0.058	0.154	0.125	0.097	0.093	0.015	0.021	0.030	0.014	0

Application I: Microcrystalline Cellulose Grouping

Table 3. 4. Grouping of MCCs II

Group	Level One		Level Two	
	$\beta = 0.5$	$IT = 0.9$	$\beta = 0.5$	$IT = 0.8$
1	Avicel PH 102, Ceolus KG 801		Celex 101, Ceolus KG 801, Ceolus KG 802*, Avicel PH 102	
2	Emcocel 50M , Prosolv 50M		Emcocel 50M, Prosolv 50M	
3	Pharmacel 101, Pharmacel 102, Viva Pur, Comprecel M101*, Comprecel M102*		Pharmacel 101, Pharmacel 102, Viva Pur, Comprecel M101*, Comprecel M102*	
—	Celex 101		Avicel PH 101	
—	Ceolus KG 802*		Avicel PH 301	
—	Avicel PH 101		Avicel PH 302	
—	Avicel PH 301		—	
—	Avicel PH 302		—	

* Additional MCCs.

Using the MDS visualization algorithm, the clustering results in 2D space, with and without PH 301 and PH 302, were presented in Figure. 3.3. Clearly, the 2 high density MCCs, PH 301 and PH 302, were distinctly different from the other MCCs (Figure. 3.3 (a)). Hence, compared to them, the other 12 grades of MCC seemed to be alike. However, the differences between the remaining 12 MCC grades were better illustrated when these 2 MCCs were excluded from the same 2D space (Figure 3.3 (b)).

Figure 3. 3 (b) shows the high degree of similarity between Comprecel M101, Comprecel M102, Pharmacel 101 and Pharmacel 102 and Viva Pur. Similarly, Emcocel and Prosolv are grouped together based on their close relative proximity. On the other hand, Ceolus KG 802 was comparatively further apart from PH 102 and Ceolus KG 801, thus they were initially placed in separate groups under Level One stringency as shown in

Application I: Microcrystalline Cellulose Grouping

the following section. A relaxation in the stringency of the clustering parameters to Level Two resulted in the assignment of Ceolus KG 802 into the group with the 2 abovementioned MCCs, as well as Celex. This observation reflected the allowance conferred to the stringency of the clustering parameters. In the case of PH 101, it was deemed to be too far apart from the other MCCs to be assigned into any groups.

(a)

(b)

Figure 3. 3. Visualization of MDS algorithm. a) Clusters without PH302, and PH301;
b) Clusters with all MCCs

3.6 Summary

The combination of the optimized wavelet network and DIC offers the opportunity for clustering MCCs into discrete groups that possess equivalent or comparable performance; thus, providing a basis for suggesting interchangeability among MCCs within the same group. This is especially advantageous for formulation scientists during the preformulation and optimization phases of product development, since it provides a preliminary assessment of new or unknown MCCs. In addition, key properties that control the performance of MCC in water-MCC interaction can be discovered.

Furthermore, the utility and value of MDS as a pre-formulation tool in the grouping of an expanded range of MCC grades was exemplified in this study. On its own, MDS provided a rapid visual comparison between the wide varieties of MCCs grades. However, its potential can be better harnessed when used in combination with other computational techniques. Our proposed pre-formulation tool which incorporates the wavelet network, data clustering and MDS, was capable of assigning newly introduced or unknown MCCs into existing groups based on their torque rheological properties. The latter, was in turn, governed by the unique interactions between their physical properties.

Other advantages of this pre-formulation tool include its simplicity of usage and flexibility in the type of data set used. By subjecting MCC grades to varying degrees of moisture, the rheological profiles of their wetted masses were able to reflect their differences and enabled their assignment into groups where the members possessed equivalent and comparable water-MCC interaction. This signifies a possible prediction of their performance in processes such as extrusion-spheronisation and confers flexibility or interchangeability in the selection of MCC grades without having to perform a laborious

Application I: Microcrystalline Cellulose Grouping

set of actual trial formulation studies. In addition to torque rheological profiles, other parameters or characteristics can also be used to group the MCC grades. This feature will undoubtedly be of great value to pharmaceutical sciences, especially in the field of pre-formulation.

Application I: Microcrystalline Cellulose Grouping

CHAPTER

4

APPLICATION II: TALKING HEAD

4.1 Introduction

Speech with facial information provides people a natural and comfortable way to communicate (Hong, *et al.*, 2002). With facial information added, the intelligibility of speech can be improved significantly, especially when the auditory speech is degraded by noise, or hearing impairment (Massaro, 1987). The visual information is especially helpful with front phonemes whose articulation we can see, such as labiodentals and bilabials (Beskow, *et al.*, 1997). For example, intelligibility between /b/ and /d/ increases significantly with visual information (Santen, *et al.*, 1997). Among various facial movements, such as smiling, eye blinking, head nodding, and eyebrow rising, the lip motion is the most important part in giving useful information in addition to speech. Therefore a talking face with lip movement will be an ideal solution for human-machine interaction. It can be used as a tool for interactive training of speech reading. It may also be used in information systems in public and noisy environments, such as airports, train stations and shopping centers. If it is possible to make the talking head look like some

Application II: Talking Head

certain individual, it may be utilized in videoconferencing or used as a synthetic newsreader. Multimedia is also an important application field of talking heads.

Meanwhile, many reports have shown that synthetic talking head systems are commonly language dependent. As Mandarin is likely to be the largest telecommunication market in the world, the advancement of Mandarin audio-visual speech system will bring a great commercial benefit as well as social benefit. In spite of that, the audio-visual speech synthesis for Mandarin has rarely been reported so far.

Before the emergence of various lip models, the easiest approach to displaying lip animation is to use pre-stored images to represent all the possible shapes and to combine these with some morphing methods for smooth boundaries, similar to concatenative speech synthesis (Bregler, 1997), for example. This method may be quite successful for some limited applications, but is very inflexible, since there is no way to control different facial features independently of each other.

In contrast to this concatenative approach, lip modeling is more flexible and becomes current mainstream. Among various models, the muscle-based model is perhaps theoretically the most elegant because it controls lip movements directly (Waters, 1987). However, the difficulties with muscle-based implementation make the parametric model more popular. Currently, most of the talking head systems are based on parametric models, commonly three-dimensional models (Hong, *et al.*, 2002; Morishima, *et al.*, 1991; Park, 1982; Yamamoto, *et al.*, 1998). In a parametric model, some control points will be determined to dynamically describe the lip shape. We employed a control point set similar to Morishima's work (Morishima, *et al.*, 1991). Users can define the region of lips, and then the control points will be automatically selected based on the regions. Thus, the

Application II: Talking Head

problem here is how to adjust these control points in accordance with other media, or how to synchronize the lip motion with other media.

The significance of media synchronization is undoubtable. Since, without the coherence of lip motions and speech signals, the intelligibility of speech may be even decreased. McGurk has shown that human beings can perceive /da/ with audio /ba/ and visual /ga/ stimuli (McGurk and MacDonald, 1976). This phenomenon implies that there is an effective audio–visual integration mechanism in the human perception of speech sounds. Therefore, the driving scheme must be constructed reasonably.

In fact, both text and speech can be used as control inputs for animation synthesis. In the speech-driven animation, the critical issue is how to build up a mapping from acoustic frames to animation frames. However, the mapping from acoustic frames to animation frames is not a simple one-to-one mapping (Yehia, *et al.*, 1998). Therefore this media mapping is supposed to represent the hidden pattern between speech signals and lip features. As an excellent tool for non-parametric estimation, the wavelet network (Zhang and Benveniste, 1992) is surely an eligible candidate for the task.

On the other hand, in the text-driven lip animation, the process generally involves determining a mapping from text (orthographic or phonetic) onto visemes (visual counterpart of phonemes) by means of vector quantization or rule-based conversion. Obviously, it is not possible to construct a precise mapping by text inputs alone. Thus, we also proposed an alternative driving scheme for lip synchronization. It makes use of both text and speech information, and results in a more precise text-viseme conversion as well as high tolerance of low quality speech signals.

4.2 Lip Modeling

a: Standard point

b: Control point of top-edge of upper lip

c: Control point of inside-edge of upper lip

d: Control point of inside-edge of lower lip

e: Control point of bottom-edge of lower lip

f: Control point of jaws

g: Control point of lip side

Figure 4. 1. Seven control points around lips and one standard points at the bottom of nose

In lip modeling, it is not necessary to control all the lattice points on the 3D model independently. These lattice points on the lips can be adjusted according to seven specific feature (control) points (shown in Figure 4.1).

Application II: Talking Head

To achieve the control point set, we have implemented a user interface for interactive region division. Since the three-dimensional view point always causes some overlap, which makes regions hard to select, we map the cylindrical projection coordinates of the three-dimensional geometry into two-dimensional space. The following figure shows the user interface of region definition based on a two-dimensional mesh model. After the definition of lips, the control points on the lips will be automatically determined. The zoom in and zoom out functions are also provided so that even a small detailed region can be defined.

Figure 4. 2. User interface for region definition

4.3 Speech-Driven with Wavelet Network

As a powerful regression estimator for nonparametric approximation, wavelet network takes advantages of both the scaling property of wavelets and the effective learning mechanism of neural networks (Zhang and Benveniste, 1992). It is a special feed-forward

Application II: Talking Head

neural network supported by the wavelet theory. The wavelet network structure is of the following form:

$$f(\mathbf{x}) = \sum_{i=1}^N w_i \psi(\text{diag}(\mathbf{s}_i)(\mathbf{x} - \mathbf{t}_i)) + \mathbf{c}^T \mathbf{x} + \bar{g} \quad (4.1)$$

where N is the number of wavelets,

w_i is the weight of the i th wavelet,

$\psi(\cdot)$ is the mother wavelet,

$\mathbf{t}_i$ is the translation factor of the i th wavelet,

$\mathbf{s}_i$ is the dilation factor of the i th wavelet,

diag denotes diagonal matrix,

$\mathbf{c}$ is a weight matrix from the input to the output,

$\bar{g}$ is a bias to the output.

With the help of Orthogonal Least Square (OLS) pruning, the wavelet network is proven to have an excellent generalization ability (Chen, *et al.*, 2004). Thus it is an ideal method for constructing the media mapping.

One of the most significant advantages of speech-driven schemes is in that the complicated linguistic concerns can be greatly reduced. The driving scheme will be much straightforward and compact. In our scheme, each acoustic frame is converted to 16 dimensional linear prediction (LPC) coefficients. Then, Cepstrum distance is used, since

Application II: Talking Head

it is good for distortion measurement of voice spectrum parameters (Morishima, *et al.*, 1991).

Therefore, the input layer of wavelet network has 16 neurons corresponding to the dimensions for LPC Cepstrum parameters. The hidden layer has 48 neurons, which is decided by the OLS pruning algorithm. The output is the moving quantity of these feature points, which are normalized according to the standard point. The audio-visual training data are required to train the wavelet network. The training data is collected in the following way. A subject is asked to pronounce each viseme. The pictures are taken serially and very quickly with a fixed time interval, generally 1/30 second. The image training data are sampled at the 500×320 pixel size, and are grayscale images. All of the image control points are picked manually. At the same time, the speech signals are recorded, and segmented with the same frame size as the image data. These segmental signals are further analyzed to calculate their LPC coefficients.

4.4 Hybrid Driving Scheme

To use text as one of the driving input, a deep understanding regarding the target language pronunciation system, e.g. the orthographic/phonetic transcription rules, is necessary. In this research, we focus on Mandarin.

Mandarin Chinese is a tonal and syllable-based language (Chen, *et al.*, 2003). In comparison with English phonetics, Mandarin Chinese phonetics have some special features: (1) Every Chinese character is a monosyllable character; (2) A typical Chinese syllable consists of a consonant and a vowel; (3) Every syllable has a tone. Generally, there are four different tones in a Mandarin pronunciation system; (4) One character may

Application II: Talking Head

have more than one syllables, and vice versa. (5) The domain of the tone is the syllable, but tones of nearby syllables often affect each other.

Table 4. 1. Chinese phoneme combination scheme. The first column is the vowel, and the first row the consonant. The number of tones for a phoneme combination is given in the corresponding cell.

		b	c	ch	d	f	g	h	j	k	l	m	n	p	q	r	s	sh	t	w	x	y	z	zh
a	5	5	2	4	5	4	4	4		2	5	5	5	4			3	3	3	5		5	3	4
ai	4	4	4	3	3		3	4		3	2	3	2	4			2	3	4	3			3	4
an	3	3	4	4	3	4	3	4		3	3	4	4	3		2	3	3	4	4		4	5	3
ang	3	3	2	4	3	4	3	3		3	4	2	3	4		4	3	4	4	4		4	3	3
ao	4	4	3	4	4		3	4		3	4	4	4	4		3	3	4	4			4	4	4
e	5		1	3	2		4	3		4	3	1	3			2	1	4	1			4	2	5
ei	3	4			1	4	1	1			5	3	2	3				1	1	4			1	1
en	2	3	2	5		4	4	3		2		4	1	3		3	1	4		4			2	3
eng		4	3	4	3	4	3	3		1	4	4	1	4		2	1	4	1	3			2	3
er	3																							
i		4	4	4	4				4		5	4	4	4	4	1	4	5	4		4	4	3	4
ia				1					5		1				3						3			
ian		4		3					3		3	3	4	4	4				4		4			
iang									3		3		2		4						4			
iao		3		3					4		4	4	2	4	4				4		4			
ie		4		2					5		4	2	2	2	4				3		4			
in		2							3		3	2	1	4	4						3	4		
ing		3		3					3		4	3	3	2	4				4		4	4		
iong									2						1						2			
iu				1					3		4	1	4		3						3			
o	3	5			1					1	4			4						4		2		
ong			2	4	3		3	4		3	4		2			2	3		4			4	3	3
ou	3		1	4	3	1	3	3		3	5	3	1	3		2	3	4	4			4	3	4
u		4	3	4	4	5	4	4	4	3	5	3	3	4		3	3	4	4	4			3	4
ua							3	3		3								3					2	
uai				3			3	3		2								3					3	
uan			3	4	3		3	4	3	2	3		1		4	1	2	2	4		4	4	3	3
uang				4			3	4		4								2						3
ue									4		2		1		3						4	2		
ui			3	2	2		3	4		4						3	4	3	4				3	2
un			4	3	3		2	3	2	3	3				2	1	2	2	4		3	4	2	2
uo			4	2	4		4	4		1	5		2			1	2	2	4				4	2
ü										3			2		5						5	4		
total	38	59	45	65	63	31	59	65	48	49	92	59	59	60	49	30	40	60	69	35	51	57	49	62

We can see the pronunciation of Chinese characters is relatively regular. All possible Chinese phoneme combinations are listed in Table 4.1. The total number of the phoneme combination of the consonants and vowels is about 406. In some cases, syllables are only formed by vowel (the second column in Table 4.1). The number in the cell stands for the number of tones for the phoneme combination. Therefore, the total number of syllables (phoneme combination plus tones) is 1294.

Application II: Talking Head

Table 4. 2. Phoneme to viseme conversion of basic phonemes

Phonemes	Viseme groups
b, p, m	vtcB
c, s, z	vtcC
ch, sh, zh, r	vtcCh
d, t, g, h, k, l, n	vtcD
f	vtcF
j, q, x, y	vtcJ
w	vtcW
a, ai, an, ang,	vtvA
e, ei, en, eng, er	vtvE
i, in, ing	vtvI
o, ong, ou	vtvO
u, ü	vtvU

Table 4. 3. Phoneme to viseme conversion of multi-vowel phonemes

Phonemes	Viseme combination
ao	vtvA, vtvO
ia, ian, iang	vtvI, vtvA
iao	vtvI, vtvA, vtvO
ie	vtvI, vtvE
iong, iu	vtvI, vtvO
ou	vtvO, vtvU
ua, uan, uang, uai	vtvU, vtvA
ue	vtvU, vtvE
ui, un	vtvU, vtvI
uo	vtvU, vtvO

Application II: Talking Head

On the other hand, phonemes with similar mouth shapes are categorized into one viseme group (shown in Tables 4.2, and 4.3). There are also two supplementary viseme groups, *vtIntermediate* and *vtClosed*.

Therefore, for each Chinese input character, its corresponding phoneme combination will be found in a Chinese character dictionary first. Then the phoneme combination is decomposed into an optional consonant part and a vowel part according to the table of character-phoneme conversion (Table 4.1). Consequentially, each phoneme is converted to a viseme (or visemes) by the mapping rules defined in Table 4.2 and Table 4.3. Commonly, a typical Chinese character will be converted into three corresponding visemes, a consonant viseme, a vowel viseme, and an end-viseme (*vtIntermediate* or *vtClosed*).

After the conversion of text to viseme, corresponding speech signals will be analyzed to locate the frame position for each viseme. For most cases, zero crossing rate is an effective way to separate vowels and consonants (Rabiner, *et al.*, 1978), especially for unvoiced fricatives (e.g. “sh” or “f”). Thus, zero crossing rate is employed to find the boundaries between consonants and vowels (shown in Figure 4.3). After the separation, the key frame of consonant viseme is simply located at the center of consonant part, whereas the frame of vowel viseme is located at the peak of energy envelope, followed by an end-viseme at the end of frames.

Application II: Talking Head

(a)

(b)

(c)

Figure 4. 3. Audio signal analysis on “sha”: a) zero-crossing rate, b) waveform, c) STFT coefficients

For those consonants unsuitable for zero crossing rate method, and multi-vowel phonemes, Short Time Fourier Transform (STFT) coefficients are employed to find the boundaries between vowels or between consonants and vowels. It is the most widely used method for studying signals. STFT decomposes a signal into small time segments, $s_m(n)$, by a window function $h(n - t_m)$, and uses Fourier transform to analyze each time segment to ascertain the frequencies that exist in that segment:

Application II: Talking Head

$$\begin{cases} s_m(n) = h(n - t_m)s(n) \\ s_m(\omega) = \frac{1}{\sqrt{2\pi}} \int e^{j\omega n} s(n)h(n - t_m)dn \end{cases} \quad (4.2)$$

The totality of such spectra indicates how the spectrum is varying in time. It overcomes the deficiency of Fourier transform in time-frequency analysis. It can be used to locate the sites of some events on time scale. With the help of STFT, the boundary of consonants and vowels can be found. After that, the key-frame of each viseme is located by the same rules as mentioned above.

4.5 Experiments and Results

Our proposed methods were integrated into a web browser system, which could provide users an online service of reading news (shown in Figure 4.4). A Mandarin Text-to-Speech (TTS) engine with wavelet enhanced Pitch Synchronous Overlap Add (PSOLA) algorithm was utilized to cooperate with the lip synchronization system.

Figure 4.4. The synthetic talking head system

Application II: Talking Head

The synthetic talking head has successfully delivered visual information to users. With the definition of several regions on lips, the whole lips were controlled separately and cooperated precisely. In addition, the speech-driven with wavelet network generated natural lip motions. Although at the training stage, it is quite time-consuming. At run time, it is able of providing a real-time synthesis of lip shape.

Meanwhile, the hybrid driving scheme also generated smooth and natural lip motions based mainly on the text information. It provides another option to convert text information to mouth shapes with 100% correctness thanks to the prior knowledge of Mandarin pronunciation rules. More importantly, it reduces the dependency of viseme selection on speech signal by making use of the prior knowledge. Thus, it is capable of producing high quality lip animation based on low quality speech signals.

In short, both driving schemes successfully provided a way to drive lip motions by using other media sources. They can be employed in different cases depending on available source media.

4.6 Summary

This chapter presents an integrated framework for systematically building a real-time synthetic talking head system. A useful region definition system for lip modeling has been described, along with the presentation of a novel approach for driving lip motion that makes use of wavelet network and a hybrid driving scheme for Mandarin. Based upon the proposed framework, a web browser with a virtual newsreader has been developed to provide a service of reading web articles. Our future work includes incorporating more facial information, such as emotion information, into current parametric model, as well as adding in more language options.

CHAPTER

5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

In this research, the construction of wavelet networks with optimized structure is studied thoroughly, and our methodology has been successfully applied to pharmaceutical microcrystalline celluloses grouping, and a talking head system as well. The work that has been reported in this thesis can be summarized as follows:

1. The wavelet network has been analyzed from the point of view of machine learning. The construction of the optimized wavelet network is fulfilled by the selection of the optimal set of hidden neurons from a wavelet pool. Our experimental results have shown that the OLS algorithm can obtain wavelet networks with higher approximation and prediction abilities than the originally proposed pruning algorithms. Based on the MSE in Table 2.2, the OLS has achieved an average 54% improvement over Algorithm 1, 8% over Algorithm 2, and 18% over Algorithm 3 in precision.
2. The optimized wavelet network has been employed to model MCCs. Then, a clustering algorithm, DIC, has further been incorporated to group the MCCs without prior knowledge of the number of clusters. Finally, MDS algorithm has

Conclusions and Future Work

been utilized to give users a more direct and visual representation of the grouping results.

3. A synthetic audio-visual system for Mandarin has been implemented. A hybrid driving scheme making use of both text and audio information has been proposed to generate the precise visual mapping.
4. An audio-driven algorithm has been integrated in the Cel animation. It can automatically adjust the motion index of inbetweening frames according to signals, and consequentially simplify the synchronization process.

5.2 Future Work

The future research includes:

1. As we all know, support vector machines have the high generalization capability, thanks to their structural risk minimization, which strikes the balance between low empirical risk and low structural complexity. Thus ideas from structural risk minimization may be borrowed to refine the construction procedure of the optimized wavelet network.
2. A faster training algorithm can be investigated to speed up the training procedure. In addition, some specialized hardware may be considered to make the wavelet network suitable for real-time applications.
3. The synthetic audio-visual system can include more language options. Also by integrating some speech recognition techniques, the system is expected to be an ideal solution to human-machine interface.

APPENDIX A: GLOSSARY

CWT	Continuous wavelet transform
DIC	Discrete incremental clustering
ERM	Empirical risk minimization
FPE	Akaike’s final prediction error criterion
FWN	Fuzzy wavelet network
LPC	Linear predictive coding
MCC	Microcrystalline cellulose
MDS	Multidimensional scaling
MLP	Multi-layer perceptron
MRA	Multiresolution analysis
MSE	Mean square error
MTR	Mixer torque rheometry
NSRMSE	Normalized square root of the mean square errors
OLS	Orthogonal least square
PSOLA	Pitch synchronous overlap add
RBF	Radial basis function
STFT	Short time Fourier transform
SRM	Structural risk minimization
TTS	Text-to-speech

REFERENCES

Achanta, A. S., J. G. Kowalski, and C. T. Rhodes, "Artificial Neural Networks: Implications for Pharmaceutical Sciences," *Drug Development and Industrial Pharmacy*, vol. 21, no. 1, pp. 119–155, 1995.

Agatonovic-Kustrin, S. and R. Beresford, "Basic Concepts of Artificial Neural Network (ANN) Modeling and Its Application in Pharmaceutical Research," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 22, no. 5, pp. 717-727, 2000.

Ahmadi, S. and A.S. Spanias, "Cepstrum-Based Pitch Detection Using a New Statistical V/UV Classification Algorithm," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 333-338, May 1999.

Akaike, H., "Fitting Autoregressive Models for Prediction," *Annals of the Institute of Statistical Mathematics*, vol. 21, pp. 243-347, 1969.

Almeida, L. B. and F. M. Silva, "Variable-Frequency Synthesis: An Improved Harmonic Coding Scheme," *Proceedings of IEEE International Conference on Acoustics, Speech, Signal processing*, vol. 2, pp. 27.5.1-27.5.4, San Diego, CA, 1984.

Arslan, L. M. and D. Talkin, "Codebook Based Face Point Trajectory Synthesis Algorithm Using Speech Input," *Speech Communication*, vol. 27, no. 27, pp. 81-93, 1999.

References

Bakshi, B. and G. Stephanopoulos, "Wave-net: A Multiresolution, Hierarchical Neural Network With Localized Learning," *American Institute of Chemical Engineers, Journal*, vol. 39, pp. 57-81, January, 1993.

Barron, A., "Universal Approximation Bounds for Superpositions of A Sigmoidal Function," *IEEE Transactions on Information Theory*, vol. 39, no. 3, 1993.

Benoit, C., T. Guiard-Marigny, B. LeGoff, and A. Adjoudani, "Which Components of the Face Do Humans and Machines Best Speechread ?," in D. Stork and M. Hennecke, editors, *Speechreading by Humans and Machines*, New York: Springer Verlag, pp. 315-328, 1996.

Benveniste, A., A. Juditsky, B. Delyon, Q. Zhang, and P. Glorennec, "Wavelets in Identification," *Proceedings of 10th IFAC Symposium on System Identification*, vol. 2, pp. 27-48, Copenhagen, 1994.

Beskow, J., "Rule-Based Visual Speech Synthesis," *Proceedings of Eurospeech '95*, pp. 299-302, 1995.

Beskow K., K. Elenius, and S. McGlashan, "The OLGA Project: An Animated Talking Agent in A Dialogue System," *Proceedings of Eurospeech 97*, 1997.

Billings, S. A. and H. Wei, "A New Class of Wavelet Networks for Nonlinear System Identification," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 862-874, 2005.

Bishop, C. M., "Improving the Generalization Properties of Radial Basis Function Neural Networks," *Neural Computation*, vol. 3, no. 4, pp. 579-588, 1991.

References

- Boser, B. E., I. M. Guyon, and V. N. Vapnik. "A Training Algorithm for Optimal Margin Classifiers," in D. Haussler, editor, *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144-152, Pittsburgh, 1992.
- Brand, M., "Voice Puppetry," *Proceedings of ACM SIGGRAPH*, pp. 21-28, 1999.
- Bregler, C., M. Covell, and M. Slaney, "Video Rewrite: Driving Visual Speech with Audio," *Proceeding of ACM SIGGRAPH*, pp. 353-360, 1997.
- Borg, I. and Groenen, P., "Modern Multidimensional Scaling: Theory and Applications," Springer-Verlag, New York, 1997.
- Burrus, C. S., R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, Upper Saddle River, New Jersey, 1998.
- Cardle, M., L. Barthe, S. Brooks, and P. Robinson, "Music-Driven Motion Editing: Local Motion Transformations Guided by Music Analysis," *EGUK'2002 Eurographics UK Conference*, Leicester, UK, pp. 38-44, June 2002.
- Carrol, S. M. and B. W. Dickinson, "Construction of Neural Nets Using the Radon Transform," *Proceedings of the International Joint Conference on Neural Networks*, vol. I, pp. 607-611, Washington D. C., June 1989.
- Chang, C. M. and T. S. Liu, "A Wavelet Network Control Method for Disk Drives," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 1, pp. 63-68, 2006.
- Chen, F., D. Shi, and G. S. Ng, "Wavelet Network with OLS Optimization for Speech Signal Processing," *Canadian Conference on AI 2004*, LNAI 3060, pp. 406-415, London, Ontario, Canada, May 17-19, 2004.

References

Chen, S., W. Lai, and Y. Wang, "A New Duration Modeling Approach for Mandarin Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, 2003.

Chen, S., E. S. Chng, and K. Alkadhim, "Regularized Orthogonal Least Squares Algorithm for Constructing Radial Basis Function Networks," *International Journal of Control*, vol. 64, no. 5, pp. 829-837, 1996.

Chen, S., Y. Wu, and B. L. Luk, "Combined Genetic Algorithm Optimization and Regularised Orthogonal Least Squares Learning for Radial Basis Function Networks," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1239-1243, 1999.

Chen, S., C. F. Cowan, and P. M. Grant, "Orthogonal Least Squares Learning Algorithms for Radial Basis Function Networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 2, pp. 302-309, 1991.

Chui, C., *Wavelets: A Tutorial in Theory and Applications*, Academic Press, San Diego, California, 1992.

Cohen, M. M. and D. W. Massaro, "Modeling Coarticulation in Synthetic Visual Speech," in N. M. Thalmann and D. Thalmann, editors, *Models and Techniques in Computer Animation*, Springer Verlag, pp. 141-155, 1993.

Cosatto, E. and H. P. Graf, "Sample-Based Synthesis of Photo-Realistic Talking Heads," *Proceedings of IEEE Computer Animation*, pp. 103-110, 1998.

Cosatto, E. and H. P. Graf, "Photo-Realistic Talking-Heads from Image Samples," *IEEE Transactions on Multimedia*, vol. 2, no. 3, pp. 152-163, 2000.

References

Cosatto, E., H. P. Graf, J. Ostermann, and J. Schroeter, "Lifelike Talking Faces for Interactive Services," *Proceedings of the IEEE, special issue on multimedia human computer interfaces*, vol. 91, no. 9, pp. 1406-1429, 2003.

Cox, T. and Cox, M., "Multidimensional Scaling," Chapman and Hall, London, 1994.

Cybenko, G., "Approximation by Superposition of A Sigmoidal Function," *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303-314, 1989.

Damper, R. I., editor, *Data-Driven Techniques in Speech Synthesis*, Kluwer Academic Publishers, 2001.

Daponte, P., G. Mercurio, and S. Rapuano, "A Wavelet Networks-Based Method for the Digital Telecommunication System Monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 6, pp. 1773-1780, 2001.

de Sousa, C., E. M. Hemerly, and R. K. H. Galvão, "Adaptive Control for Mobile Robot Using Wavelet Networks," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 32, no. 4, pp. 493-504, 2002.

Ezzat T. and T. Poggio, "MikeTalk: A Talking Facial Display Based on Morphing Visemes," *Proceedings of IEEE Computer Animation*, pp. 96-102, 1998.

Ezzat, T., G. Geiger, and T. Poggio, "Trainable Videorealistic Speech Animation," *Proceedings of ACM SIGGRAPH*, pp. 388-397, 2002.

Fu, S., R. Gutierrez-Osuna, A. Esposito, P. K. Kakumanu, and O. N. Garcia, "Audio/Visual Mapping With Cross-Modal Hidden Markov Models," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp.243- 252, 2005.

References

Funahashi, K., "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*, vol. 2, pp. 183-192, 1989.

Galvão, R. K. H. and T. Yoneyama, "A Competitive Wavelet Network for Signal Clustering," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 34, no. 2, pp. 1282-1288, 2004.

Gasperlin, M., L. Tusar, M. Tusar, J. Kristl, and J. Smid-Korbar, "Lipophilic Semisolid Emulsion Systems: Viscoelastic Behaviour and Prediction of Physical Stability by Neural Network Modeling," *International Journal of Pharmaceutics*, vol. 168, no. 2, pp. 243–254, 1998.

Gomm, J. B., and D. L. Yu, "Selecting Radial Basis Function Network Centers with Recursive Orthogonal Least Squares Training", *IEEE Transactions on Neural Networks*, vol. 11, pp. 306-314, 2000.

Gratch, J., J. Rickel, E. Andre, N. Badler, J. Cassell, and E. Petajan, "Creating Interactive Virtual Humans: Some Assembly Required," *IEEE Intelligent Systems*, vol. 17, no. 4, pp. 54-63, 2002.

Gutierrez-Osuna, R., P. K. Kakumanu, A. Esposito, O. N. Garcia, A. Bojorquez, J. L. Castillo, and I. J. Rudomin, "Speech-Driven Facial Animation with Realistic Dynamics," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 33-42, 2005.

Hartman, E. J., J. D. Keeler, and J. M. Kowalski, "Layered Neural Networks with Gaussian Hidden Units as Universal Approximations," *Neural Computation*, vol. 2, no. 2, pp. 210-215, 1990.

References

- Ho, D., P. Zhang, and J. Xu, "Fuzzy Wavelet Networks for Function Learning," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 1, pp. 200-211, 2001.
- Hong, J., "Identification of Stable Systems by Wavelet Transform and Artificial Neural Networks," *Ph.D thesis*, University of Pittsburg, Ptissburgh, PA, 1992.
- Hong, P., Z. Wen, and T. S. Huang, "Real-Time Speech-Driven Face Animation with Expressions Using Neural Networks," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 916-927, 2002.
- Hornik, K., "Multilayer Feedforward Networks are Universal Approximators," *Neural networks*, vol. 2, 1989.
- Huang, Y. and C. Huang, "Evolving Wavelet Networks for Power Transformer Condition Monitoring," *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 412-416, 2002.
- Huang, Y., "A New Data Mining Approach to Dissolved Gas Analysis of Oil-Insulated Power Apparatus," *IEEE Transactions on Power Delivery*, vol. 18, no. 4, pp. 1257-1261, 2003.
- Hunt, A. J. and A. W. Black, "Unit Selection in a Concatenative Speech Synthesis System Using a Large Speech Database," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 373-376, Atlanta, 1996.
- Hussain, A. S., X. Yu, and R. D. Johnson, "Application of Neural Computing in Pharmaceutical Product Development," *Pharmaceutical Research*, vol. 8, no. 10, pp. 1248-1252, 1991.

References

Haykin, S., *Neural Networks, A Comprehensive Foundation*, Prentice Hall, New Jersey, USA, second edition, 1999.

Jiao, L., J. Pan, and Y. Fang, "Multiwavelet Neural Network and Its Approximation Properties," *IEEE Transactions on Neural Networks*, vol. 12, no. 5, pp. 1060-1066, 2001.

Kadambe, S. and G. F. Boudreaux-Bartels, "A Comparison of Wavelet Functions For Pitch Detection of Speech Signals", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal processing*, pp. 449-452, 1991.

Kortekaas, R. and A. Kohlrausch, "Psychoacoustical Evaluation of the Pitch-Synchronous Overlap-and-Add Speech-Waveform Manipulation Technique Using Single-Formant Stimuli," *Journal of the Acoustical Society of America*, JASA, vol. 101 (4), pp. 2202-2213, 1997.

Kshirsagar, S. and N. M. Thalmann, "Lip Synchronization Using Linear Predictive Analysis," *Proceedings of IEEE International Conference on Multimedia and Expo*, New York, August 2000.

Laroche, J., Y. Stylianou, and E. Moulines, "HNS: Speech Modification Based on a Harmonic + Noise Model," *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 550-553, 1993.

Lavagetto, F., "Converting Speech into Lip Movements: A Multimedia Telephone for Hard of Hearing People", *IEEE Transactions on Rehabilitation Engineering*, vol. 3, no. 1, pp. 90-102, 1995.

References

- Le Goff B. and C. Benoit, "A Text-to-Audiovisual-Speech Synthesizer for French," *Proceedings of International Conference on Spoken Language Processing '96*, pp. 2163-2166, Philadelphia, October 1996.
- Lee, Y., D. Terzopoulos, and K. Waters, "Realistic Modeling for Facial Animations," *Proceedings of ACM SIGGRAPH*, pp. 55-62, 1995.
- Lewis, J. P. and F. I. Parke, "Automated Lipsynch and Speech Synthesis for Character Animation," *Proceedings of Human Factors in Computing Systems and Graphics Interface*, pp.143-147, 1987.
- Li, Z., E. C. Tan, I. McLoughlin, and T. T. Teo, "Proposal of Standards for Intelligibility Tests of Chinese Speech," *IEE Proceedings Vision, Image & Signal Processing*, vol. 147, no. 3, June 2000.
- Ljung, L., *System Identification: Theory for the User*. Printice Hall Information and System sciences Series, Prentice Hall, 1987.
- Lu, J., H. S. Seah, and F. Tian, "Computer-Assisted Cel Animation: Post-processing After Inbetweening," *International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, February 2003.
- Lytle, W., "Driving Computer Graphics Animation From A Musicalsore," *Scientific Excellence in Supercomputing: The IBM 1990 Contest Prize Papers*, 1990.
- Mallat, S. G. and S. Zhong, "Characterization of Signals From Multiscale Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710-732, July 1992.

References

Massaro, D. W., *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, New Jersey: Lawrence Erlbaum, 1987.

McAulay, R. J. and T. F. Quatieri, "Magnitude-Only Reconstruction Using A Sinusoidal Speech Model," *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 27.6.1-27.6.4, 1984.

McAulay, R. J. and T. F. Quatieri, "Speech Analysis/Synthesis Based on A Sinusoidal Representation," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. ASSP-34, pp. 744-754, August 1986.

McGurk, H. and J. MacDonald, "Hearing Lips and Seeing Voices," *Nature*, pp. 746-748, December 1976.

Meulman, J., "A Distance Approach to Nonlinear Multivariate Analysis," DSWO Press, Leiden, 1986.

Mishra, S. and J. Hahn, "Mapping Motion to Sound and Music in Computer Animation and VE," *Invited Paper, Proceedings of Pacific Graphics*, 1995.

Moody, J. and C. J. Darken, "Fast Learning in Networks of Locally Tuned Processing Units," *Neural Computation*, vol. 1, no. 2, pp. 281-294, 1989.

Morishima, S. and H. Harashima, "A Media Conversion from Speech to Facial Image for Intelligent Man-Machine Interface," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 4, pp. 594- 600, 1991.

References

Moulines, E. and F. Charpentier, "Pitch Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones," *Speech Communication*, vol. 9, pp. 453-467, December 1990.

Moulines, E. and J. Laroche, "Non-Parametric Techniques for Pitch-Scale Modification of Speech," *Speech Communication*, vol. 16, pp. 175-205, 1995.

Muller, K., A. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. Vapnik, "Using Support Vector Machines for Time Series Prediction", in B. Scholkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*, pp. 243-253, Cambridge, MA: MIT Press, 1999.

Murtoniemi, E., P. Merkkü, P. Kinnunen, K. Leiviskä, and J. Yliruusi, "Effect of Neural Network Topology and Training End Point in Modelling the Fluidized Bed Granulation Process," *International Journal of Pharmaceutics*, vol. 110, pp. 101-108, 1994.

Nahas, M., H. Huitric, and M. Saintourens, "Animation of A B-Spline Figure," *The Visual Computer*, vol. 3, no. 5, pp. 272-276, 1988.

Nakamura, S., "HMM-based Transmodal Mapping from Audio Speech to Talking Faces," *Proceedings of IEEE International Workshop on Neural Networks for Signal Processing*, vol. 10, pp. 33-42, 2000.

Nielson, G. M., "Scattered Data Modeling," *IEEE Computer Graphics and Applications*, vol. 13, no. 1, pp. 60-70, 1993.

References

- Ohman, T., "An Audio-Visual Speech Database and Automatic Measurements of Visual Speech," *Quarterly Progress and Status Report*, Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm, Sweden, 1998.
- Osuna, E. and F. Girosi, "Reducing the Run-Time Complexity of Support Vector Machines," *Proceedings of the 14th International Conference on Pattern Recognition*, Brisbane, Australia, 1998.
- Parke, F. I., "A Parameterized Model for Facial Animation," *IEEE Computer Graphics and Applications*, vol. 2, no. 9, pp. 61-70, 1982.
- Pati, Y. and P. Krishnaprasad, "Analysis and Synthesis of Feedforward Neural Networks Using Discrete Affine Wavelet Transformations," *IEEE Transactions on Neural Networks*, vol. 4, pp. 77-85, January 1993.
- Pelachaud, C., N. I. Badler, and M. Steedman, "Generating Facial Expressions for Speech," *Cognitive Science*, vol. 20, no. 1, pp. 1-46, 1996.
- Pelachaud, C., "Visual Text-to-Speech," in I. S. Pandzic and R. Forchheimer, editors, *MPEG-4 Facial Animation: The Standard, Implementation and Applications*, pp. 125-140, Chichester, UK: John Wiley and Sons, 2002.
- Poggio, T. and F. Girosi, "Networks for Approximation and Learning," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1481-1497, 1990.
- Rabiner, L. R. and R. W. Schafer, *Digital Processing of Speech Signal*, Englewood Cliffs, New Jersey: Prentice Hall, 1978.

References

Resnikoff, H. L. and R. O. Wells, *Wavelet Analysis: The Scalable Structure of Information*, Springer Verlag, October 1998.

Robel, A., "Scaling Properties of Neural Networks for the Prediction of Time Series," *Proceedings of the 1996 IEEE Workshop on Neural Networks for Signal Processing VI*, 1996.

Ruskai, M., G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, editors, *Wavelets and Their Applications*. Boston: Jones and Bartlett Books in Mathematics, 1992.

Santen J., R. Sproat, J. Olive, J. Hirschberg, editors, *Progress in Speech Synthesis*, New York: Springer-Verlag, 1997.

Scholkopf, B., C. Burges, and V. Vapnik, "Incorporating invariances in support vector learning machines," in C. von der Malsburg, W. von Seelen and B. Sendhoff, editors, *Proceedings of the International Conference on Artificial Neural Networks*, vol. 1112, pp. 47-52, 1996.

Scholkopf, B., K. K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2758-2765, November 1997.

Scholkopf, B., P. Simard, A. Smola, and V. Vapnik, "Prior Knowledge in Support Vector Kernels," in M. K. M. Jordan and S. Solla, editors, *Advances in Neural Information Processing System* vol. 10, pp. 640-646, Cambridge, MA: MIT Press, 1998.

References

Shih, C. and G. P. Kochanski, "Chinese Tone Modeling with Stem-ML," *International Conference on Spoken Language Processing*, Beijing, China, 2000.

Sjoberg, J., H. Hjalmarsson, and L. Ljung, "Neural Networks in System Identification," *Proceedings of 10th IFAC Symposium on System Identification*, vol. 2, pp. 49-72, Copenhagen, 1994.

Sontag, E. D., "Feedforward Nets for Interpolation and Classification," *Journal of Computer and System Sciences* 45, pp. 20-48, 1992.

Stinchcombe, M. and H. White, "Universal Approximation Using Feedforward Networks with Non-sigmoid Hidden Layer Activation Functions," *Proceedings of the International Joint Conference on Neural Networks*, vol. I, pp. 613-617, Washington D.C., June 1989.

Stylianou, Y., "A Pitch and Maximum Voiced Frequency Estimation Technique Adapted to Harmonic Models of Speech," *IEEE Nordic Signal Processing Symposium*, 1996.

Stylianou, Y., "Applying the Harmonic Plus Noise Model in Concatenative Synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 21-29, 2001.

Syrdal, A., Y. Stylianou, L. Garrison, A. Conkie, and J. Schroeter, "TD-PSOLA versus Harmonic Plus Noise Model in Diphone Based Speech Synthesis," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. I, pp. 273-276, May 1998.

Takayama, K., M. Fujikawa, and T. Nagai, "Artificial Neural Network as A Novel Method to Optimize Pharmaceutical Formulations," *Pharmaceutical Research*, vol. 16, no. 1, pp. 1-6, 1999.

References

- Tung, W. L. and C. Quek, "GenSoFNN: A Generic Self-Organizing Fuzzy Neural Network," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1075–1086, 2002.
- Vapnik, V., *Estimations of Dependences Based on Empirical Data*, New York: Springer-Verlag, 1982.
- Vapnik, V., *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.
- Waters, K., "A Muscle Model for Animating Three-Dimensional Facial Animation," *Computer Graphics*, vol. 21, no. 4, pp. 17-24, 1987.
- Waters, K. and T. M. Levergood, "DECface: An Automatic Lip-Synchronization Algorithm for Synthetic Faces," *Technical Report CRL 93/4*, Cambridge Research Lab, 1993.
- Weigend, A. S. and N. A. Gershenfeld. "Time Series Prediction: Forecasting the Future and Understanding the Past," *Addison-Wesley Pub. Comp*, 1993.
- Yamamoto, E., S. Nakamura, and K. Shikano, "Lip Movement Synthesis from Speech Based on Hidden Markov Models," *Speech Communication*, vol. 26, no. 1-2, pp. 105-115, 1998.
- Yehia, H., P. E. Rubin, and E. Vatikiotis-Bateson, "Quantitative Association of Vocal-Tract and Facial Behavior," *Speech Communication*, vol. 26, pp. 23-43, 1998.
- Young, R. K., *Wavelet Theory and Its Applications*, Boston: Kluwer Academic Publishers, 1992.

References

Zhang, J., G. G. Walter, Y. Miao, and N. W. Lee, "Wavelet Neural Networks for Function Learning," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1485-1497, 1995.

Zhang, Q. and A. Benveniste, "Wavelet Network," *IEEE Transactions on Neural Networks*, vol. 3, no. 6, pp. 889-898, Nov 1992.

Zhang, Q., "Using Wavelet Network in Nonparametric Estimation," *IEEE Transactions on Neural Networks*, vol. 8, no. 2, pp. 227-236, March 1997.