Automatic Segmentation, Learning and Retrieval of Melodies Using A Self-Organizing Neural Network

Steven Harford
School of Computing
Dublin City University
Dublin 9
Ireland
sharford@computing.dcu.ie

Abstract

We introduce a neural network, known as SONNET-MAP, capable of automatic segmentation, learning and retrieval of melodies. SONNET-MAP is a synthesis of the SONNET (Self-Organizing Neural NETwork) architecture (Nigrin, 1993) and an associative map derived from ARTMAP (Carpenter, Grossberg, and Reynolds, 1991). SONNET-MAP automatically segments a melody based on pitch and rhythmic grouping cues. Separate SONNET modules represent the pitch and rhythm dimensions of each segmented phrase independently, with two associative maps fusing these representations at the phrase level. Further SONNET modules aggregate these phrases forming a hierarchical memory structure that encompasses the entire melody. In addition, melodic queries may be used to retrieve any encoded melody. As far as we are aware, SONNET-MAP is the first self-organizing neural network architecture capable of automatically segmenting and retrieving melodies based on both pitch and rhythm.

1 Introduction

Typically, music information retrieval (MIR) systems represent melodies as strings of symbols drawn from a finite alphabet, thereby reducing the retrieval process to the task of approximate string matching. We introduce a self-organizing neural network architecture, known as SONNET-MAP, that offers an alternative approach to MIR. SONNET-MAP does not represent a melody as a string of symbols but, instead, as a learned hierarchy of automatically segmented melodic phrases that are encoded in long-term memory (LTM) by adjustable connection weights. Furthermore, the retrieval process is accomplished by propagating cell activities through the network rather than by the use of traditional string matching algorithms.

SONNET-MAP is primarily based on the SONNET architecture. Recent work has applied SONNET to the problem of learning pitch sequences (Page, 1993) and to the interpretation of rhythmic structure (Roberts, 1996). We expand on this research by applying SONNET-MAP to the automatic segmentation and retrieval of melodies based on both pitch and rhythm.

2 SONNET-MAP

2.1 Architecture

SONNET-MAP is composed of two parallel SONNET modules; one which represents pitch sequences and the other which represents rhythms (see Figure 1). These modules are connected via two associative maps, derived from ARTMAP, that bind pitch sequence representations with their corresponding rhythm representations, thereby forming rigid, two-dimensional representations of melodic phrases. An additional SONNET module aggregates these phrases forming melodic phrase sequences which, in turn, are aggregated by further SONNET modules forming a hierarchical memory structure that encompasses the entire melody. With such an organization, a single classification cell at the most abstract level in the hierarchy represents the whole melody.

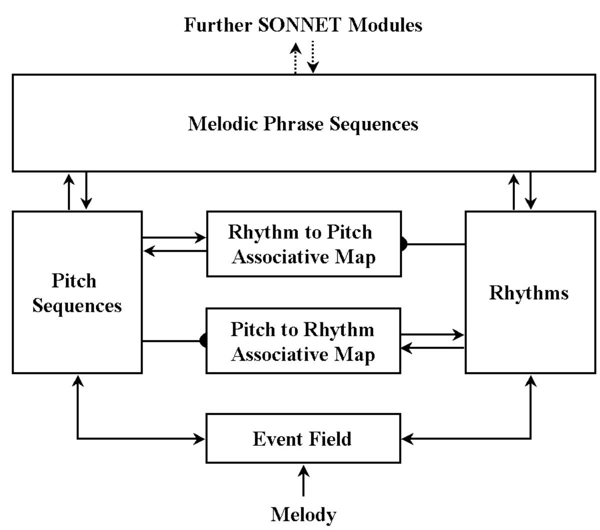

Figure 1: The SONNET-MAP Architecture

2.2 Melody Representation

Melodies are presented to SONNET-MAP's event field as a sequence of {Invariant Pitch-Class, Onset Time} duples. With the invariant pitch-class representation (Bharucha, 1991), all sequences are normalized into a common set of invariant pitch categories by coding them relative to a tonal center. In the case of Western tonal music employing equal-tempered tuning, every input cell of the pitch sequence module represents one of twelve chromatic pitch classes. Rhythms are represented using inter-onset intervals (IOIs). By allowing the input cells of the rhythmic module to represent onset times and by allowing the cell activations to continuously change, the relative activity levels associated with consecutive onsets represent IOIs (Roberts, 1996).

2.3 Segmentation

SONNET-MAP segments melodies using a number of grouping mechanisms that are based on aspects of human melody perception (Roberts, 1996). Some of these mechanisms are described below:

- An extended response time is afforded to the first items in a sequence. Consequently, a primacy effect is exhibited whereby the first pattern in a sequence is generally more likely to be encoded than later patterns.
- The limited capacity and depth of STM restricts the size and duration of melodic segments.
- Bottom-up weight initialization biases the network towards grouping by temporal proximity.
- Generally, relatively long IOIs tend to indicate phrase boundaries. Such IOIs provide an extended response time allowing SONNET-MAP to segment patterns ending with relatively long IOIs without difficulty.
- Frequently occurring musical patterns tend to represent significant melodic phrases. Repetitions of such phrases allow them to be easily segmented by SONNET-MAP.

Melucci and Orio (2002) argue that automatic segmentation based on melodic features is more effective than algorithms that do not use such information. The approach outlined above is consistent with this argument and preliminary SONNET-MAP simulations produced segmentations comparable to manual segmentation.

2.4 Retrieval

The parallel and hierarchical organization of SONNET-MAP provides practical benefits when applied to content-based retrieval (CBR). The initial separation of pitch sequences and rhythms allows retrieval to be based on either pitch, or rhythm, or both. SONNET-MAP also provides a set of vigilance parameters that may be used to allow either exact or approximate melody matching. Unlike other CBR systems, SONNET-MAP does not retrieve melodies using traditional string matching algorithms. Instead, by presenting melodic queries to SONNET-MAP, previously encoded melodies are retrieved by the propagation of cell activities throughout the network. The most highly activated cell assemblies in the most abstract SONNET module are used to form a ranked list of candidate melodies.

3 Preliminary Simulations

Preliminary simulations of SONNET-MAP were tested using a set of contemporary melodies. The segmentations formed were consistent with the grouping mechanisms described above and were comparable to manual segmentations, which we performed. SONNET-MAP was then applied to the task of CBR. Melodic queries were presented to SONNET-MAP and in each case SONNET-MAP retrieved the desired melody. Generally, melodies were retrieved using fewer notes when the melodic queries were aligned with the segmentations performed by SONNET-MAP. This further demonstrates the value of utilizing segmentation based on melodic features.

4 Conclusions and Future Work

We introduced a system, known as SONNET-MAP, capable of automatic segmentation, learning and retrieval of melodies. Preliminary testing revealed promising results regarding SONNET-MAP's ability to segment and retrieve melodies. However, these simulations were based on exact pattern matching using perfect queries. SONNET-MAP needs to be tested using more realistic queries, such as those generated by query-by-humming (QBH) systems, in order to evaluate how well it copes with imperfect input.

Roberts (1996) shows how a beat-tracking signal can be transmitted to SONNET in order to achieve tempo invariant pattern recognition. No beat-tracking signal is provided in the network simulations described above; instead the tempo of each melody is supplied in advance (in the form of a tactus-span) in order to achieve tempo invariance. This method has the disadvantage of preventing SONNET-MAP from operating in real-time. Therefore, future versions of SONNET-MAP should be provided with a beat-tracking signal.

SONNET-MAP was designed in a manner that allows the different dimensions of melody to be independently processed and subsequently combined at more abstract levels. The addition of other parallel SONNET modules representing other melodic dimensions (e.g. duration, intensity, timbre) would further enhance SONNET-MAP's ability to segment and retrieve melodies.

References

Bharucha, J. J. (1991). Pitch, harmony, and neural nets: A psychological perspective. In Todd, P. M. & Loy, D. G. (Eds.), *Music and Connectionism*. The MIT Press.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991). Artmap: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, 4(5).

Melucci, M. & Orio, N. (2002). A comparison of manual and automatic melody segmentation. In *Proceedings of the 3rd International Conference on Music Information Retrieval*.

Nigrin, A. (1993). *Neural Networks for Pattern Recognition*. The MIT Press.

Page, M. P. A. (1993). *Modelling Aspects of Music Perception Using Self-Organizing Neural Networks*. PhD thesis, University of Wales.

Roberts, S. C. (1996). *Interpreting Rhythmic Structures Using Artificial Neural Networks*. PhD thesis, University of Wales.