DESIGNING A NEURAL NETWORK BASED AUDIO CLASSIFICATION SYSTEM

KHALED ABDALGADER MOHAMED OMAR

UNIVERSITI UTARA MALAYSIA
2004

DESIGNING A NEURAL NETWORK BASED AUDIO CLASSIFICATION SYSTEM

A thesis submitted to the Graduate School in partial fulfillment of the requirement for the degree Master of Science (Intelligent System),

Universiti Utara Malaysia.

(Northern University of Malaysia)

By

KHALED ABDALGADER MOHAMED OMAR

© KHALED ABDALGADER, June 2004. All rights reserved.

JABATAN HAL EHWAL AKADEMIK
(*Department of Academic Affairs*)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

KHALED ABDALGADER MOHAMED OMAR

calon untuk Ijazah
(candidate for the degree of) **MSc. (Int. Sys.)**

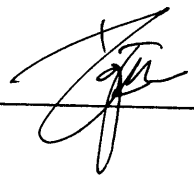
telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

DESIGNING A NEURAL NETWORK BASED AUDIO CLASSIFICATION SYSTEM

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the filed is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **MRS. NORAZIAH CHE PA**

Tandatangan
(Signature) : 

Tarikh
(Date) : June 16, 2004

PERMISSION TO USE

In presenting this thesis as major requirement for a pos-graduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may take it freely available for inspection after being submitted for a year. I further agree that permission for copying of this thesis in any manner in whole or in part, for scholarly purposes may be granted by my supervisor or, in this absence, by the Dean of the Graduate School. It is understood that any copying or publication or use of this thesis or part thereof for financial gain by given to me and to Universiti Utara Malaysia for any scholarly use, which may be made of any material from my thesis.

Request for permission to copy or to make other use of material in this thesis, in whole or in part, should be addressed to:

Dean of Graduate School

Universiti Utara Malaysia

06010 Sintok

Kedah Darul Aman

MALAYSIA

ABSTRACT

Artificial neural networks have found profound success in the area of pattern recognition. The collections of digital music have become increasingly common over the recent years. As the amount of data increases, digital content classification is becoming more important. In this thesis, we are studying content-based classification of digital musical signals according to their musical genre (e.g., jazz, rock, pop and blues) and the features used. The purpose of this thesis is to propose of designing a neural network based audio classification system "*Model*", and analyze the requirements that needed to classifying it. This thesis covers a literature review on human musical genre recognition, neural network technique, signal processing, and related works of research. In addition, the methodology that used in designing audio classification model using neural network is introduced. The method was follow in this thesis is content analysis, and the designing of the model has through several phases: requirements analysis, knowledge representation and model designing. The theory behind the used features is reviewed and the fining from the proposed designing is presented.

Keywords: Neural Network; Digital Audio; Classification System.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and credit to the following:

- Universiti Utara Malaysia for the well-equipped resources, facilities and excellent professionalism especially by the Graduate School's staff, and the Faculty of Information Technology's staff.
- My idealistic and tolerant supervisor, *Puan Noraziah Che Pa* for her motivation and structured guidance in helping me to conduct the evaluation and writing this thesis.
- I wish to thank my parents for understanding the long periods I have been working and not having time to visit them. Also I wish to thank them for all support and endless love. Big thanks also to my friends for understanding my little time for them during this work.
- All lecturers and colleagues who helped to strengthen this work.

KHALED, June 2004.

TABLE OF CONTENTS

Permission to Use..... I

Abstract..... II

Acknowledgement..... III

Table of Contents..... IV

List of Figures..... XV

Chapter 1: Introduction

1.1 Overview..... 1

1.2 Problem Statement..... 2

1.3 Objective of the Study..... 3

1.4 Scope of the Study..... 3

1.5 Significant of the Study..... 4

1.6 Organization of the Thesis..... 4

Chapter 2: Literature Review

2.1 Overview..... 5

2.2 Signal Processing and Synthesis..... 5

2.3 Digital Audio Files Formats..... 6

2.3.1 Wave Format..... 6

2.4 Music Genre..... 7

2.4.1 Music Genre Recognition..... 8

2.4.2 Music Genre Taxonomy..... 9

2.4.3 Music Genre Hierarchy..... 10

2.5 Difficulties..... 11

2.6 Artificial Neural Networks..... 12
2.6.1 Biological Neuron..... 12
2.6.2 Artificial Neuron..... 13
2.6.3 Neural Networks Learning..... 14
2.6.4 Backpropagation Algorithm..... 16
2.6.5 Using Neural Networks..... 16
2.6.6 Multilayer Perceptron..... 18
2.6.7 Neural Networks in Audio Recognition..... 20
2.7 Related Works..... 21
2.7.1 Speech Recognition..... 21
2.7.2 Speaker Recognition..... 22
2.7.3 Music vs. Speech Classification..... 22
2.7.4 Video Content Analysis..... 22
2.7.5 Speech/Music Discrimination and General Audio Classification..... 23

Chapter 3: Methodology

3.1 Overview..... 25
3.2 Requirements Analysis..... 26
3.2.1 Features Description..... 26
3.2.2. Features Approach..... 27
3.2.2.1 Choosing the Right Feature..... 27
3.2.3 Music Characteristic and Description..... 29
3.3 Knowledge Representation..... 30
3.3.1 Digital Audio..... 30
3.3.1.1 Digital Audio File..... 30
3.3.2 Digital WAVE Format..... 31
3.4 Model Designing..... 33
3.4.1 Determining Model Architecture..... 34

3.4.2 Determining Features..... 35
3.4.2.1 Frequency Domain..... 35
3.4.2.2 Time Domain..... 35
3.4.3 Determining Neural Network Architecture..... 36
3.4.3.1 Determining Learning Method..... 37
3.4.3.2 Determining Activation Function..... 38

Chapter 4: Findings

4.1 Overview..... 39
4.2 Features Needed to Classify Music..... 39
4.2.1 Frequency Domain (MFCC, BER)..... 39
4.2.1.1 Mel-Frequency Cepstral Coefficients..... 39
4.2.1.2 Band Energy Ratio (BER)..... 42
4.2.2 Time Domain (Amplitude, Duration)..... 43
4.3 Classification Model..... 43
4.3.1 Pre-processing (Normalization)..... 46
4.3.2 Feature Extraction..... 47
4.3.3 Pre-processor (Transformation)..... 48
4.3.4 Classification..... 48
4.4 Discussion..... 50

Chapter 5: Conclusion

5.1 Overview..... 51
5.2 Conclusions..... 51
5.3 Suggestions and Limitations..... 52
5.3.1 More Advance Feature Extraction..... 52
5.3.2 MP3 Files Instead of WAV Files..... 53
5.3.3 More Classification Categories..... 53

5.3.4 Model Evaluation..... 54

References..... 55

Appendix-A: A Musical genre hierarchy (audio taxonomy)..... 60

Appendix-B: Backpropagation Algorithm..... 61

LIST OF FIGURES

Figure 2.1: Wave Format Representation..... 6

Figure 2.2: Simplified Biological Neuron..... 13

Figure 2.3: Basics Artificial Neuron..... 14

Figure 2.4: Multilayer Perceptron Structure..... 19

Figure 3.1: Research Design and Objective Achieved..... 25

Figure 3.2: Structure of Digital Audio File..... 31

Figure 3.3: Waveform Diagram..... 31

Figure 3.4: Technical Terms..... 32

Figure 3.5: Structure of the Digital WAVE File..... 33

Figure 3.6: Model Designing Phases..... 34

Figure 3.7: Neural Network Architecture Used for Classification..... 36

Figure 4.1: Extracting the MFCC Features..... 41

Figure 4.2: Proposed Model of Neural Network Based Music Classification System..... 44

Figure 4.3: Use Case of Model Processing..... 45

Figure 4.4: Normalization Process..... 45

Figure 4.5: Feature Extraction Process..... 45

Figure 4.6: Classification Process..... 46

CHAPTER 1

INTRODUCTION

1.1 Overview

Classification of digital audio "*signals*" according to their contents has been a major concern in recent years. There have been many studies on *Audio Content Analysis* (ACA), using different features and different methods. It is a well-known fact that audio signals are baseband, one-dimensional signals (Aksoy, 2002). General audio consists of a wide range of sound phenomena such as music, sound effects, environmental sounds, and speech and non-speech signals.

With the development of multimedia technology, classification is increasingly used in audio applications (Wold, 1996). In general, many research efforts high accuracy audio classification is only achieved for the simple cases such as speech, music discrimination. Pfeiffer (1996) have presented, a theoretical framework and application of automatic audio content analysis using some perceptual features. Saunders (1996), presented a speech music classifier based on simple features such as *Zero Crossing Rate* (ZCR) and short time energy for radio broadcast.

A *Neural Network* (NN) classifier is an artificial intelligent network with parallel processing units working together. The most common neural network model is the *Multilayer Perceptron* (MLP). This type of neural network is known as a supervised network (Haykin, 1999), because it requires a desired output in order to learning. The goal of this type of network is to create a model that correctly maps the input to the

The contents of the thesis is for internal user only

References

Aksoy, M., Ayaz, H. & Konukoglu, E. (2002). Content-based Classification of Musical Instrument Sounds Using Gaussian Methods. EE 473 DIGITAL SIGNAL PROCESSING.

Aucouturier, J. & Pachet, F. (2003). Representing Musical Genre: A State of the Art. Journal of New Music Research, vol. 32, no. 1, pages 83-93.

Batlle, E. & Cano, P. (2001). Automatic Segmentation for Music Classification using Competitive Hidden Markov Models. Audiovisual Institute. University Pompeu Fabra Rambla 31, 08002 Barcelona Catalunya – Spain.

Becchetti, C. & Ricotti, P. L. (1999). Speech Recognition, Theory and C++ Implementation. John Wiley & Sons.

Breure, L. (2001). Development of the genre concept. New Music Journal. Retrieved April 4th, 2004, form: http://www.cs.uu.nl/people/leen/GenreDev/GenreDevelopment.htm.

Burred, J. & Lerch, A. (2003). A Hierarchical Approach to Automatic Musical Genre Classification. Communication Systems Group Technical University Berlin, Germany, Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK, September 8-11.

Daniel, K. (1998). Artificial Neural Networks. An individual project within MISB-420-0. Saint Louis University. St. Louis November. Retrieved March 22nd, 2004, from: http://hem.hj.se/~de96klda/NeuralNetworks.htm

Erickson, T. (1999). Rhyme and punishment: The creation and enforcement of conventions in an on-line participatory limerick genre.

Eronen, A. (2003). Musical instrument recognition using ICA-based transform of features and discriminatively trained HMMs. vol. 2, pages 133-136. Seventh International Symposium on Signal Processing and Its Applications.

Foote, J. *a* (1999). An Overview of Audio Information Retrieval. Multimedia Systems, 7(1): 2–10.

Foote, J. *b* (1997). A Similarity Measure for Automatic Audio Classification. In Proc. AAAI Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora.

Grimaldi, M., Cunningham, P. & Kokaram, A. (2001). An Evaluation of Alternative Feature Selection Strategies and Ensemble Techniques for Classifying Music. Computer Science Department, Trinity College Dublin, Ireland and Electronic Engineering Department, Trinity College Dublin, Ireland.

Hartmann, W., Pitch, P. & Auditory O. (1996). Journal of the Acoustical Society of America, vol. 100, no. 6, pages 3491-3502.

Haykin, S. (1999). Neural networks: a comprehensive foundation. 2nd edition, Prentice-Hall.

Houtsma, A. J. M. (1997). Pitch & Timbre: Definition, Meaning and Use. Journal of New Music Research, vol. 26, no. 2, pages 104-115, 1997.

Houtsma, A. J. M. (1995). Hearing, handbook of perception and cognition, chapter 8: Pitch Perception, pages 267-296. Academic Press Inc., San Diego, CA, USA, 2nd edition.

Hussey, J. & Hussey, R. (1997). Business Research: A Practical Guide for Undergraduate and Postgraduate Students, 1st ed. MACMILLAN PRESS LTD.

Li, D., Sethi, I., Dimitrova, N. & McGee, T. (2001). Classification of general audio data for content-based retrieval. Pattern Recognition Letters, vol. 22, no. 5, pages 533-544.

Liu, Z., Wang, Y. & Chen, T. (1998). Audio Feature Extraction and Analysis for Scene Segmentation and Classification. Journal of VLSI Signal Processing System, 20: 61–79.

Logan, B. (2000). Mel Frequency Cepstral Coefficients for music modeling. International Symposium on Music Information Retrieval.

Oravec, M. (2001). Multilayer Perceptron in Face Recognition. Faculty of Electrical Engineering and Information Technology, Bratislava.

Pachet, F. & Cazaly, D. (2000). Taxonomy of Musical Genres. Paris, France. Content-Based Multimedia Information Access Conference (RIAO).

Perelmuter, G., Enrique, V., Vellasco, M. & Pacheco, M. (2000). Recognition of Industrial Parts Using Artificial Neural Networks. Department of engineering, PUC Rio, Brazil.

Pfeiffer, S., Fischer, S. & Effelsberg, W. (1996). Automatic Audio Content Analysis. Proceedings of the fourth ACM international conference on Multimedia, pp. 21-30.

Rabiner, L. and Juang, B. H. (1993). Fundamentals of Speech Recognition. PTR Prentice-Hall Inc., New Jersey.

Reynolds, D. & Rose, R. (1995). Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pages 72-83.

Robert, S. T. (2001). Digital Audio Technology. Center for Audio Recording Arts (CARA). MI 313. Retrieved March 25th, 2004, from: http://cara.gsu.edu/courses/MI_313/digi2.htm

Sadie, S. editor. (2001). The New Grove Dictionary of Music and Musicians. MacMillan Publishing Company.

Saunders, J. (1996). Real-time Discrimination of Broadcast Speech/ Music. Proc. ICASSP96, vol.II, pp.993-996, Atlanta.

Saunders, J. (1996). Real-time Discrimination of Broadcast Speech/Music, vol. 2, pages 993-996, Atlanta, GA. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

Scheirer, E. (1998). Tempo and beat analysis of acoustic musical signals. Journal of the Acoustical Society of America, vol. 103, no. 1, pages 558-601.

Scheirer, E. & Slaney, M. (1997). Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator, vol. 2, pages 1331-1334. IEEE International Acoustics, Speech, and Signal Processing (ICASSP).

Silvia, P., Stephan, F. & Wolfgang, E. (1996). Automatic Audio Content Analysis. In ACM Multimedia 96, pages 21–30, Boston.

Scott, P. (2001). Music Classification Using Neural Network. EE373B Project.

Wold, E., Blum, T. & Wheaton, J. (1996). Content-based Classification, Search and Retrieval of Audio. IEEE Multimedia, 3(3), pp.27-36.

Zhang, T. & Kuo, C. (1998). Content-based Classification and Retrieval of Audio. In SPIE's 43rd Annual Meeting - Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations VIII, San Diego.

Zhang, T. & Kuo, J. (1999). Content-based Audio Classification and Retrieval. University of Southern California. Retrieved March 25th, 2004, from: http://viola.usc.edu/extranet/Projects/database-audio/

Zhang, T. & Kuo, C. (2001). Audio Content Analysis for Online Audio Visual Data Segmentation and Classification. IEEE Transactions on Speech and Audio Processing, vol. 9, no. 4, pages 441-457.