COMPUTATIONAL MODELING OF IMPROVISATION IN TURKISH FOLK MUSIC USING VARIABLE-LENGTH MARKOV MODELS

A Thesis
Presented to
The Academic Faculty

by

Sertan Şentürk

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Music Technology in the
School of Music

Georgia Institute of Technology
December 2011

COMPUTATIONAL MODELING OF IMPROVISATION IN TURKISH FOLK MUSIC USING VARIABLE-LENGTH MARKOV MODELS

Approved by:

Dr. Parag Chordia, Committee Chair
School of Music
Georgia Institute of Technology

Dr. Gil Weinberg
School of Music
Georgia Institute of Technology

Dr. Jason Freeman
School of Music
Georgia Institute of Technology

Date Approved: 25 August 2010

Benim her derdime ortak sen oldun

Ağlarsam ağladın gülersem güldün

Sazım bu sesleri turnadan m'aldın

Pençe vurup sarı teli sızlatma

Aşık Veysel Şatıroğlu

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Parag Chordia. Without his incredible guidance and "know-how" skills, I am doubtful that I would be able to learn and achieve this much throughout the masters program. I would also like to thank Jason Freeman for the very same reason. I will remember him as one of the most knowledgeable, dedicated and helpful teachers I have ever had. I am very thankful to Prof. Gil Weinberg for his supportive guidance and comments in the review of the thesis.

I would like to thank Turkish Radio and Television Corporation (TRT), *Türk Müzik Kültürünün Hafızası Score Archive* and the numerous musicians, transcribers and archivers for their efforts in building the *TRT Turkish folk music database*. I would also like to thank Prof. Erdal Tuğcular for his invaluable contributions to the *Uzun Hava Humdrum Database*, Avinash Sastry for the multiple viewpoints framework, and Prof. Okan Murat Öztürk for his assistance in Turkish music theory and ethnomusicology. Their efforts simply form the foundations of my thesis.

Additionally, I would like to thank Prof. Nezihe Şentürk, Prof. Xavier Serra, Prof. Barış Bozkurt and Dr. Andre Holzapfel for their precious comments about the thesis and suggestions for future work. I am indebted to Nathan Weitzner, Brian Blosser and Andrew Colella for proofreading my English. Last but not the least, I want to thank all my family (especially my mother, father and my sister) and my friends (especially Emre Can Dağlıoğlu, Can Öktemer and Avinash Sastry). Without their selfless support, I would have never even made it to a new day. On the contrary, without my ridiculous ways, they would have a much better life.

This material is based upon work supported by the National Science Foundation under Grant No. IIS-0855758.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The thesis describes a new database of *uzun hava*s, a non-metered structured improvisation form in Turkish folk music, and a system, which uses Variable-Length Markov Models (VLMMs) to predict the melody in the *uzun hava* form. The database consists of 77 songs, encompassing 10849 notes, and it is used to train multiple viewpoints, where each event in a musical sequence are represented by parallel descriptors such as *Duration*s and *Note*s. The thesis also introduces *pitch-related* viewpoints that are specifically aimed to model the unique melodic properties of *makam* music. The predictability of the system is quantitatively evaluated by an entropy based scheme. In the experiments, the results from the *pitch-related* viewpoints mapping 12-tone-scale of Western classical theory and 17-tone-scale of Turkish folk music are compared. It is shown that VLMMs are highly predictive in the note progressions of the transcriptions of *uzun hava*s. This suggests that VLMMs may be applied to *makam*-based and non-metered musical forms, in addition to Western musical styles. To the best of knowledge, the work presents the first symbolic, machine-readable database and the first application of computational modeling in Turkish folk music.

CHAPTER I

INTRODUCTION

1.1 *Overview and Remarks*

The thesis presents a new symbolic database of *uzun hava*s, a non-metered structured-improvisation form in Turkish folk music, and a machine learning system used for computational modeling of *uzun hava*s.

This introduction gives a brief presentation of the thesis and addresses the motivation. Chapter 2 gives a brief explanation of traditional Turkish music, related works in music information retrieval (MIR), and the contributions and novelty of the thesis. Chapter 3, presents the symbolic database and the conceptual and practical difficulties faced during it's creation. Chapter 4 brings the hypothesis, the computational modeling framework and the evaluation process. Next, Chapter 5 explains the experimental setup, evaluation method and the results obtained from the evaluation of the computational modeling. This chapter also presents the novel representation proposed to model Turkish folk music. Chapter 6 discusses the results, and approaches taken throughout the research. Chapter 7 suggests future works to be completed. Finally, Chapter 8 concludes the thesis work.

Throughout the thesis, even though English translations are typically provided, Turkish terms are more emphasized. This is due to the fact that traditional Turkish music is mostly an oral tradition, which cannot be explained by Western classical music theory. For this reason, English interpretations hold the danger of being "lost in translation" and sometimes being completely misunderstood.

As a final remark, there are some mentions to *Türk Sanat Müziği (Musikisi)*, which is a sub-genre under traditional Turkish music and has emerged from the Ottoman palace [69]. *Turkish classical music*, *Ottoman classical music* and *Turkish art music* (and all other possible variants) are possible translations. In order to emphasize the origin of the music and as a homage to the Arabic, Armenian, Greek, Kurdish, Jewish, Persian, Polish and all other musicians, who played the music in the Ottoman court, I have chosen to use the term, "Ottoman classical music."

1.2 Motivation

Musical improvisation is a complex phenomenon, and there have been many attempts to describe and model it [8, 14, 90]. Moreover, there is a lack of understanding the "music" in the current MIR research with respect to how humans actually perceive it [101]. Previous work on Western melodies showed that variable-length n-gram models and human judgments of melodic continuation are highly correlated [77]. We hope this research will bring clues about how we actually anticipate music [53] outside the Occidental boundaries.

Through the understanding of a musical style by computational methods, predictive or generative systems based on the style may be built. Such systems can be used as machine performers which would be able to improvise on-the-fly in interactive performances, meta-composers that would suggest improvisational ideas to other performers [55, 75] or as an educational tool that can help musicians to play and improvise in this particular style.

The vast unchartered aspects of the world musics remains as a major challenge in the field of music information retrieval (MIR) [66]. In order to further advance the state-of-the-art in MIR, the unique challenges brought by world musics should be considered [46]. Research involving paradigms such as heterophony in music in Far Asia [67], polyrhythms in African percussions [7] or *makam* theory in Turkish music [89] would immensely expand our knowledge and tools in music. Further computational research into the diverse musical genres throughout the world will deepen our knowledge of universal versus genre-specific aspects of music, and allow us to truly evaluate the generality of various modeling strategies. Moreover, the findings from various cultures might open up new paths for musical creativity, expressivity and interaction.

CHAPTER II

BACKGROUND AND RELATED WORK

2.1 Traditional Turkish Music

Throughout history, Anatolia and Thrace have been home to many civilizations and the crossing bridge between cultures. As a result, music in modern Turkey is as diverse as the numerous groups that have stepped foot on its soil. Traditional Turkish music has been one of the most important and influential traditions in the world [13, 15]. Traditional Turkish music has been a "melting pot", incorporating elements from numerous other musical traditions such as Hellen, Hittite, Byzantine, Armenian, Kurdish, Jewish, Arabic and Persian [45]. Western music theory and practice is insufficient to explain the uniqueness and the richness of the melodic structures (*makam*, Section 2.1.1.1) and metric structures (*usul*, Section 2.1.1.2) in traditional Turkish music. In order to comprehend traditional Turkish music, it is necessary to understand some basic concepts in Turkish music theory.

2.1.1 Basic Concepts in Turkish Music Theory

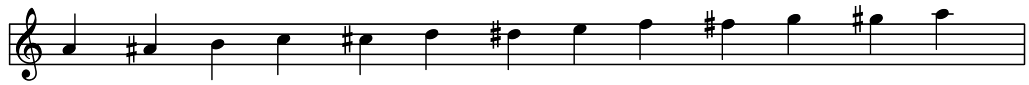
2.1.1.1 Makam

In Western classical music theory, the octave is divided into 12 pitches (Figure 1a). On the other hand, in traditional Turkish music, an octave can be divided into more than 12. At present, there is no theory that is completely agreed upon in Turkish music due to the differences in theory and practice [97], and the suggested number of pitches in an octave ranges from 17 to 79 [104]. Currently, education in *makam* music is based on Arel-Ezgi-Uzdilek theory [5, 44]. However, Arel-Ezgi-Uzdilek theory is highly criticized among contemporary scholars [73, 80, 89, 97]. As a result, this section tries to form a basic picture of Turkish music theory that draws from various contemporary theories. Nevertheless, some of the contradictions between theory and practice will be pointed out throughout this section in order to inform the reader and also to explain some of the decisions undertaken in the thesis work.

Throughout history, traditional Turkish musics has predominantly been an oral tradition. Nonetheless, there have been different notation systems used throughout centuries, and Western staff notation was adapted by the start of the 20^{th} century [80, Chapter 1]. To indicate intervals smaller than semitones, flat and sharp symbols are altered to make special symbols. It is acceptable to use the note naming conventions coming from Western classical music (*La, Si, Do* or *A, B, C*) and the traditional names interchangeably. The traditional names may indicate the octave of the note (*tiz* to indicate one octave higher and *kaba* to indicate one octave lower) or it may completely change in different octaves (see *Dügah* and *Muhayyer* in Figure 1b). Traditional names are also emphasized more in practice, and it is crucial to learn the traditional names not only to understand the cultural background and but also the musical structures [97, Page 133].

According to the Arel-Ezgi-Uzdilek theory, a whole tone is divided into 9 intervals named *koma*. Out of 53 *koma*s in an octave, only flats and sharps of the 1^{st}, 4^{th}, 5^{th}, 8^{th} and 9^{th} *koma*s are used to discretize an octave into 24 consequent tones [69]. In Turkish folk music, due to the selection of instruments (section 2.1.2.1), there may be a single note between the semitones. This tone is notated either by a special flat ($\flat^2$) or a sharp ($\sharp^3$) symbol adopted from Western classical notation. The number written on the right top of the accidental symbols indicates the *koma* distance from the natural note [1]. It should be noted that in the Turkish folk music practice, the *koma* distance is not important; it is merely used to indicate that the pitch lies between a semitone. Therefore, it makes more sense to treat these as quarter-tones

[1] Notice that *koma* values of 2 and 3 are not used in Arel-Ezgi-Uzdilek theory.



(a) Western classical music

(b) Turkish folk music

Figure 1: Notes in an octave in Western classical music and in Turkish folk music. The traditional names of the notes in traditional Turkish music are given below each symbol.

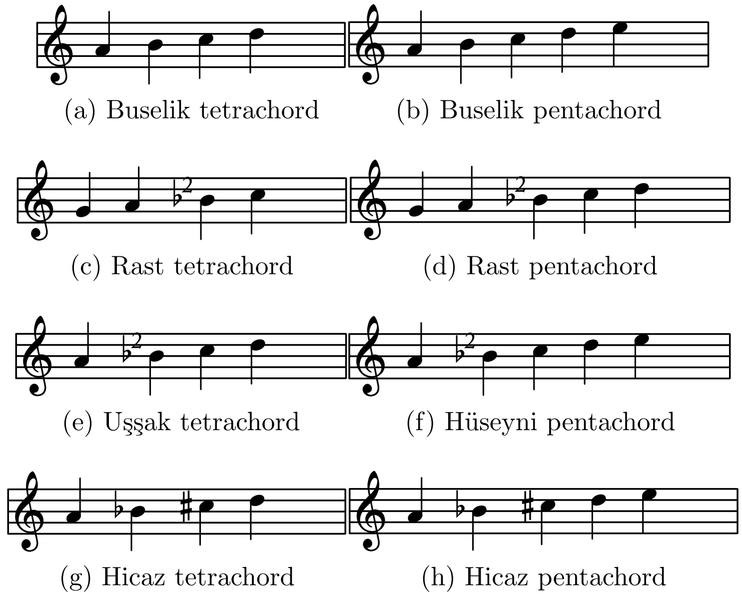
[2] with a non-deterministic deviation from its neighbors. When all of the notes are arranged, there exists 17 notes in an octave (Figure 1b). It should also be noted that the notes are not tuned according to the reference note, $A = 440Hz$, and they are not well-tempered.

The melodic structure of traditional Turkish music is explained by *makam*s. *Makam*s may be considered as the modes of traditional Turkish music [89]. Since the music is based on modality rather than tonality, it makes more sense to talk about a modal center rather than using terms such as tonic, dominant [3]. Typically, melodies in *makam*s have a *başlangıç* (starting, initial) tone and a *karar* (ending, final) tone [73].

The melodies in *makam*s are built by using tetrachords (*dörtlü*) or pentachords (*beşli*) [69] (Figure 2). Tetrachords and pentachords are explained by the traditional

[2]Harvard Dictionary of Music defines a quarter tone as ”an interval equal to half a semitone” and a microtone as ”an interval smaller than a semitone [83].” Therefore, ”microtone” is more suitable for the English explanation since the tone may not be equal to half a semitone. However, in Turkish folk music discourse, the term *çeyrek ses*, which can be literally translated as quarter-tone, is used to indicate the single tone smaller than a semitone. As stated in Section 1.1, I would prefer using the Turkish terminology.

[3]On the contrary, Arel-Ezgi-Uzdilek theory uses tonal terminology such as dominant (güçlü) [69].



(a) Buselik tetrachord (b) Buselik pentachord

(c) Rast tetrachord (d) Rast pentachord

(e) Uşşak tetrachord (f) Hüseyni pentachord

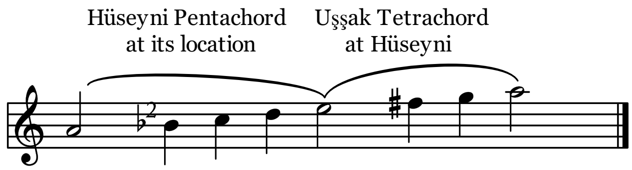
(g) Hicaz tetrachord (h) Hicaz pentachord

Figure 2: Some important pentachords and tetrachords in Turkish traditional music. They are written "at their locations."

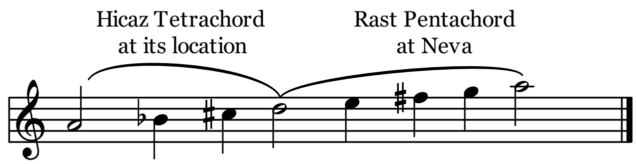
names associated with them and the specified starting note. When a tetrachord or pentachord is said to be played "at its location" (*yerinde*), it will start from the default starting note of the sequence. For example, the default starting note of *Rast* tetrachord/pentachord is *Rast* (G), whereas *Hicaz* tetrochord/pentachord at its location starts from *Dügah* (A) [4].

Each *makam* has some peculiarities in melodies, which are explained as "melodic nuclei" (*ezgi çekirdeği*) [70], "characteristic motifs" [97] and "tunes specific to a particular makam" [27] by different Turkish music scholars. It can be said that *makam*s are formed by "navigating" (noun: *seyir*) around these melodic progressions. *Makam*s can have ascending, descending or ascending/descending *seyir*s such that two pieces having the same key signature but having different *seyir*s might also have different *makam*s [69].

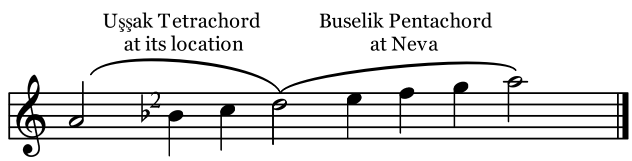
[4]Since a lot of makams, tetrachords, pentachords and notes share the same traditional name, readers should be careful to understand what is being referred. As an example "Hicaz pentachord in Rast" means the Hicaz pentachord starting from the Rast note (G).



(a) Hüseyni *makam* at its location



(b) Hicaz *makam* at its location



(c) Uşşak *makam* at its location

Figure 3: Hüseyni, Hicaz and Uşşak *makam*s at their locations.

To make the explanations more concrete, let's investigate three of the most used *makam*s in traditional Turkish music: Hüseyni, Hicaz and Uşşak at their locations (Figure 3). All of these *makam*s have the same *karar* (ending) note, Dügah (A), however they differ from each other in key signatures, *seyir*s and *başlangıç* (starting) notes. Hüseyni makam (Figure 3a) is formed by Hüseyni pentachord at its location (A) and Uşşak tetrachord at Hüseyni note (E). The *başlangıç* (starting) note of the makam is Hüseyni (E). *Seyir* of the *makam* may be both ascending and descending. Hicaz (Figure 3b) is formed by Hicaz tetrachord at its location (A) and Rast pentachord at Neva note (D). The *başlangıç* note of the Hicaz makam is Neva (D). Again, the *seyir* can be both ascending and descending. In both Hüseyni and Hicaz, Mahur ($F\sharp$), may be played as Eviç ($F\sharp^3$) [5] or Acem ($F\natural$). Acem ($F\natural$) is a typical case in the descending *seyir*. Uşşak *makam* (Figure 3c) is formed by Uşşak tetrachord at its

[5]Notice that in both Hicaz and Hüseyni makams, to comply with the intervals in the upper tetrachord/pentachord, the note should actually be $F\sharp^3$. Also in practice of Turkish folk music, typically $F\sharp^3$ is the played note. This shows another contradicting representation between Arel-Ezgi-Uzdilek theory and folk music practice.

location (A) and Buselik pentachord at Neva note (D). Notice that the key signature of Uşşak *makam* is almost identical to Hüseyni. However, the *başlangıç* note of this *makam* is Neva (D).

2.1.1.2 Usul

In traditional Turkish music, the metric structure is explained by *usul*. *Usul* is "the structure of musical events which are coherent with respect to time [69]." *Usul* can be roughly translated to "meter." Nevertheless, *usul* has a broader meaning than meter: The *usul* of a piece may emphasize the *makam*, and a change in *usul* may affect the *seyir*s and the *makam* [97]. An *usul* can be as short as two beats or as long as 128 beats. However, it should always have at least one strong and one weak beat [69].

Analogous to different *makam*s having the same accidentals, there can be different *usul*s with the same number of beats due to the difference in the timings of the beats. Turkish music also makes a rich use of *usulsüz* (non-metered) progressions. *Usulsüz* sections are typically the improvised parts of the pieces in traditional turkish music.

2.1.2 Turkish Folk Music

Turkish folk music is a profound music style that is the product of the emotions, thoughts, humor and social life of Turkish people. Turkish folk music is typically anonymous, and the songs have been carried from generation to generation as an oral tradition.

In Turkish folk music, variations in performance practice and expression may correspond to regional differences. Every region in Anatolia has a peculiar style, explained by the term *tavır*. *Tavır* constitutes the playing, singing styles and techniques particular to the region. However, regional folk artists often devise their own styles and do not strictly adhere to the regional style.

Figure 4: Aşık Veysel, one of the most famous folk artists of 20^{th} century, playing bağlama

2.1.2.1 Instruments

Turkish folk music uses a large number of plucked, bowed, wind and percussive instruments. The most characteristic instrument family of Turkish folk music is the *saz* family, which consists of plucked string instruments native to Anatolia and the surrounding geographies (Greece, Balkans, Caucasus, Iran...).

The most common *saz* played in Anatolia is bağlama (Figure 4). It typically has 17 notes in an octave [71]. The frets are tied to the fretboard instead of pinning them. As a result, frets are easily moveable, and microtonal adjustments in the temperament can be made to play in different *makam*s and/or to emphasize *tavır*. Typically, other instruments (if any) are tuned with respect to bağlama, therefore it is safe to say that theory and practice in Turkish folk music is centralized around bağlama [97, Chapter 17]. A thorough analysis of bağlama also suggests the theory behind Turkish folk music is the same with the Ottoman classical music [97, Chapter 17–18] [6].

[6] Historically Arel-Ezgi-Uzdilek theory leaves Turkish folk music outside its scope. On the other hand, contemporary scholars agree that the melodic structures in Turkish folk music are explainable by *makam*s [72, 93, 104]. This issue points out another weakness in the Arel-Ezgi-Uzdilek theory.

2.1.2.2 Uzun Hava

In Turkish folk music, the pieces can be categorized into two groups with respect to *usul*. Pieces with definite *usul* are named *kırık hava*s, whereas pieces that incorporate *usulsüz* (without any meter) sections are typically named as *uzun hava*s. In the *usulsüz* sections of an *uzun hava*, the performer improvises in notes and timings while maintaining certain *seyir*s. The improvisations usually converge to modal centers and therefore *uzun hava*s can be explained by *makam* theory. In this sense, *uzun hava*s can be considered as structured-improvisation pieces. They are typically played in Hüseyni [54].

Typically *uzun hava*s are performed by a single performer, who also plays bağlama. The music is usually sad; the lyrics (if any) are generally about the daily struggles and emotions of Anatolian folk. There are various types of *uzun hava*s across Anatolia, which may differ from each other *tavır*s and choice of *makam*s.

2.2 Related Works in Music Technology

Computational models of music are typically used to understand various aspects of music through statistical means. The modeling is mainly aimed at two applications: 1) predictive systems, 2) generative systems. Predictive systems aim to guess future events in music by taking peculiar aspects of the musical style into consideration, whereas generative systems attempt to create music. Note that the practical applications may take both roles in its implementation and execution: a predictive system might form the basis of the framework of a real-time interactive system, which keeps the track of previous events, and generates the next event by prediction [74]; and a generative system may be used to assess the consistency of the model with respect to human expectations [82].

Computational modeling of musical styles is not a new topic in the field of music technology. One of the earlier attempts of computational modeling is done by Xenakis [103], where he generates music through statistical distributions. Ebcioğlu has developed a system named CHORAL, which can harmonize four part chorales [41]. CHORAL decides on musical events by consulting parallel representations of music. The rules are set from baroque music theory. Another approach is the recombinant model. In this model, musical patterns are predetermined; by various alternation and combination algorithms, there patterns are stacked together to form the music. The musical snippets may be, again, defined by the human developer, or they might be gathered by (supervised or unsupervised) data mining techniques. Recombinant modeling is one of the techniques adapted in David Cope's EMI, which is one of the earlier attempts (and probably the most controversial so far) of a machine composing in a particular composers style [35, 36].

The modeling system in the thesis follows another scheme, in which the analysis and modeling of the musical style in music is mostly left to the computer. The algorithms are typically machine-learning algorithms. One of the noteworthy examples is the artificial neural network (ANN) approach taken by Toiviainen [94]. By using ANNs, he has modeled bebop style improvisation, and showed that ANNs are able to create variances of the trained music.

Markov models and n-gram modeling are two closely related and common techniques in computational modeling. Ames states that Markov models are common tools in algorithmic composition [3], and explains various methods to incorporate the models into musical applications. Assayag and Dubnov, with their colleagues, have extensively worked on the performance of Markov models and dictionary-based methods for computational modeling of musical styles. Dictionary-based methods may be interpreted as different representations of Markov models, which may give better performances than a regular Markov model. In [9, 58, 39], the authors have described and compared two dictionary named prediction methods for automatic music style modeling, namely the Incremental Parsing (IP) method and the Prediction Suffix Trees (PSTs). Later, they have included factor oracles [8] in performance comparisons, and developed the so-called "audio oracles" [38] from the the factor oracles. They have applied these methods successfully to various musical styles from J.S. Bach's chorales to improvised jazz pieces.

Pachet's the Continuator [74] is one of the state-of-art interactive generative systems. The system is based on varible-length Markov models (VLMMs). The Continuator is able to interact with human performers: It listens to the MIDI streams played by the performer, and generates continuations of the input. The user can choose the generative scheme from a number of continuations derived from the parameters of the MIDI data. Moreover, the outcomes from the VLMM is processed so that the Continuator's output is consistent with the polyphony, rhythm and musical progressions.

The framework used in the thesis is mostly based on the so-called "multiple viewpoints" modeling (MVM) (Section 4). It is introduced by Conklin et al. [30, 31, 32, 33, 34] and further developed by Pearce et al. [76, 77, 79]. MVM has made a major impact in the machine-learning algorithms on music. MVM basically uses parallel descriptors to represent music. It is a general framework, and its power comes from the flexibility of viewpoints defined to represent the musical phenomenon. Moreover, long-term and short-term modeling is integrated to capture the general context of the musical style along with peculiar characteristics of the current song [32]. The system uses entropy-based methods to merge the long-term and short-term models, and also to quantitatively evaluate the predictions given by the computational model [63]. Recently, Pearce et al. has also showed that this n-gram modeling scheme may show a significant resemblance to the musical expectations in the human mind [78].

Even though information retrieval in world musics has recently started to attract attention in academia, there has already been substantial amount of research in the field [61, 99]. The research topics range from the automatic segmentation and transcription of tabla [26, 49] to robotic musicianship and hyperinstruments in Indian classical music [56], and mode recognition in North Indian classical [22] and African musics [65]. European folk musics have been studied in detail, probably because they are similar to and culturally shared with Western musics. Chai and Vercoe have presented a classification method based on HMMs for European folk musics [19]. The physical properties of Elgin Avloi, an ancient Greek instrument has been modeled and simulated by Tsaxalinas et al [96]. Eerola et al. has shown that statistical similarity methods capture the listeners' similarity ratings by conducting experiments on European folk songs from different ethnics. Conklin and Anagnostopoulou have done melodic pattern mining on Cretan folk songs [29]. By using a measure defined as "relative empirical probability," the authors present short patterns peculiar to the song types and locations across Crete. Krumhansl et al. has presented a comprehensive study of melodic expectation in Finnish spiritual folk hymns [57]. They have observed that the musical familiarity and expertise in the style alters the melodic expectations, and the neural network models of the self-organizing map (SOM) emit expectations similar to the reactions of human subjects.

Most definitely, Chordia et al.'s research on tabla [21, 24, 25], an Indian percussion, stands out as the most related research to the thesis work. The research is aimed at the computational modeling of the tabla sequences by using multiple viewpoints modeling. Tabla possesses a relatively simple musical language, where the name of each stroke indicates a peculiar timbre [40]. Moreover, in some forms of tabla music (such as *qaida*), the melodic instrument keeps on playing a rhythmic loop, while the percussion plays solo improvisations centered around a theme [40]. Therefore, some of the tabla music may be interpreted as structured improvisation pieces played by a quasi-melodic instrument. Apart from the rhythmic dissimilarities, we can draw some parallelism in the symbolic analysis of *uzum hava*s and *qaida*s. What is more, the multiple viewpoints framework used in the thesis is adapted from this series of studies with minor changes. Even though the success of the multiple viewpoints modeling lies in the choice of parallel representations (Section 4.3), the conceptual similarities between the two musical forms imply that the results might be consistent with each other.

Parallel to the other world musics, computational research on traditional Turkish music has recently emerged. Erkut has worked on physical modeling of tanbur, a traditional Turkish stringed instrument [43]. Holzapfel's PhD. thesis deals with similarity measures of ethnic musics, and puts an emphasis on traditional musics of Greece and Turkey [51]. The most comprehensive research in traditional Turkish music is done in evaluating current music theories [1, 18, 47] and classifying makams [17, 48, 60] through the use of pitch class histograms. Additionally, a novel *makam* classification algorithm based on n-gram modeling has been presented by Alpkoçak and Gedik [2].

To the my best of my knowledge, there has been a single work on the analysis of melodies in traditional Turkish music. Güngör Gündüz and Ufuk Gündüz have made mathematical analysis of 4 Ottoman classical music and 2 Turkish folk music pieces [50]. Similar to the thesis, the symbolic notations provided by Turkish Radio Television Corporation (TRT) are used for analysis. The paper checks some of the properties of the songs such as fractal dimensions, self-similarities, note progressions and organizational behaviours. However, the musical explanation of the traditional Turkish music is strictly based on the Arel-Ezgi-Uzdilek theory. For example, they state "the folk music is free of all makams." As thoroughly explained in Section 2.1, this is a false claim. Moreover, the songs are treated as "complex mathematical systems", and some of the results are not discussed thoroughly with respect to their musical meanings: The authors show that there are very frequent transitions from order to disorder in the songs, however they do not discuss whether these transitions might correspond to a significant change in the symbolic notations or in human anticipation. Therefore, the current state of the research is not as fruitful from a musical or an ethnomusicological point of view.

Beyond debate, databases are crucial elements of statistical analysis of any kind of data. Unfortunately, there is a lack of machine-readable databases dedicated to traditional Turkish musics. To the best of my knowledge, there are currently two databases for traditional Turkish music that are ready to be used for machine processing. The first one is the traditional Turkish music MIDI database, which is distributed along with the music notation software Mus2 [7]. The second one is the *Parametric Turkish Music Database* [16], which consists of the pitch tracks of the audio recordings of a vast number of Ottoman classical music pieces. I believe, this absence is a key factor in the lack of statistical research involving traditional Turkish folk music.

Parallel to Conklin et al. and Pearce et al.'s research [33, 76], the computational framework in this thesis (Section 4) incorporates multiple viewpoints modeling (Section 4.3) with both long-term and short-term models (Section 4.4). Variable-length Markov modeling (Section 4.2.1) is used to model the sequences, and the training data is stored as Prediction Suffix Trees (Section 4.2.2). The evaluation of the system is done by entropy-based calculations (Section 5.3). In a limited fashion, the system is capable of generating melodic patterns by either picking the mostly likely or a random event from the probability distribution in the next step (Section 5.4).

2.3 Contributions and Novelty

To the best of my knowledge, the *Uzun Hava Humdrum Database* is the first symbolic notation database of *uzun hava*s in machine readable format. Though the database cannot be considered as a novel contribution by itself, it would hopefully help to

[7] http://www.mus2.com.tr/en/

satisfy the increasing demand in academia of accessing various musical traditions.

The computational modeling framework is based on multiple viewpoints modeling (MVM) a general, flexible modeling technique for melodic sequences. From a technical point of view, this work stands out as the first usage of variable-length Markov models (VLMMs) and multiple viewpoints modeling (MVM) in traditional Turkish music. It also presents the first attempt of computational modeling of traditional Turkish music. The novelty of the thesis lies in the representations specifically defined for the analysis of *uzun hava*s: The work proposes novel *pitch-related* viewpoints that addresses the key relationships in the 17-tone scale of Turkish folk music.

CHAPTER III

SYMBOLIC DATABASE

For the computational modeling, an extensive database is built, which is comprised of symbolic notations of *uzun hava*s. To the best of my knowledge, the database is the first machine-readable, symbolic notation database of *uzun hava*s in Turkish folk music, a relatively untouched phenomenon. Therefore, the database is an important contribution to academia.

The database is aimed at enabling easy access to the *uzun hava* form. The database may help scholars from various disciplines to focus on the analysis of the Turkish folk music rather than setting up a database, and diversify the statistical research in traditional Turkish music. I believe that the database will be especially useful for scholars, who are unable to gather machine-readable data by themselves, and who do not have the knowledge or the resources to construct a database of Turkish folk music. I also hope that the database will open up a path for the analysis of Turkish folk music in cross-cultural and cross-genre music research, especially in research dealing with improvisation.

3.1 Overview

The *Uzun Hava Humdrum Database* is a collection of symbolic notations of *uzun hava*s, a structured improvisation form in Turkish folk music [1]. The database currently encompasses 77 songs from different regions of Anatolia, Iraq and Caucasus. It consists of 10849 notes in total, in 8 *makam*s (Table 1). Unsurprisingly, the *makam*s of the songs are biased towards Hüseyni (Section 2.1.2.2). The notations are encoded

[1] The *Uzun Hava Humdrum database* is available online at http://sertansenturk.com/uploads/uzunHavaHumdrumDatabase

Table 1: *Makam*s in the *Uzun Hava Humdrum Database* and the number of songs per *makam*

***Makam*s**	**# of Songs**
Hüseyni	40
Hicaz	15
Uşşak	13
Rast	3
Kürdi	2
Nikriz	2
Segah	1
Karciğar	1
Total	77

in the Humdrum based syntax called **kern format [52]. The database is constructed with the help of Prof. Erdal Tuğcular (Department of Music Education, Gazi University, Ankara, Turkey).

The original source of the symbolic notations is the Turkish folk music database of the Turkish Radio and Television Corporation (TRT) [2]. The TRT folk music database is the largest symbolic database of Turkish folk music, having more that 7250 pages of sheet music, notated in modified Western staff notation and saved in .tiff image format. The database holds *kirik hava*s and *uzun hava*s picked from different regions of Anatolia, Thrace, Middle East and Azerbaijan. There are 123 scores of *uzun hava*s in the Turkish Radio and Television Corporation's (TRT) Turkish Folk Music Database.

3.2 Problems and Decisions in Setting Up the Database

Although the TRT database provides a nice set of symbolic notation, on the basis of comments from Prof. Erdal Tuğcular and oral discussions with Okan Murat Öztürk (Başkent Üniversity State Conservatory, Ankara, Turkey), it can be said that the

[2] The *TRT Turkish folk music database* is available online at "*Türk Müzik Kültürünün Hafızası" Score Archive* (http://www.sanatmuziginotalari.com/), which is freely accessible via http://devletkorosu.com/.

TRT database is sometimes unrealistic, and it contains considerable errors. First of all, the notation is an adapted version of Western symbolic notation. In a musical culture, a notation system has to hold certain features and represent the practice in a satisfactory manner in order to be accepted in practice [80, page 13]. Therefore the sufficiency of the symbolic notation might be debatable, especially when expressivity of a musical culture is transferred from the teacher to the student by oral tradition. Moreover, for world musics, there are some intrinsic problems of using symbolic notation and making deductions solely based on them [68, Chapter 2–3]. This may also raise questions about the validity of using Western notation as a valid representation of improvisation in Turkish folk music. In fact, representation of traditional Turkish music have undergone a series of contradictions during the adaptation of Western symbolic notation [80, 97]. For these reasons, some musicologists suggest that audio recordings should be the basis of analysis in ethnomusicology [6, 12]. Yet, since audio analysis is generally not as easy and straightforward as processing symbolic data, symbolic notation is at least an adequate choice for initial steps in computational analysis.

Second, the quality of transcriptions between the transcribers and the pieces varies considerably. Moreover, there are some transcription mistakes such as missing key signatures and temporary accidentals (such as $F\sharp^3$'s in Hüseyni and Hicaz), which were corrected manually in the *Uzun Hava Humdrum Database*. As another example, in a couple of songs the *usul* of the piece changes in every measure, while the piece should be *usulsüz* (as an example check U0218 in the TRT database): They clearly show that the transcriber attempts to divide the piece into melodic phrases in a way that totally disregards the *usulsüz* nature of *uzun hava*s! while acknowledging these facts, symbolic notation was chosen as the input since the thesis aims to be the first step of computational modeling in Turkish folk music.

To read the image files in .tiff, three optical music recognition (OMR) softwares were tested. At first, Audaveris and SharpEye were tried. Upon checking some files, Audaveris was not found to be satisfactory. For SharpEye, the accuracy was fine, however the lyrics recognition made the files too bulky and difficult to clean-up. Finally, the built-in SmartScore 5 Lite in Finale 2010 was chosen. Apart from some handwritten scores, the optical character recognition in SmartScore 5 Lite was fairly accurate. Nevertheless, all of these software are constrained within Western classical music theory. As a result, there were some major conceptual problems in the accuracy of the recognition software.

The simplest (and expected) error is that the OMR system will not be able to recognize characters special to the music. Moreover, the system may misidentify these special characters. The recognition of $\flat^2$ as $\flat$ lies in this category. Note that this type of errors does not bring any critical failures, and it is easy to automate the correction.

The second type of errors is due to hierarchy. Typically, the character recognition system is forced to conceive the music under the assumption of Western classical music tonality and metric structure. Therefore, even though the system can recognize the musical symbols in low-level, high-level algorithms force these symbols to be disregarded or altered in an inappropriate manner. Without a surprise, OMR fails in recognizing *makam*s and *usul*s in Turkish.

As the computer is constrained to typical metric structures seen in Western music, it is either unable to or not confident in recognizing compound rhythms (5/8, 9/8, 11/8 etc.). Moreover, if there are frequent changes in meter, OMR may find it hard to track the meter. Finally, free improvised sections might be a major problem. While Western classical music until 20^{th} century does not incorporate such elements extensively, they are inseparable elements of traditional Turkish music. The main issue arises when the algorithm tries to restrain the music to simple meters: once it believes the duration of measure has been filled, it may disregard the upcoming notes or write them as harmony. In both cases, the *usual* and *seyir*s will be disrupted, which requires careful manual corrections.

Another ominous error happens when the system tries to recognize the modal structure under the rules of tonality. When the system faces a mode with a key signature which is not present in Western classical music tonality, it will try to match the key signature into one of the scales known to it. In order to do that, it might either try to add or remove some accidentals. As a result, the piece may sound completely different. As a practical example, this problem is inevitable in the Hicaz *makam*. Since this makam has both flat and sharp signs (Figure 3b), OMR typically deciphers the signature as B♭, which is the key signature of *F major* (or *D minor*) and adds a sharp to the first note it recognizes on the first measure of the staff and also attaches the same accidental to all of the consequent notes with the same pitch. Unless corrected manually, this would prove to be a fatal error, especially in research involving melodic analysis.

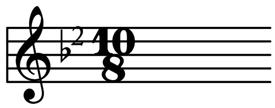
Consequently, quarter tone accidentals, *makam*s which do not follow Western tonality, and non-metered sections, not only remained unrecognized but also caused confusions in meter, scale and notation. Moreover, recognition of ties were problematic in OMR, and even though nearly all of the transcriptions are monophonic, the OMR system has occasionally created erroneous parallel harmonies. Due to these problems with OMR, handwritten scores and the scores with highly problematic character recognition are eliminated, and the number of pieces is reduced from 123 to 107.

After the recognition phase, the songs are saved in MusicXML 2.0 format, which is supported by Finale 2010. Apart from that, Finale 2010 was not used for any kind of processing. As the format of the database, the Humdrum based syntax called **kern format was chosen. The syntax provides ease in readability with broad search, comparison and editing capabilities, and it also supports microtonal deviations [52]. These features make Humdrum a well known and widely used toolkit in academia. In fact, the simple and systematic syntax of the **kern format proved very useful to correct the mistakes in OMR and add the missing aspects specific to *uzun hava*s: to take care of these issues in OMR some fully or semi-automated tasks were run on the **kern file. Almost all of the tasks are coded in either Python or bash scripting.

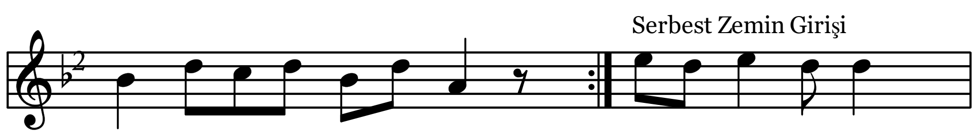
The first task was to convert the scores in MusicXML 2.0 format to **kern notation by using xml2hum [86]. Next, by using regular expressions, again in bash scripting, the **kern syntax was cleaned out of mistakes caused by ties and parallel harmonies. In the *Uzun Hava Humdrum Database*, the name, region, *makam*, key signatures and *usul* are printed to the start of the file as comments, which are filled manually.

Usulsüz (non-metered) sections in *uzun havas* are treated as cadenzas such that the sections start with "*MX/X", indicating the following notes will be played in a non-metered fashion. Each note is proceeded by the letter "Q", which is used to indicate gruppettos in **kern format [52]. The meter changes are manually entered in the **kern encoding, and then gruppetto symbols are added to the corresponding notes by using Python.

In order to comply with the standard Humdrum notation, quarter tones are indicated as deviations in cents in a second spine. This second spine is created in Python with the default value of 0 deviation for notes and "not-applicable" deviation for rests (indicated by "."). Then, the quarter tone accidentals are corrected song by song via regular expressions in TextWrangler. In the TRT database, there are accidentals, which have different koma deviations from the same tone ($B\flat^2$, $B\flat^3$, $B\flat^4$ etc.). However, as explained in Section 2.1, the most common instrument played in *uzun havas* is bağlama, and it has 17 notes per octave. Moreover, the theoretical and actual pitch values in Turkish folk music do not match, and they might even deviate from region to region and even from performer to performer. Therefore, all *koma* values lying between semitones are mapped into a single quarter-tone with 50 cents deviation from the original note to match the 17-tone scale. The missing temporary accidentals in the TRT database are also included in this step (Figure 5).



```
!!!Rep U0368
!!!Isim(Name): Yar Yad Oldum
!!!Yore(Region): Elazığ
!!!Makam: Uşşak
!!!Donanim(Key Signature): b-2
!!!Usul: 10/8, Serbest
**kern  **Dcent
*staff1 *.
=1- =1-
*clefG2 *.
*k[b-]  *.
*M10/8  *.
```




```
=11 =11
!!linebreak:default
4b- 50
8dd 0
8cc 0
8dd 0
8b- 50
8dd 0
4a  0
8r  .
=12:!   =12:!
*MX/X   *.
8eeQ    0
8ddQ    0
4eeQ    0
8ddQ    0
4ddQ    0
*-  *-
```

Figure 5: The key signature, *usul* at the start and the last two measures of the *uzun hava*, U0368, followed by the corresponding **kern syntax. The word "Serbest" (tr: free) indicates the start of the *usulsüz* (non-metered) section

CHAPTER IV

COMPUTATIONAL MODELING

This chapter is dedicated to the explanation of the computational modeling framework. The framework aims to construct a model of the metric and melodic structures in *uzun hava*s, a structured non-metered improvisation form in Turkish folk music by using various parallel descriptors for the pitches and durations of the notes. The model is then used to predict the melodic continuations in *uzun hava*s.

The framework is based on n-gram modeling (Section 4.1), and variable-length Markov models (VLMM, Section 4.2.1) are used to train the computational model of *uzun hava*s. VLMMs are stored in Prediction Suffix Trees (Section 4.2.2) for better performance. During the selection of the next symbol, the predictions from each level of the tree are smoothed (Section 4.2.3) to include both the general structure and specific patterns inside the trained sequences. If the system is asked to predict a symbol which has never occurred before, it assesses its confidence by checking the the number of single occurrences (Section 4.2.4). The real power of the system comes from the so-called "multiple viewpoints modeling" (Section 4.3), where each event in a musical sequence are represented by parallel descriptors. The predictions from each viewpoint is obtained by consulting a long and a short term model (Section 4.4), which are trained on the entire database and the specific song respectively.

4.1 n*-gram Modeling*

N-gram modeling is a commonly used technique to probabilistically model sequences of elements such as phonemes in natural language processing [63], and music [37, 92]. A n-gram is simply a subsequence consisting n items from a given sequence.

To demonstrate n-gram modeling, it is convenient to study a simple musical example: Let's take the last two measures of U0368 again, ignore the repeat sign between the measures. Also, let's change the *Dügah* (A) note at the 7^{th} step to *Çargah* (C) to decrease the sparsity in the sequence (Figure 6) [1]. If we are to count the notes and rests in this sequence, we will observe the unigrams shown in Table 2. The counts of unigrams brings an interesting outcome: in this sequence, the note *Neva* (D) is played the most, implying it has more importance than the others. In fact, as indicated in the database (Figure 5), U0368 is in *Uşşak makam*, in which *Neva* (D) is one of the modal centers.

While even the simple unigrams present a very powerful point, they do not tell much about how the melody progresses. Just by looking in the unigram counts, we cannot conclude whether the sequence was formed just by repeatedly playing a note and then moving to the next note (i.e. $B\flat^2B\flat^2$CCDDDDDDEE) or there is a tendency to converge to *Neva* (D). Therefore, it might be a good idea to increase the size of the n-gram. Let's check the bigrams and their counts (Figure 3). Now, we can argue that the melody is indeed converging to *Neva* (D). Moreover, we can see that the bigrams starting with *Segah* ($B\flat^2$) or *Hüseyni* (E) always end up in *Neva* (D). This brings us into another intriguing observation. By only looking at the unigrams, we would always conclude that the next note should be 46.15% *Neva* (D), regardless of the previous note. However, bigrams suggest that if the current note is either *Segah* ($B\flat^2$) or *Hüseyni* (E) then the following note will 100% be *Neva* (D)! Evidently, increasing the size of the n-gram might help us to point out the peculiarities not only in the *makam* but also in its *seyir*.

On the other hand, as easily seen in these examples, as the size of the n-grams are increased, number of possible n-grams also increases. Since the length of the

[1]This sequence will be used in all examples throughout this section, and it will be called as "the modified version of the last two measures of U0368."

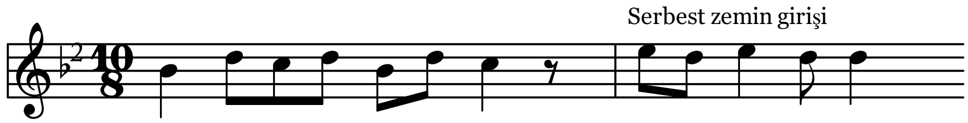


Figure 6: The modified version of the last two measures of U0368 from the *Uzun Hava Humdrum Database*. The repeat sign between the two measures taken out, and the Dügah note at the 7^{th} step of the original melody is changed to Çargah.

Table 2: Unigrams of notes and rests observed in the modified version of the last two measures of U0368 and their counts

Unigrams	B♭^{2}	C	D	E	Rest(R)
Counts	2	2	6	2	1

sequence does not change, this will result in a sparser space. Therefore, there is a limitation to the size of a n-gram model. Formally, as the order n increases, the maximum number of possible n-grams would increase to n^k, where k is the number of the possible symbols. However, even in large databases, most of the sequences will not be present or seen with a few examples. Notice that for the sequence in Figure 6, only 9 bigrams out of the 32 possible bigrams are observed, and there is typically a single count on most of these bigrams. This sparsity issue might lead to the so-called zero frequency problem [28] (explained in Section 4.2.4).

4.2 Markov Models

A Markov model is a causal, discrete random process. In a Markov model, every possible outcome is represented with a symbol, s_k, where $1 \leq k \leq N$ and N is the total number of the symbols. Each of these symbols is assigned to a state, which can simply be denoted as, k, the index of the symbol. The model can change from

Table 3: Bigrams of notes and rests observed in the modified version of the last two measures of U0368 and their counts

Bigrams	B♭^{2}D	CD	CR	DB♭^{2}	DC	DD	DE	ED	RE
Counts	2	1	1	1	2	1	1	2	1

one state to the other, and the probabilities of the next state, called the *transition probabilities*, depend only on the probabilities of the current and the previous states [81].

If the sequences are directly observable, i.e. the states are visible, most of the problems can be directly solved by dealing with transition probabilities. To familiarize with the concept, let's consider the simplest case: 1^{st} order Markov model. Mathematically speaking, the state transition probability, a_{ij}, of a 1^{st} order Markov model can be written as:

$$a_{ij} = P(\omega_{t+1} = j | \omega_t = i) \tag{1}$$

where ω is the state at the given time (i.e. the current state at time t, the next state at $t+1$), while i and j indicate the possible states from 1 to N. These transition probabilities can be arranged to form a NxN transition matrix. Since the transition probabilities have to hold the definitions of probability theory, the coefficients of the transition matrix should satisfy:

$$0 \leq a_{ij} \leq 1 \qquad \text{and} \qquad \sum_{j=1}^{N} a_{ij} = 1 \tag{2}$$

To further solidify, let's go back to the modified version of the last two measures of U0368 (Figure 6), and train a 1^{st} order Markov model from this sequence. Since there are 5 symbols in the sequence, there will be 5 states in the model. Let's define the symbol set as $S = \{B\flat^2, C, D, E, Rest\}$. From the counts found in Table 3, we can easily calculate the transition probabilities between the states (Equation 3). The graphical visualization of the Markov model is shown in Figure 7. The figure clearly shows this short sequence is centered around the *Neva* (D) note.

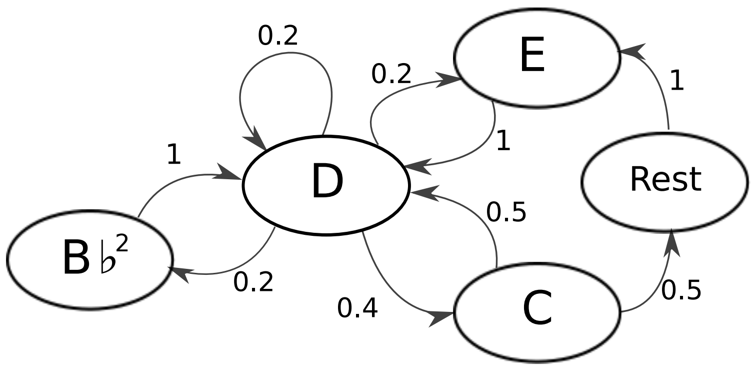


Figure 7: The first order Markov model trained on the notes of the modified version of the last two measures of U0368.

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.4 & 0 & 0 \\ 1 & 0.5 & 0.2 & 1 & 0 \\ 0 & 0 & 0.2 & 0 & 1 \\ 0 & 0.5 & 0 & 0 & 0 \end{bmatrix} \tag{3}$$

Nevertheless, it should not be forgotten that the training is done over a very sparse set of data. Clearly, with only 4 notes and the rest states, this model will not be of any practical use. This model will not be able to predict any other states accurately: it will not even be able to recognize *Dügah* (A), the *karar* (ending) note, of the *Uşşak makam* [2]. To overcome this problem, a Markov model should be trained over a large set of data, especially in music, which has a vast sample space. Moreover, escape probabilities might be introduced to the model to compensate the so called "zero-frequency problem" (explained in Section 4.2.4).

Up until now, we have considered a 1^{st} order Markov model. To generalize, in a n^{th} order Markov model, next state depends on the last n states. A $(n-1)^{th}$ Markov

[2]If the *D*ügah note was not changed to Çargah in the modified version, the model would obviously recognize the note. Nevertheless, the model will still be impractical.

model can be trained by n-grams with a size of n. In this case, the probability of the next state is defined as:

$$P(\omega_{t+1} = s_{t+1} | \omega_t = s_t, \omega_{t-1} = s_{t-1}, \omega_{t-2} = s_{t-2}, \ldots, \omega_1 = s_1) = \\ P(\omega_{t+1} = s_{t+1} | \omega_t = s_t, \omega_{t-1} = s_{t-1}, \omega_{t-2} = s_{t-2}, \ldots, \omega_{t-n} = s_{t-n}) \quad (4)$$

where s denotes the possible state at the given time. Higher order Markov models provide more specific information about the trends in a sequence. However, this improvement also brings a major disadvantage: As the order of the model is increased, possible transitions between states increases exponentially. A n^{th} order Markov model with N states will have a transition matrix of size N^n, rendering the model computationally expensive.

4.2.1 Variable Length Markov Models

As explained above, increasing the order of the Markov model might reveal more details about the data stream. However, specific patterns will get extremely uncommon as the order of the model gets higher, even with very big data sets. Moreover, while observing specific patterns is very helpful, sometimes general information might be as crucial. As an example, as higher order models tend to be sparse, the n-grams will be highly related to the training sequences, thus they might be less reliable. Therefore, in a generative algorithm, integrating lower order models to the system might also be useful to provide some regularity.

In order to capture the generality of lower order models and specificity of the sequences in higher order models, we can use an ensemble of Markov models with different orders to form a variable length Markov model (VLMM). The variable length of memory in contrast with fixed Markov model yields a rich and flexible description of sequential data.

4.2.2 Prediction Suffix Tree

As mentioned in Section 4.2, while higher order reveals more specific patterns in a data stream, the possible number of transition probabilities are increased enormously. The computational expense will even be greater, if we are working on VLMMs, where the predictions will be dependent of an ensemble of Markov models with different orders. Even though most of the probabilities in higher orders may probably be equal to zero, a straightforward implementation of a variable length Markov model requires to check all possible transition probabilities.

This problem of dimensionality might be eased by using Prediction Suffix Tree (PST) [84]. PST can be as depicted an alternative representation of VLMM. PSTs have also been applied to music [39, 95]. In PST, each symbol is represented as a node along with its count and probability. The root of the tree holds the current state of the model. In every level, a node (parent) may be connected to children nodes in the next level, which are the possible picks as the next state. By traversing from one level to the next one via one of these these branches, we can observe the n-grams, seen in the sequence, in increasing size. As an example, the first level is composed of the unigrams observed in the sequence, the second level indicates the possible bigrams starting in the states given in the first level, the third level constitutes the last symbol of the trigrams, and so on. To visualize the data structure, a diagram of the PST trained on the modified version of the last two measures of U0368 (Figure 6) is shown in Figure 8.

Calculation of the probabilities are pretty straightforward by advancing in the tree. Assume that a sequence has the symbol space of $S = \{s_1, s_2, \ldots, s_k, \ldots, s_N\}$, where N is the total number of the symbols. The probability of observing two arbitrary states, s_t and s_{t+1}, consequently can be computed as:

$$P(w_{t+1} = s_{t+1}, w_t = s_t) \triangleq P(s_t s_{t+1}) = P(s_t)P(s_{t+1}|s_t) \tag{5}$$

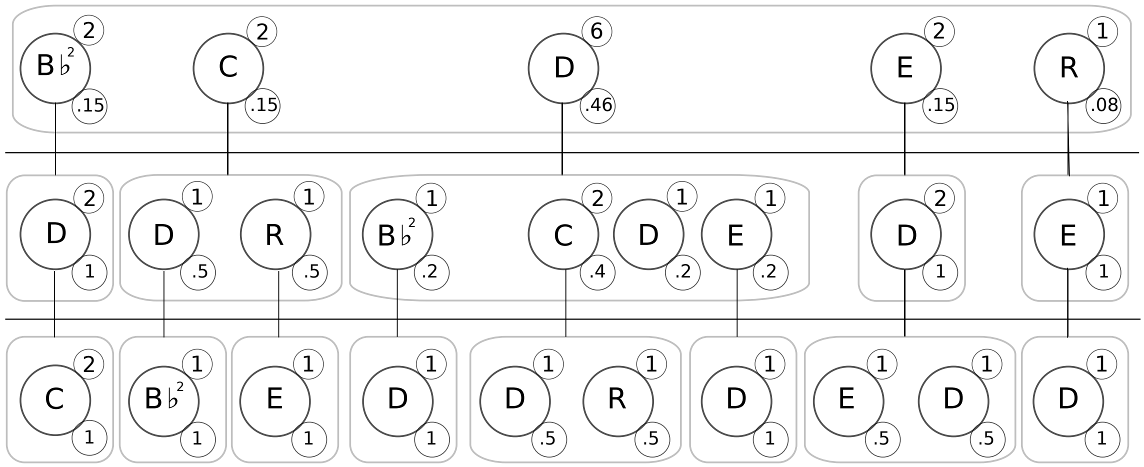


Figure 8: Prediction suffix tree representing the Markov models with a maximum order of 2, trained on the modified version of the last two measures of U0368. Bubbles on the top right and bottom right of each node denotes the count and the probability of the node respectively.

where $P(s_t)$ is the probability of observing s_t in the first level and $P(s_{t+1}|s_t)$ is the probability at the children node marking s_{t+1}, branching from s_t. To generalize, the probability of a particular subsequence of length m is given by:

$$P(s_t s_{t+1} \ldots s_{t+m}) = P(s_t)P(s_{t+1}|s_t)P(s_{t+2}|s_t s_{t+1})\ldots P(s_{t+m}|s_t s_{t+1} \ldots s_{t+m-1}) \tag{6}$$

which is simply calculated by starting from s_t at the first level, following the consequent nodes and multiplying the probabilities seen in these nodes. For example, probability of observing *Çargah* (C) followed by *Neva* (D) and *Segah* (B$\flat^2$) in the PST given in Figure 8 is simply:

$$\begin{aligned} P(CDB\flat^2) &= P(C)P(D|C)P(B\flat^2|CD) \\ &\approx 0.15 \cdot 0.5 \cdot 1 \\ &\approx 0.075 \end{aligned} \tag{7}$$

The result obtained from the PST will be exactly identical to a $m-1^{th}$ order Markov model trained on the same sequence. Moreover, since PST does not store the unseen n-grams, the calculation will be much faster than a regular VLMM implementation. Therefore we can benefit from this computational gain by including much higher orders and consequently the performance will be increased.

4.2.3 Smoothing

Even if a PST shows some characteristic patterns, the probabilities of such n-grams might be very small in higher orders due to the enormous size of the transition matrices of the higher order Markov models in the VLMM. Due to this problem, unless compensated, lower order Markov models will always dominate higher orders in a VLMM and making the system insensitive to context-specific patterns. In order to make up for the sparseness of the chains in higher order models, a method called smoothing is applied. There are two basic types of smoothing algorithms: back-off models and interpolation models.

Starting from the maximum order in the tree, back-off models search for the a sequence with a count exceeding a certain threshold. If there are no matches, the first element in the sequence is dropped, and the new sequence is searched again in the n-grams one size smaller. This process is continued until a positive match is found.

In interpolation models, the predictions given by each Markov model in the VLMM are given a weight, which is proportional to the models order. Mathematically speaking, given the subsequence $\{s_{t-n+1}, s_{t-N}, \ldots, s_{t-1}, s_t\}$ of length n, probability of observing s_{t+1} in the next state by using a VLMM with a maximum order of n is given as:

$$
\begin{aligned}
P(\omega_{t+1} = s_{t+1}|c_{t+1}) &\triangleq P(\omega_{t+1} = s_{t+1}|\omega_t = s_t, \omega_{t-1} = s_{t-1}, \ldots, \omega_1 = s_1) \\
&= \Big(w_0 P(\omega_{t+1} = s_{t+1}) + w_1 P(\omega_{t+1} = s_{t+1}|\omega_t = s_t) \\
&\quad + w_2 P(\omega_{t+1} = s_{t+1}|\omega_t = s_t, \omega_{t-1} = s_{t-1}) + \ldots \\
&\quad + w_n P(\omega_{t+1} = s_{t+1}|\omega_t = s_t, \omega_{t-1} = s_{t-1}, \ldots, \omega_{t-n+1} = s_{t-n+1}) \Big) / \sum_{i=0}^{n} w_i \\
&= \frac{w_0 P(\omega_{t+1} = s_{t+1}) + \sum_{i=1}^{n} w_i P(\omega_{t+1} = s_{t+1}|\omega_t = s_t, \ldots, \omega_{t-i+1} = s_{t-i+1})}{\sum_{i=0}^{n} w_i} \qquad (8)
\end{aligned}
$$

where c_{t+1} denotes to conditions given by the preceding states in the VLMM, and w_i is the weight to be multiplied with the probability provided by the i^{th} order Markov model in a VLMM with a maximum order of N. Though this calculation seems complex, it is actually pretty straightforward. Starting from the unigram s_{t+1}, the n-grams forming the end of the subsequence is inspected one by one (i.e. $\{s_{t+1}, s_t\}$, $\{s_{t+1}, s_t, s_{t-1}\}$, ..., $\{s_{t+1}, s_t, \ldots, s_{t-N+1}\}$). By referring to the PST for these n-grams, the probabilities of the outer node (s_{t+1}) multiplied by the weight of the order of the fixed model (in other words, size of the n-gram-1) are summed altogether. Finally, this value is divided by the summation of weights up to the maximum order of the VLMM to get a single probability value for the particular state. This calculation is repeated for each possible state and a discrete probability distribution for the next state is obtained. If the mostly likely outcome is required, the state with the highest smoothed probability is picked.

In language modeling, there are several choices of smoothing methods [20]. In previous research [21], two smoothing methods termed as Kneser-Ney (KN) and $1/N$ were compared for musical sequences. KN was adopted directly from language processing because earlier work showed it to be a superior smoothing method in the context of natural language processing [20]. By using an entropy-based evaluation of predicted outputs (Section 5.3), the results suggest that $1/N$ scheme might be better in musical applications.

Later, a simple back-off model and a parametric approach based on generalizing the 1/N technique were compared [24, 25]. In terms of average perplexities (Section 5.3), the interpolated models outperformed the back-off model, yet the back-off model provided lower median perplexities. This means that back-off model works well as long as it finds a matching pattern in the high orders, but interpolation methods are preferable if occasional bad misses are not tolerated. Additionally, when $1/N$ and the parametric model were compared, there wasn't a significant increase in performance.

In this work, $1/N$ method is picked in order to minimize the bad misses. In this smoothing method, the weight for the model with a order of i is given by:

$$w_i = \frac{1}{(N - i + 1)} \tag{9}$$

where N is the maximum order of the VLMM. $1/N$ method provides a greater weight to higher orders relative to the lower orders.

Let's give a solid example to make the procedure clearer. Given the PST in Figure 8, and the "smoothed" probability of observing the note *Segah* ($B\flat^2$) after *Çargah* (C) and *Neva* (D) is given as:

$$\begin{aligned} P_s(\omega_{t+1} = & B\flat^2 | \omega_t = D, \omega_{t-1} = C) \\ &= \frac{w_0 P(B\flat^2) + w_1 P(B\flat^2|D) + w_2 P(B\flat^2|DC)}{\sum_{i=0}^{N} w_i} \\ &= \frac{\frac{1}{3}\cdot 0.15 + \frac{1}{2}\cdot 0.2 + 1\cdot 1}{\frac{1}{3} + \frac{1}{2} + 1} \\ &\approx 0.63 \end{aligned} \tag{10}$$

It is easily seen that smoothing amplifies predictions given by higher order models, while the lower order models are also contributing. By smoothing the predictions by a weight proportional to the order of the model, the system would be able to match longer musical structures, specific to the training dataset, while keeping the consistency of the general model by referring to the lower order models.

4.2.4 Zero Frequency Problem

As mentioned in Section 4.1, a considerable amount of the n-grams might not appear on training. As the size of the n-grams increase, the possible number of n-grams grows exponentially and this sparsity issue becomes inevitable. Practically speaking, even if a large database is at hand, a PST will not be able to cover all possibilities, and consequently the system might argue the probability of very obvious, context-specific patterns should be zero! For example, the PST in Figure 8 will not be able to predict any symbols apart from Segah (B$\flat^2$), Çargah (C), Neva (D), Hüseyni (E) and rest. Using this PST is therefore impractical since it will not even be able to properly understand, Uşşak, the makam it was trained on. This sparsity problem is addressed as the zero frequency problem [28].

In order to overcome the zero frequency problem, escape probabilities may be introduced. For each level of the PST, escape probabilities assigns a small amount of probability to symbols, which have never been observed before. When an event is observed at a branch, at which it has never occurred before, the escape probability is returned instead of 0. In the system, the Poisson distribution was approximated to calculate the escape probabilities [102]. The escape probability, $e(n)$, at the n^{th} level is given as:

$$e(n) = \begin{cases} \dfrac{T_1(n)}{N(n)} & \text{if } T_1(n) > 0 \\ \dfrac{1}{N(n)+1} & \text{if } T_1(n) = 0 \end{cases} \tag{11}$$

where T_1 is the number of symbols that have occurred exactly once and N is the total number of tokens at the n^{th} level so far. By allowing escape probabilities customized to the levels of the PST, we can get a better assessment whether the system would expect new n-grams. If the counts of the nodes at a level are typically one (which will be a tendency as the order gets higher), the escape probability will be high, whereas a lower escape probability will be emitted from a level holding common patterns. Notice that there is still a chance that every node in a level has a count higher than 1, i.e. $T_1(n) = 0$. To take care of this case, a special escape probability is emitted (Equation 11).

Going back to the PST example in Figure 8, the escape probabilities when a unseen event takes place, will be $\frac{1}{13}$, $\frac{1}{2}$ and $\frac{9}{11}$ in the 1^{st}, 2^{nd} and the 3^{rd} levels respectively. Notice how the system's expectation of encountering a new event is increasing in the deeper levels.

4.3 Multiple Viewpoints

Some types of data may be divided into parallel representations. These representations might be useful to distinguish different aspects and assess some of the peculiar properties of the data stream. Treating the data in multiple representations can also be useful to predict the next symbol when one of the representations might be suitable for a particular sequence, whereas another representation might bring advantages in other situations.

Music is a good case of such analysis. The notes in a musical progression, in the simplest case, may be divided into pitches and durations. We can also work on more "advanced" relations such as scale degree, position in the measure, fermata. In terms of predicting next event, some of these representations might outperform others under different conditions. As an example, scale degree would be very useful if all the musical context is in the same key, however melodic interval might prove more suitable if the predictions are required in a transposed key.

The representation of data in multiple ways was first coined by Conklin and Witten [31] and later developed by Pearce, Conklin and Wiggins [76] as the so-called "multiple viewpoints modeling" (MVM). In this technique each of the representations are called a "viewpoint." A collection of viewpoints form the "multiple viewpoint system." In the system, the next sequence is predicted based on the information incorporated from these viewpoints.

The viewpoints can be divided into 3 types:

- Basic types: The simplest viewpoint that do not depend on other viewpoints. E.g. MIDI note numbers, duration of the notes
- Derived type: The viewpoints that are derived from basic types. E.g. Melodic interval, pitch-class distribution
- Cross type: Viewpoints that are constructed by taking the cross-product of two or more types. The tuples forming the parallel viewpoints are mapped into unique tokens to obtain this viewpoint. E.g. Notes $\otimes$ Durations; in this cross type a quarter C, a quarter D, a eighth C will all be mapped to different symbols.

Multiple viewpoints modeling can be seen as a general modeling scheme: it can be used to model anything that can be expressed in multiple representations such as music, finance or reactions of artificial intelligence in computer games. On the other hand, the power (or the weaknesses) comes from the choice of viewpoints picked to describe the phenomena. Apart from obvious differences in the viewpoints for completely different situations (representations of finance and computer games would of course be totally different), subproblems in a phenomenon might also require dissimilar viewpoints. As an example, the position of a stroke in a rhythmic cycle of tabla music would be helpful to predict the next stroke [25]; on the contrary it wouldn't make sense to use this viewpoint to model the *usulsüz* (non-metered) sections in an *uzun hava* as there are no cycles in these sections. Evidently, we have to be careful about the viewpoints defined to gather computationally meaningful data.

Another important point is, even though actual events might be significantly different for two phenomena, their representation might be quite similar. The notes in Bach's chorales can be easily modeled in a viewpoint showing the associated MIDI number with the note [32]. Even though frequencies of the notes in Turkish folk music and Baroque music do not match, Turkish folk music notes can indeed be modeled as floating point MIDI numbers (Section 5.2) [3]. From a symbolic point of view, they might seem very similar, however this does not mean the results would be the same (Section 6): familiar ears wouldn't want to hear Turkish music in Western classical tuning, and vice versa.

4.4 Long-term and Short-term Modeling

A common limitation of training the predictive models over large amount of data is the model is rendered too general to effectively predict patterns specific to the current song: If the song has a peculiar recurring phrase, but this phrase is not seen very frequently in the training database, the patterns generated might be totally irrelevant, even though they are supposed to match the context. Therefore, to obtain predictions which are trained over a particular style and also sounds like a specific song, a long-term-model (LTM) and a short-term-model (STM) may be constructed [32]. The long-term-model is built on the entire training set and a short-term-model (STM), which is trained on the current song that is being evaluated. Only symbols up to the current time are used in the STM; looking ahead is not permitted when making a prediction.

When a prediction is to be made at a given time-step, the LTM and STM are

[3]The notes in Baroque music does not necessarily have to be tuned according to $A = 440Hz$. In this sense, they also show a deviation from the frequency values mapped to MIDI numbers.

combined and normalized to a single predictive distribution for each of the viewpoints [32]. The probability of the next symbol for a possible state in the weighted distribution is obtained by:

$$P_{merged}(w_{t+1} = s_{t+1}) = \frac{\sum_{m \in M} \omega_m P_m(w_{t+1} = s_{t+1})}{\sum_{m \in M} \omega_m} \tag{12}$$

where ω_m is the weight associated with the model, and M is the set of the models in the system. In this case, M corresponds to $\{LTM, STM\}$. Weight associated with the model is related to the, H_m, the entropy of the probability distribution of the model. H_m is defined as:

$$H_m \triangleq -\sum_{k=1}^{N} P_m(w_{t+1} = s_k) \log_2 \Big(P_m(w_{t+1} = s_k) \Big) \tag{13}$$

where s_k is an element in the symbol set, $S = \{s_1, s_2, \ldots, s_k, \ldots, s_N\}$, and N is the total number of the symbols. As the probability distribution gets biased over fewer outcomes, the entropy will decrease. Therefore, lower entropies implies higher predictability. Entropy of a distribution is constrained by an upper bound, H_{m-max}. The possible maximum entropy occurs when the probability is uniformly distributed, implying that uniform probability distribution possesses the no predictive power. Substituting $P_m(w_{t+1} = s_k) = 1/N$ to Equation 13, H_{m-max} is found as:

$$H_{m-max} = \log_2(N) \tag{14}$$

Finally, weight associated with the model is defined as:

$$\omega_m \triangleq \frac{H_{m-max}}{H_m} \tag{15}$$

CHAPTER V

EXPERIMENT AND RESULTS

5.1 Hypothesis

I hypothesize that multiple viewpoint modeling (MVM), which have been shown to be effective for computational modeling of musical voices in Western music [30, 31, 33, 34, 76, 77], can be effectively adapted to predict Turkish folk music. The adaptation will largely be concerned with finding appropriate representations that address key relationships in the *uzun hava*s, a structured and non-metered musical form in Turkish folk music. We believe that the experiments will show that MVMs are a flexible computational modeling tool that can be applied to various musics through the use of appropriate representations. To verify the hypothesis, comparative experiments will be carried between the viewpoints previously defined for Western melodies [31] and the novel viewpoints defined for Turkish folk music. The viewpoints will be trained on the transcriptions of the improvised melodies in *uzun hava*s given in extended Western staff notation. For the evaluation of the system, a quantitative entopy-based scheme [63] will be performed at the song level and through all experiments. I believe the results will bring relatively low entropies that would show that MVMs can effectively model *uzun hava*s. Moreover, by comparing the entropies obtained from the viewpoints previously defined for Western melodies and the novel viewpoints presented in the thesis, I hope to show that taking appropriate representations into consideration might be a key factor to successfully model a musical style.

5.2 Experimental Setup

From the *Uzun Hava Humdrum database*, the *uzun hava*s in Hüseyni and Hicaz *makam*s, the two *makam*s with the most songs are chosen. From the database, 5 songs in the Hüseyni *makam* and one song in the Hicaz *makam* are taken out, because they are either not played "at their locations" or they make a transition (*geçki*) to another *makam* [1]. Although Uşşak *makam* is represented nearly as much as Hicaz, it is not included because of its closeness to the Hüseyni *makam*, and it might drastically worsen the predictions of LTM. Moreover, other *makam*s in the database are also disregarded because they are represented by very few songs, and again LTM would not be able to give satisfactory results. In brief, the experiment is carried on a set of 49 *uzun hava*s of which 35 are in Hüseyni and 14 are in Hicaz *makam* respectively (Table 7). The total number of musical events (i.e. notes and rests) in the experiment is 7538.

For the experiments, 15 viewpoints were defined. The 8 viewpoints without the *Cents-Deviation* are chosen from a subset of the viewpoints used by Conklin and Witten [31] so that parallel observations can be made. The remaining 7 viewpoints incorporating the *Cent-Deviations* are the novel contributions of the thesis which are aimed at addressing the key relationships in the *uzun hava*s. From the set of viewpoints provided by Conklin and Witten, the viewpoints that incorporate the position of the note in a cycle are taken out, since they would be problematic in the *usulsüz* sections of *uzun hava*s. Moreover, *Fermata*, *Time-Signature*, *Key-Signature* and related viewpoints are left out from the experiments due to time constraints, and they will be included in the future work (Section 7). The viewpoints are:

- Durations (D): A basic viewpoint indicating the duration of the note relative to a whole note. The duration of a whole note is defined as one and shortest

[1] Even though it wouldn't matter in terms of computational modeling, two of the pieces were actually taken out due to their politically incorrect lyrics.

duration is be zero, which is the duration of a grace note.

- Notes (N): A basic viewpoint indicating the pitch of the note. This is simply the MIDI number of the note.

- Cents-Deviation (CD): A basic viewpoint indicating the deviation of a note from semi-tone in cents. The value of the viewpoint taken as either 0 or 50 cents. This viewpoint is not used by itself but to form the novel *pitch-related* viewpoints [2], since decoupling it from pitch might disrupt the *makam* structures (explained later in section 6 in more detail). Viewpoints with *Cents-Deviation* constitutes the context-specific aspects of the *uzun hava* modeling.

- Notes with Cents-Deviation (NwCD): A viewpoint indicating the pitch of the note with quarter tones added to the scale. This viewpoint can be interpreted as the floating MIDI number of the note.

- Contour (C): A derived viewpoint showing whether the current note is ascending, descending or stationary with respect to the previous note. It can take the values of {-1, 0, 1, null}.

- Melodic-Interval (MI): A derived viewpoint marking the relative change in pitch with respect to the previous note. This viewpoint can take any positive or negative integer within the MIDI range.

- Melodic-Interval with Cents-Deviation (MIwCD): A viewpoint specifying the relative change in pitch with respect to the previous note with quarter tone

[2]From a mathematical point of view, this operation can be interpreted as the crossing the 12-tone scale (*Notes* viewpoint) and *Cents deviation* to bring the 17-tone scale in Turkish folk music. However, as emphasized in Chapter 2.1 treating *makam* theory, as any kind of extension from Western music theory would be an utter mistake! Therefore the *pitch related* viewpoints without cent deviations are an incomplete (and unfortunate) method of describing traditional Turkish music. In fact, those viewpoints are only added for comparison of the results when the target symbols are increased from the Western constraints to encompass *makam*s (Section 5.4). Under the light of this discussion, the viewpoints constructed from *Cents-Deviation* viewpoint will not be addressed as cross types from pitch and cents informations throughout the thesis.

precision. This viewpoint can take any positive or negative floating number within the MIDI range.

- Scale-Degree (SD): A derived viewpoint denoting the relation of the note with respect to the *karar* (ending) tone of the *makam*. The notes are wrapped into one interval such that the viewpoint can take an integer value between 1 (for *karar tone*) and 12. Quarter tones are ignored, i.e. $B\flat^2$'s is treated as $B\flat$.

- Scale-Degree with Cents-Deviation (SDwCD): A viewpoint comprising the relation of the note with respect to the *karar* (ending) tone of the *makam* with quarter tones included.

- Durations $\otimes$ Notes (D $\otimes$ N): A cross viewpoint incorporating the duration and the MIDI note.

- Durations $\otimes$ Notes with Cents-Deviation (D $\otimes$ NwCD): A cross viewpoint joining the duration and the floating MIDI note.

- Durations $\otimes$ Melodic-Interval (D $\otimes$ MI): A cross viewpoint linking the duration and the melodic interval.

- Durations $\otimes$ Melodic-Interval with Cents-Deviation (D $\otimes$ MIwCD): A cross viewpoint combining the duration, the melodic interval and the cents deviation.

- Durations $\otimes$ Scale-Degree (D $\otimes$ SD): A cross viewpoint putting the duration and the scale degree together.

- Durations $\otimes$ Scale-Degree with Cents-Deviation (D $\otimes$ SDwCD): A cross viewpoint incorporating the duration, the scale degree and the cents deviation.

The *Durations*, *Notes*, *Cents-Deviation* viewpoints are fetched from the songs **kern format by calling regular expressions via a bash script. Then, using MATLAB, all of the viewpoints except the cross viewpoints are extracted from these viewpoints,

Table 4: Basic and derived viewpoints corresponding to the events in the last two measures of U0368. "N/A" indicates situations where obtaining a value for the viewpoint is not applicable and "-" indicates the value of the viewpoint is null.

	1	2	3	4	5	6	7	8	9	10	11	12	13
D	.250	.125	.125	.125	.125	.125	.250	.125	.125	.125	.250	.125	.250
N	70	74	72	74	70	74	72	-	76	74	76	74	74
CD	.5	0	0	0	.5	0	0	-	0	0	0	0	0
NwCD	70.5	74	72	74	70.5	74	72	-	76	74	76	74	74
C	N/A	1	-1	1	-1	1	-1	-	-	-1	1	-1	0
MI	N/A	4	-2	2	-4	4	2	-	-	-2	2	-2	0
MIwCD	N/A	3.5	-2	2	-3.5	3.5	2	-	-	-2	2	-2	0
SD	2	6	4	6	2	6	4	-	8	6	8	6	6
SDwCD	2.5	6	4	6	2.5	6	4	-	8	6	8	6	6

and the symbols in the viewpoints are mapped to unique floating numbers. Going back to the example in Figure 6, Table 4 shows the non-cross type viewpoints that would be obtained from this sequence.

The modeling and evaluation (Chapter 4, Section 5.3) part of the framework was implemented in C++ [21]. Aside from being a computational model, the algorithm is also aimed to be used for generative music, and therefore it is compiled as as an external object in Max/MSP. The framework consists of the Max/MSP external along with some supporting patches. Currently, the framework accepts two viewpoints in a single run. For convenience, the viewpoints of all songs are combined into text files and fed into Max/MSP. Each text file holds two columns of floating numbers which are the corresponding viewpoints. The framework can be compiled either to treat these viewpoints separately or to internally form a cross viewpoint from them. The end of each song is marked with a special character in order to reset the stream in the end of each song during training.

5.3 Evaluation

For evaluation of the prediction system, leave-one-out cross-validation was performed on the subset picked from the *Uzun Hava Humdrum database* explained in Section 5.2. During the experiment, each song is picked as the testing data, and LTM is trained over the other songs. STM is built while the testing data is fed to the system. At each time step t, the true symbol is noted. Then the predictions carried in the previous step $t-1$ are checked and $P(\omega_t = s'_t)$, the probability of the true symbol, s', at t is recorded.

From the probabilities, average cross-entropy [63] is calculated at the song level and through all experiments. Cross-entropy is a common domain-independent approach used for evaluating the quality of model predictions. Assume that we have observed a sequence of length N. Average cross-entropy is defined as:

$$H_c = -\frac{1}{N}\sum_{i=1}^{N}\log_2\Big(p_T(e_i|c_i)\Big) \tag{16}$$

where $p_T(e_i|c_i)$ is the probability of e_i given the context c_i with respect to the probabilistic predictive theory T [88]. Rewriting the equation to match our system:

$$H_c = -\frac{1}{N}\sum_{t=1}^{N}\log_2\Big(P_s(\omega_t = s'_t|c_t)\Big) \tag{17}$$

where, $P_s(\omega_t = s'_t|c_t)$, the "smoothed" probability of the true symbol, given by the computational modeling algorithm, at each time step t, and c_t refers to conditions given by the preceding states in the VLMM (Equation 8). Assuredly, escape probabilities (Equation 10) are also considered in the calculations.

Upon inspecting this equation, it can be seen that the higher the probability of the true symbol is, the lower the cross-entropy will get. This behavior allows us to interpret the average cross-entropy as a way to evaluate the confidence of the system. Also, notice that if the probability of any event in the sequence is predicted as 0, the average cross-entropy will diverge to $+\infty$. Nevertheless, escape probabilities at each level of the PST (section 4.2.4) ensure such an occurrence would never happen.

In a predictive system, average cross-entropy is a more reliable criteria than the prediction accuracy of the true symbol. While calculating the symbol recognition rate, a wrong but likely outcome will be treated as bad as an unlikely choice. On the other hand, average cross-entropy will distinguish a confident prediction from an unsure one by penalizing the former less. Therefore average cross-entropy is preferable to prediction accuracy, especially in applications, which can be used in generative processes, where we would not want the exact replica of the music but alternations of it. The Max/MSP framework can predict the next note by picking either the most-likely or a random symbol from the probability distribution of the next symbol space. The most likely prediction and the true symbol are also recorded during the experiment, and the symbol recognition accuracy is provided in the Section 5.4.

During the experiment, the Max external outputs instantaneous cross entropies (the $\log_2$ values of the probabilities) of the true symbol at each prediction step. Later, the values are averaged in MATLAB to obtain average cross-entropy and they are also converted to perplexities. Perplexity is a measure of the number of choices that the model has picked among the true symbol [63]. Average perplexity is defined as $P = 2^{H_c}$. Average perplexity is found for each validation (i.e. for each song) and also for the whole experiment. In addition to average perplexities, median perplexities are recorded. The prior probabilities of the symbols are used to obtain a baseline for evaluating perplexity results. In other words, average perplexity of the 0^{th} order model LTM is used as the baseline.

5.4 Results

In the experiments, classification accuracies, average and median perplexities over the whole dataset and in the song-level are recorded for STM, LTM and combination of LTM and STM with different maximum orders. For all results below, the term "significant" has the following meaning: the claim is statistically significant at the 0.01

level as determined by a multiple comparison test using the Tukey-Cramer statistic [3].

The accuracies of most-likelihood classification for a maximum order of 14 is given at Table 5. The table shows that the system has a low classification accuracy, and apart from the recognition of *Duration* viewpoints, there is hardly an useful increase in the classification accuracy with respect to the *a-priori* classification. Though, generally STM gives slightly worse results, it is not possible to say whether LTM, STM or the combination of the two models consistently outperforms others. Moreover, in some cases, the baseline (classification by prior information) surpasses the model. Also notice the classification accuracy greatly decreases in the *Contour* viewpoint.

Nonetheless, as explained in Section 5.3, classification accuracy is not as dependable as entropy based evaluation methods. Therefore, average and median perplexities for the predictions of the multiple viewpoints given by the LTM, STM and combined model for a number of different maximum VLMM orders are calculated. Figure 9 shows that for the *Durations* viewpoint, the average perplexity decreases almost monotonically with increasing order. This trend is true for all of the viewpoints. This result was expected since increasing the order would allow us to locate more context-specific patterns, and the system would be more confident with it's predictions. STM gives the lowest perplexities in every order. It is also seen that there is only a slight change in perplexity after order 14, therefore checking back more than 14 durations is not necessary. Note that there is not a significant decrease in the cross-entropies after a maximum order of 5.

Table 6 shows the average and median perplexities for a maximum order of 14. It can be easily seen that the average and median perplexities greatly decreases with respect to the baseline. Therefore, although the system typically fails to give an exact

[3] The experimental data and the complete set of results are available at http://sertansenturk.com/uploads/publications/senturk2011Improv/.

Table 5: Classification accuracies in percentage for the multiple viewpoints using a maximum order of 14. The first row in each cross type reports the classification accuracy of the unique tokens obtained by the cross product of the two viewpoints. The second and the third rows report the classification accuracy of the first and the second viewpoints forming the cross type.

	Priors	**LTM**	**Com.**	**STM**
D	35.45	70.96	69.09	61.46
N	24.22	26.57	27.66	26.94
NwCD	24.22	23.27	27.09	26.36
C	38.58	18.19	21.32	25.21
SD	24.64	22.92	23.27	21.60
SDwCD	24.09	26.70	27.31	26.64
MI	24.64	26.84	27.28	26.61
MIwCD	24.08	23.16	23.18	21.19
D $\otimes$ **N**	8.42	14.91	14.01	12.19
	35.40	64.23	59.72	54.60
	24.22	25.51	25.78	24.24
D $\otimes$ **NwCD**	8.53	15.60	14.39	12.48
	35.45	63.64	59.75	54.62
	24.22	25.84	25.18	23.59
D $\otimes$ **SD**	13.16	13.16	13.20	11.78
	26.43	64.83	59.27	53.01
	24.09	21.72	21.94	21.29
D $\otimes$ **SDwCD**	15.10	15.10	14.08	12.23
	35.45	64.66	60.21	55.01
	24.64	26.65	25.44	24.05
D $\otimes$ **MI**	8.53	15.79	14.49	12.48
	35.45	65.40	60.27	55.00
	24.64	26.35	25.33	24.08
D $\otimes$ **MIwCD**	8.15	12.88	12.19	10.71
	26.43	63.33	57.83	51.76
	24.08	21.70	20.88	20.07

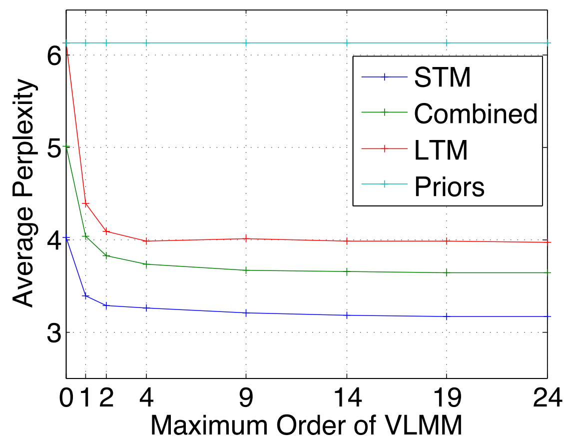


Figure 9: Average perplexities for duration prediction using LTM, STM and combined models for orders 0-25

match to the notes in the melodic progressions, it is giving more confident predictions.

When comparing the LTM, STM and their combination, although LTM is a noticable improvement over the baseline, STM always delivers the most confident results. Even combining the LTM and the STM is not as effective as STM. LTM significantly outperforms the Combined model, while STM significantly outperforms both. Comparing results of STM and the baseline, it can easily be seen that there is a remarkable decrease in the average and median perplexities. The power of STM is even more obvious in the cross types, where the average perplexities of the baseline and LTM are enormous with respect to the average perplexities of STM and, average perplexities of STM are the approximately half of the average perplexities given by the combined model! This means that STM may reduce the number of symbols to choose the true symbol as much as to the half.

Table 6 also shows that there is typically not much of difference between the average perplexities between *pitch related viewpoints* without *Cents-Deviation* and *pitch-related* viewpoints with *Cents-Deviation* incorporated. This means there is not a significant penalty if the quarter tones seen in Turkish folk music are added up to the possible symbols of the 12-note target space of Western classical music (implications are explained in Section 6.).

Table 6: Average and median perplexities for the multiple viewpoints using a maximum order of 14

	Priors		**LTM**		**Combined**		**STM**	
	Av.	Med.	Av.	Med.	Av.	Med.	Av.	Med.
D	6.14	3.00	3.99	2.30	3.65	2.01	3.18	2.00
N	10.61	5.18	5.00	3.51	4.12	2.84	3.87	2.60
NwCD	11.51	5.18	5.12	3.54	4.16	2.84	3.90	2.62
C	2.95	2.75	2.57	2.04	2.56	1.85	2.38	1.81
SD	8.36	5.00	4.59	3.40	3.94	2.78	3.75	2.54
SDwCD	9.10	5.00	4.69	3.43	4.01	2.78	3.77	2.55
MI	8.94	6.00	5.14	3.54	4.54	2.96	4.13	2.93
MIwCD	14.19	7.50	5.86	3.79	5.01	3.18	4.57	3.10
D $\otimes$ **N**	61.10	12.50	29.55	9.98	15.07	9.98	7.61	6.00
D $\otimes$ **NwCD**	65.60	12.50	30.85	19.59	15.45	10.1	7.64	6.00
D $\otimes$ **SD**	50.23	12.50	26.13	9.65	14.26	9.65	7.62	6.00
D $\otimes$ **SDwCD**	54.87	12.50	27.73	9.81	14.61	9.81	7.64	6.00
D $\otimes$ **MI**	49.92	12.00	26.75	10.70	16.31	10.70	7.46	5.90
D $\otimes$ **MIwCD**	75.58	12.81	35.69	13.38	20.39	13.38	7.53	5.63

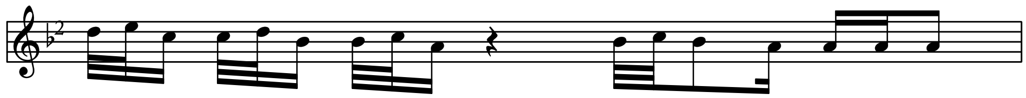
When the average perplexities are checked song by song (Table 7), it was observed that some songs have exceptionally high perplexities. Upon examining the songs, it was seen that the reason for the relatively high average perplexities seen in these songs is mostly due to the uncommon durations such as dotted notes (especially the double dotted notes), triplets, 64^{th} notes and gruppettos. When other *pitch related* viewpoints were inspected song by song, no critical problems were observed: usually if a song had a high perplexity, it was due to the length of the song being too short so that even a single ripple in the perplexity affected the average substantially. Also, by comparing the average perplexities of from *Durations* $\otimes$ *Scale-Degree-with-Cents-Deviation* and *Durations* $\otimes$ *Melodic-Interval-with-Cents-Deviation* at Table 7, it can be observed that one viewpoint can be favorable to the other for different patterns.

As stated in Section 5.3, the Max/MSP framework have the limited generative capability of predicting next symbol. The most-likely predictions at each step were checked empirically to observe whether the results have some validity with the source material. Figure 10 shows the ending of U0057, the predicted phrases by *Durations* $\otimes$

Table 7: Average perplexities given by evaluating *Durations* ⊗ *Scale-Degree-with-Cents-Deviation* and *Durations* ⊗ *Melodic-Interval-with-Cents-Deviation* viewpoints for each song in the experiment using a VLMM of maximum order of 14.

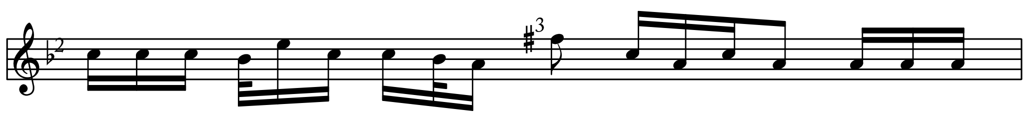
			D ⊗ SDwCD				**D ⊗ MIwCD**			
#	**Song**	**Makam**	**Priors**	**LTM**	**Com.**	**STM**	**Priors**	**LTM**	**Com.**	**STM**
1	**U0002**	Hicaz	37.35	17.87	9.93	7.38	65.86	27.76	15.81	9.70
2	**U0020**	Hicaz	34.28	16.95	9.19	5.79	56.50	30.59	15.23	5.78
3	**U0023b**	Hicaz	41.22	22.57	10.63	6.08	59.51	32.85	16.26	6.22
4	**U0031**	Hüseyni	73.74	36.22	27.11	7.27	92.11	33.07	30.50	7.56
5	**U0037**	Hüseyni	45.07	24.71	13.13	5.36	79.07	40.99	19.99	4.01
6	**U0049**	Hicaz	46.63	26.17	6.17	3.07	81.91	41.50	8.42	2.95
7	**U0049a**	Hicaz	39.68	18.08	12.67	9.04	69.56	29.94	20.38	9.98
8	**U0051**	Hüseyni	31.36	18.78	16.79	7.18	47.76	29.35	25.47	7.53
9	**U0057**	Hüseyni	51.98	24.52	17.55	8.62	61.59	35.50	26.01	8.85
10	**U0072**	Hicaz	84.73	32.24	22.70	12.10	122.49	47.26	36.33	11.06
11	**U0080**	Hüseyni	56.59	25.00	12.04	6.55	65.39	33.21	17.77	6.84
12	**U0120**	Hüseyni	33.21	18.09	14.71	9.99	64.79	34.13	30.16	10.15
13	**U0143**	Hüseyni	40.17	17.88	12.70	6.40	44.73	26.41	19.94	5.46
14	**U0181**	Hüseyni	38.41	15.89	10.92	6.29	43.09	28.55	13.23	5.51
15	**U0184**	Hicaz	67.54	39.11	8.90	5.56	83.33	48.34	11.10	5.73
16	**U0208**	Hüseyni	120.25	86.20	27.08	10.84	62.15	33.42	26.51	10.81
17	**U0218**	Hüseyni	54.95	27.26	21.98	10.59	54.39	31.38	26.35	7.45
18	**U0272**	Hüseyni	48.19	37.06	31.92	12.71	80.69	52.57	49.50	8.50
19	**U0285**	Hicaz	69.23	45.48	15.11	6.78	108.00	69.10	20.09	6.75
20	**U0333**	Hüseyni	57.04	34.11	20.24	9.85	106.12	61.89	38.95	9.35
21	**U0396**	Hüseyni	74.43	64.88	96.12	2.55	101.41	94.95	91.13	3.20
22	**U0410**	Hüseyni	43.74	15.90	11.73	6.00	49.04	18.72	13.37	5.51
23	**U0418**	Hüseyni	42.52	30.29	15.99	6.40	65.18	49.96	27.45	5.95
24	**U0460a**	Hüseyni	32.61	18.44	7.19	5.44	57.42	28.33	9.76	6.20
25	**U0485**	Hicaz	154.02	126.54	20.48	11.38	138.19	87.89	29.91	11.60
26	**U0561**	Hüseyni	41.43	16.99	17.67	7.20	85.64	32.21	33.69	6.61
27	**U0573**	Hüseyni	30.36	18.86	9.96	5.45	57.04	32.22	16.81	5.84
28	**U0605**	Hüseyni	50.37	26.84	25.11	9.65	68.82	36.07	31.87	8.73
29	**U0611**	Hüseyni	56.06	24.15	21.16	9.95	82.88	28.21	26.87	10.75
30	**U0624**	Hüseyni	131.09	67.91	30.55	9.37	152.43	73.69	46.60	8.22
31	**U0628**	Hüseyni	107.38	43.30	19.80	11.14	136.54	42.46	25.80	10.56
32	**U0647**	Hicaz	43.62	13.44	7.49	5.49	63.95	18.18	9.64	5.65
33	**U0648**	Hüseyni	51.54	27.19	36.84	9.55	67.20	34.98	56.07	6.45
34	**U0668**	Hüseyni	43.16	18.76	19.53	7.35	66.34	27.42	35.80	6.49
35	**U0670**	Hüseyni	41.95	18.49	20.11	5.17	43.04	18.96	18.52	4.98
36	**U0697**	Hicaz	39.92	18.85	11.27	6.87	57.87	22.20	18.81	8.29
37	**U0706**	Hüseyni	62.30	20.71	12.71	10.28	60.01	24.86	15.66	9.06
38	**U0711**	Hicaz	48.42	25.32	17.06	12.57	71.36	33.39	23.27	11.22
39	**U0718**	Hüseyni	43.73	29.01	13.44	5.13	58.87	36.23	17.74	6.20
40	**U0723**	Hüseyni	96.44	71.32	41.15	7.31	129.70	92.44	49.22	6.89
41	**U0724**	Hüseyni	92.10	59.16	28.11	11.71	153.45	65.26	39.10	11.44
42	**U0730**	Hicaz	57.11	26.77	8.35	4.44	66.79	26.60	9.36	4.08
43	**U0741**	Hüseyni	39.04	17.25	9.84	7.60	81.23	29.05	15.01	9.21
44	**U0745**	Hüseyni	61.56	15.81	10.15	7.20	37.23	21.75	14.20	4.30
45	**U2002**	Hüseyni	33.25	17.99	18.71	5.42	49.24	28.32	24.88	6.80
46	**U2007**	Hicaz	61.06	51.95	13.15	3.86	85.14	61.28	15.84	4.00
47	**U2008**	Hüseyni	43.05	20.87	14.19	4.86	54.87	20.00	11.09	4.27
48	**U2009**	Hüseyni	45.49	12.66	8.90	3.99	67.74	16.26	9.47	4.70
49	**U2010**	Hüseyni	32.82	13.45	10.84	4.06	50.04	18.08	15.92	3.45

Scale-Degree-with-Cents-Deviation, and the cross-entropy profile of each prediction. At a glance, it can be seen that all the predictions lie within the key signatures of *Hüseyni*, the *makam* of U0057. The only exception is observed when the LTM is asked to predict the next note in the place of the rest at the 10^{th} step (Figure 10b). There, LTM predicts $B\flat$, which is in the key signature of *Hicaz*. On the other hand, STM and Combined models both predict $F\sharp^3$ (Figure 10c, Figure 10d), where the true symbol emits a slightly less instantaneous cross-entropy (Figure 10e). Figure 10a shows that the start of the subsequence is highly structured in terms of both durations and the pitches. However, all of the models fail to capture this descending path adequately. Nonetheless, transcribed melody and all of the predictions except the predictions at step 10 consistently stay inside the Hüseyni pentachord at its location (Figure 2f). Moreover all melodies converge to Dügah, the *karar* (ending) tone of *Hüseyni makam*, and the entropy profile (Figure 10e) shows that the models are relatively confident in this prediction. Another interesting point is in terms of average perplexities Combined model (24.41) gives better results in this short pattern compared to LTM (28.25) and STM (28.32). Notice that these average perplexities are much higher than the average perplexities of the song (Table 7).



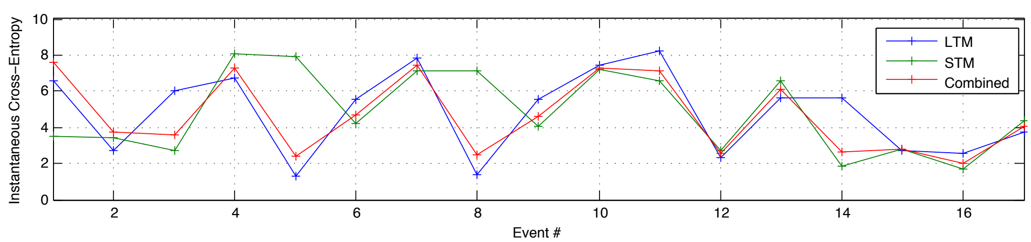
(a) Ending of U0057

(b) Prediction of **D⊗SDwCD** using LTM



(c) Prediction of **D⊗SDwCD** using STM

(d) Prediction of **D⊗SDwCD** using Combined Model



(e) Instantaneous cross-entropies of **D⊗SDwCD** for each model emitted by the true symbol at each prediction step

Figure 10: Ending of U0057, predicted patterns by using *Durations* ⊗ *Scale-Degree-with-Cents-Deviation* viewpoint and the instantaneous cross-entropies of the true symbols at each model.

CHAPTER VI

DISCUSSIONS

Even though the classification rates are very low (Table 5), the average perplexity results (Table 6) show the computational modeling is confident of its predictions: Compared to the baseline, the system is able to pick the true symbol among significantly fewer symbols using the STM with 14 order VLMM. Since our aim in the system is a predictive model instead of a classification model, on the basis of the average perplexity values, we can argue the computational modeling has been successful in modeling the *uzun hava*s. Therefore, it can be argued that multiple viewpoints modeling, which has been shown to be effective for computational modeling of Western music [30, 31, 33, 34, 76, 77], can be effectively adapted to predict Turkish folk music.

Moreover, STM significantly outperforms LTM, and the combination of both of the models. The success of STM indicates that the songs typically have strong patterns. These patterns peculiar to each song are either not observed in no to few songs, therefore LTM cannot effectively track them. Since *seyir*s are an integral part in the explanation of *makam*s, finding context-specific patterns might be easier if a medium term model (MTM) is introduced to the system. A MTM would have parallel PSTs, each of which are only trained on a single *makam*. It should also be noted that this finding is in parallel with the results in the previous research on tabla sequences [21, 24, 25].

In order to model the melodies in traditional Turkish music, the selection of multiple viewpoints might be crucial. For example, the *Cents-Deviation* information cannot be used without being integrated to the *pitch-related* viewpoints such as *Notes* or *Scale-Degree*. For a generative system, decoupling them might still give good average perplexities. However, when the pitch and the cents deviation are predicted independently from each other, the results might introduce notes with wrong accidentals. These erroneous pitches would disrupt the melodic intervals and the *makam* structure.

One of the most prominent observations is that extending the possible set of pitches from Western music to Turkish music results in a slight, insignificant increase in perplexity values. At first, this finding may be misinterpreted as incorporating *Cents-Deviations* is meaningless. On the contrary, one should not fail to note that when the quarter tones are included, i.e. the symbol indicating both the quarter tone and the neighboring tone is decoupled to create two unique symbols. For a given song, the occurrence of one of the symbols is usually very close to zero, and almost all of the counts are accumulating on the other note since the transcriptions strictly obey the key signature of their *makam*s. By inspecting the most-likely predictions, it is seen that the predictions typically stay at the key signatures or the temporary accidentals of the *makam*. Moreover, the instantaneous perplexities show that the confidence of the system does not differ when it is asked the predict a quarter-tone or a semi-tone (Figures 10a, Figure 10e). Therefore, the multiple viewpoint system is able to model the context-specific pitches in *makam*s and distinguish them from the neighboring tones present in Western music virtually without any penalty. If the system's symbolic output is sonified, the consequences will be much clearer: music generated in the 12-tone scale of Western classical theory is expected to sound much different and less "Turkish" than the 17-tone scale of *makam* theory. As an example, think of a sequence generated from a song trained on Uşşak makam: predictions from *pitch-related* viewpoints without *Cents-Deviation* will have $B\flat$'s instead of $B\flat^2$. As a result, the generated melodies will probably not sound like Uşşak; they might sound more like the modern Phrygian mode on A.

On the other hand, the slight increase in perplexities brings a negative criticism. As explained in Section 2.1.1.1, the notes played in ascending and descending *seyir*s are typically different in practice. In detailed transcriptions, it would cause a scatter to the neighboring notes around quarter tones, and we would expect some significant increase in perplexities of the predictions. However, the *Uzun Hava Humdrum Database* does not typically show these deviations in ascending and descending *seyir*s. It is one of the reasons why 12-tone and 17-tone predictions give such close results. To detect these changes in *seyir*s, it is almost certain that the research should be extended into the audio domain.

As explained in Section 4.3, multiple viewpoints are a general way of modeling parallel representations of a sequence. Once the framework is set, it is relatively straightforward to use the concept in completely different problems. Yet, the power of the model comes from the the viewpoints picked to describe the sequence. As no-free-lunch theorem suggests, the viewpoints have to be decided after thorough considerations. In our experiments, *pitch related viewpoints with Cents-Deviation* somewhat fulfill the necessity of context-dependent descriptors, and as explained above, such viewpoints are as confident as the ones without *Cents-Deviation* while presenting us with a much more precise melody modeling.

Viewpoints based on absolute pitch might give poor results. As an example, if the training songs in Hüseyni *makam* are entirely played at their location, the system would not be able to predict a piece in Hüseyni transposed to a different *karar* note. However, since the experiment set is removed of such transposed pieces, this problem is not encountered. In the model, prediction of the next note in a song would be inclined to follow the branches in the PST which were trained on the same *makam*. As a result, the average perplexities given by the *Notes* and *Notes-with-Cents-Deviation* viewpoints brings very similar results to the other *pitch related* viewpoints.

As mentioned in Section 5.4, the system cannot properly predict some less frequent states of the *Durations* viewpoint such as the gruppettos, 64^{th} notes and dotted notes. In LTM, since these notes are rarely encountered, they typically do not possess high counts in any n-grams. As a result, the presence of these notes aggravates the predictions. On the other hand, STM, which is trained on the particular song, is affected less from this problem: in fact, the patterns formed by these note durations may even be as prominent as patterns formed by fourth notes, eighth notes and such. However, these durations also increase the number of possible states in the system; leaving STM less confident in its predictions. This fact can be easily seen from the increase in the prior average perplexities given in Table 7. Also, gruppettos and very fast notes such as 64^{th} notes may be interpreted as the embellishments and ornamentations transcribed from listening to a piece. As explained in Section 3.2, the transcriptions in the TRT database typically do not represent these musical elements adequately. Consequently, since these symbols occur very infrequently in the *Uzun Hava Humdrum Database*, the system finds it very hard to recognize these symbols. We can conclude that the *Uzun Hava Humdrum Database* is incapable of representing the ornamentations and embellishments in the original performances of the *uzun hava*s, and therefore the computational model is unable to predict these improvisational elements, which are inseparable from the *uzun hava* form.

Up to now, the results obtained from the symbolic notation have been discussed. However, the biggest potential criticism towards the thesis work is whether these results show any actual relevance to the *uzun hava* form. In Section 3.2, some of the dangers and drawbacks of using transcriptions and Western symbolic notation to represent non-Western music are given. Moreover, the notations provided by TRT are known to contain critical errors [97], and there is a noticeable difference between the transcription styles of the transcribers.

At the current stage of the thesis, it can be observed that the predictions may have some consistencies with the transcriptions (Figure 10). Moreover, the average perplexities emitted by the true symbols are relatively low. Nevertheless, as long as the input type stays in the notation format, it is not clear whether the system is adequately able to describe the actual music. On the other hand, in order to conduct systematic research in any topic, especially in ones where very little previous research is available, it makes more sense to keep the complexity as simple as possible rather than to dive into the problem blindly. Accordingly, the usage of symbolic data and transcriptions is a necessary, initial step to discover the hidden aspects of traditional Turkish music.

CHAPTER VII

FUTURE WORK

While, the current viewpoints with *Cents-Deviation* have been pretty useful in predicting the melodic sequences in *uzun hava*s, it is acknowledged that constructing more viewpoints and introducing new cross types might give us a better understanding of the *uzun hava* form. As an example, Table 6 shows that the average perplexities are reduced for the *Contour* viewpoint using STM, and this decrease is significant. Nonetheless, it might be more helpful to cross the *Contour* viewpoint with *Scale-Degree-with-Cents-Deviation* or *Melodic-Interval-with-Cents-Deviation* viewpoints to get a better view in *seyir*s. In addition to *Scale-Degree-with-Cents-Deviation* viewpoint, which shows the distance of the note with respect to the *karar* (ending) note, it might be useful to construct another viewpoint, which shows the distance of the note with respect to the *başlangıç* (starting) note. Moreover, adding viewpoints such as *Fermata* and *Time-Signature* [31] might bring more knowledge about *usul* and its effects on the melody. *Time-Signature* might be especially useful to predict and distinguish the melodic continuations in the *usullü* and *usulsüz* sections of *uzun hava*s. Parallel to the novel *pitch-related* viewpoints used in the thesis, *Time-Signature* and related viewpoints might be extended to address the distinct *usul*s having the same number of beats (Section 2.1.1.2).

In order to claim a stronger relevance between the model and the actual music, the research has to be extended in some ways. One crucial step is to include audio in the computational model along with the symbolic notation. First, we would ideally be able to learn more aspects of *uzun hava*s such as the embellishments by directly working on audio, Also, the relevance of the music and the symbolic notation may be evaluated by comparing the results from symbolic notations and audio. From the MIR point of view, this can be done by incorporating note segmentation and automatic transcription algorithms and setting up a model by using variable-length hidden Markov models (VLHMM). Automatic segmentation and transcription algorithms may also be beneficial to automatically gather more consistent and reliable transcriptions of *uzun hava*s. The VLHMM model is already coded in our previous research [25] by Avinash Sastry. However, an extensive implementation and integration of the automatic segmentation and transcription algorithms stand out as major challenges.

The next step would be to convert the setup into a practical generative system. By learning from both audio and symbolic notations, the generative system would be able to play or print improvisational ideas based on the computational modeling of *uzun hava*s. Then, Turkish folk music virtuosos and ethnomusicologists expert on Turkish folk music might be consulted to point out the "interesting" and the "failed" parts in the generated patterns. They may be asked to write out and play the patterns in their own style. Later, the original, generated and reinterpreted scores and audio recordings may be cross-compared. Moreover, cognitive experiments might be carried out in parallel to scientifically present the relevancy of the model according to the expectations of humans. I hope that the parallelism in the quantitative results between Conklin et al. and Pearce et al.'s research [30, 31, 33, 34, 76, 77] and this thesis, may be generalized to Pearce's findings in music perception and cognition [78].

Within such feedback, I believe there is a substantial room for the computational modeling to improve. Moreover, if the modeling is improved above a certain level, the model might be either used as a core component of an educational software, which might help beginner-to-intermediate students to learn how to play Turkish folk music or improvisation in general, and as a meta-musician/composer, which can improvise along with human musicians or provide them improvisational ideas on-the-fly in human-computer interactive performances. Such applications would open up new paths for musical expressivity and help spreading the ideal of "liberation of sound" [100].

Another interesting aspect of including audio recordings is investigating *tavırs*. To the best of my knowledge, there is a lack of extensive research on how the music of Turkey changes with respect to factors such as geographical regions, ethnic groups, languages and religions. Later, the research might be extended to include musical cultures from the neighbors of Anatolia such as Balkan, Armenian, Persian and Arabic, which share some musical connections with traditional Turkish music such as *makam*s, musical forms and sometimes even the songs. While a plain computational approach would lack the depth of social analysis, it might still indicate musical similarities in a geotagged and multi-cultural context.

CHAPTER VIII

CONCLUSION

Within the thesis, a symbolic database, named *Uzun Hava Humdrum Database*, is constructed from the transcriptions of *uzun hava*s in the TRT Turkish folk music archive with the collaboration of Prof. Erdal Tuğcular (Department of Music Education, Gazi University, Ankara, Turkey). The database is encoded in the Humdrum **kern format [52] and it encompasses 77 songs, 10849 notes in 8 *makam*s. To the best of knowledge, the database is the first symbolic database of *uzun hava*s in machine-readable format. The conceptual problems and the practical hardships of creating a symbolic database of a non-Western musical style is also presented along with the explanation of the database. We hope that the database will help to fill the lack of availability of examples from world musics for academic research purposes.

The second contribution of the thesis is the computational modeling of *uzun hava*s. The system is based on the multiple viewpoints modeling (MVM) framework developed at the Georgia Tech Center for Music Technology (GTCMT) [21]. A subset of pieces from *Uzun hava Humdrumdatabase* is picked to train the computational model. The novelty of the thesis lies within the viewpoints constructed to model the 17-tone scale of Turkish folk music. These viewpoints and viewpoints previously defined for Western music [31] are experimented on to predict the duration and pitch of the next note. The average and median perplexities show that the system is highly predictive. It also shows that the multiple viewpoints modeling, which has previously been applied to Western music [30, 31, 33, 34, 76, 77], may also be used to model *makam* music. The results also suggest that the transcriptions hold highly context-specific patterns that are not easy to catch in the long-term model (LTM). On the other hand, the melodic patterns in the *uzun hava*s are self-consistent since the short-term model (STM) outperforms LTM. To the best of my knowledge, the thesis brings the first attempt of modeling melodies and improvisations in traditional Turkish music, and it is the first usage of varable0length Markov models (VLMMs) and MVMs in the statistical analysis of traditional Turkish music.

Even though the current stage of the thesis requires more depth in the modeling scheme and different methodologies, as the first step of computational analysis of melodic structures in Turkish folk music, it is very promising. Future work, opened by this research, may help us better understand musical structures in Turkish folk music, and lead to practical applications that might be integrated in music education and performances.

Finally I would like to point out that there are next to no considerations of world musics. This lack of interest may be conceived as as maybe an inevitable occidental inclination in the music technology area. I believe this constrained point of view should be eliminated if the MIR community aims to work on music in general. Moreover, research on different musical styles might not only widen our perspective on musical creativity, but current (Eurocentric) MIR technologies might also benefit from the findings from other traditions. I hope this work will bring inspiration and motivation to both myself and other colleagues, who desire to understand musical phenomenon, pursue new horizons in musical interactions, and embrace human creativity in a multicultural context.

REFERENCES

[1] Akkoç, C., "Non-deterministic scales used in traditional Turkish music," *Journal of New Music Research*, vol. 31, no. 4, pp. 285–293, 2002.

[2] Alpkoçak, A. and Gedik, A., "Classification of Turkish songs according to makams by using n-grams," in *Proceedings of the 15. Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN)*, 2006.

[3] Ames, C., "The Markov process as a compositional model: a survey and tutorial," *Leonardo*, vol. 22, no. 2, pp. 175–187, 1989.

[4] Antonopoulos, I., Pikrakis, A., and Theodoridis, S., "Locating rhythmic patterns in music recordings using hidden Markov models," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 1, pp. I–221, IEEE, 2007.

[5] Arel, H., *Türk musikisi nazariyatı dersleri.* Hüsnütabiat Matbaası, 1968.

[6] Arom, S., "The use of play-back techniques in the study of oral polyphonies," *Ethnomusicology*, vol. 20, no. 3, pp. 483–519, 1976.

[7] Arom, S., Thom, M., Tuckett, B., and Boyd, R., *African polyphony and polyrhythm: musical structure and methodology.* Cambridge university press Cambridge, 1991.

[8] Assayag, G. and Dubnov, S., "Using factor oracles for machine improvisation," *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, vol. 8, no. 9, pp. 604–610, 2004.

[9] Assayag, G., Dubnov, S., and Delerue, O., "Guessing the composers mind: Applying universal prediction to musical style," in *Proceedings of the International Computer Music Conference*, pp. 496–499, Citeseer, 1999.

[10] Aucouturier, J., Pachet, F., and Sandler, M., "The way it sounds: Timbre models for analysis and retrieval of music signals," *Multimedia, IEEE Transactions on*, vol. 7, no. 6, pp. 1028–1035, 2005.

[11] Bainbridge, D. and Bell, T., "The challenge of optical music recognition," *Computers and the Humanities*, vol. 35, no. 2, pp. 95–121, 2001.

[12] Bartók, B., Herzog, G., and Lord, A., *Serbo-Croatian folk songs.* Columbia University Press, 1951.

[13] Bartók, B. and Suchoff, B., *Béla Bartók Essays.* Univ of Nebraska Pr, 1976.

[14] Berkowitz, A. and Ansari, D., "Generation of novel motor sequences: The neural correlates of musical improvisation," *NeuroImage*, vol. 41, no. 2, pp. 535–543, 2008.

[15] Betzwieser, T., "Orientalism, masquerade and Mozart's Turkish music," *Journal of the American Musicological Society*, vol. 57, no. 3, pp. 656–663, 2004.

[16] Bozkurt, B., Gedik, A., and Karaosmanoğlu, M., "TMVB: Klasik Türk müziği icra analiz çalışmaları için bir veri bankası," in *Proceedings from Türk Müziğinde Uygulama-Kuram Sorunları ve Çözümleri Çağrılı Kongre, Maçka, İstanbul*, 2008.

[17] Bozkurt, B., Gedik, A., and Karaosmanoglu, M., "Music information retrieval for Turkish music: problems, solutions and tools," in *Signal Processing and Communications Applications Conference, 2009. SIU 2009. IEEE 17th*, pp. 804–807, IEEE, 2009.

[18] Bozkurt, B., Yarman, O., Karaosmanoğlu, M., and Akkoç, C., "Weighing diverse theoretical models on Turkish maqam music against pitch measurements: A comparison of peaks automatically derived from frequency histograms with proposed scale tones," *Journal of New Music Research*, vol. 38, no. 1, pp. 45–70, 2009.

[19] Chai, W. and Vercoe, B., "Folk music classification using hidden Markov models," in *Proceedings of International Conference on Artificial Intelligence*, vol. 6, 2001.

[20] Chen, S. and Goodman, J., "An empirical study of smoothing techniques for language modeling," in *Proceedings of the 34th annual meeting on Association for Computational Linguistics*, pp. 310–318, Association for Computational Linguistics, 1996.

[21] Chordia, P., Albin, A., Sastry, A., and Mallikarjuna, T., "Multiple viewpoints modeling of tabla sequences," in *Proceedings of International Conference on Music Information Retrieval*, 2010.

[22] Chordia, P. and Rae, A., "Raag recognition using pitch-class and pitch-class dyad distributions," in *Proceedings of International Conference on Music Information Retrieval*, Citeseer, 2007.

[23] Chordia, P. and Rae, A., "Understanding emotion in raag: An empirical study of listener responses," *Computer Music Modeling and Retrieval. Sense of Sounds*, pp. 110–124, 2009.

[24] Chordia, P., Sastry, A., and Albin, A., "Evaluating multiple viewpoint models of tabla sequences," in *Proceedings of 3rd international workshop on Machine learning and music*, pp. 21–24, ACM, 2010.

[25] Chordia, P., Sastry, A., and Şentürk, S., "Predictive tabla modelling using variable-length markov and hidden markov models," *Journal of New Music Research*, vol. 40, no. 2, pp. 105–118, 2011.

[26] Chordia, P., *Automatic Transcription of Solo Tabla Music*. PhD thesis, Stanford University, 2005.

[27] Cihat, C., "Türk müziğinde makam üzerine bir inceleme," Master's thesis, Erciyes Üniversitesi, 1993.

[28] Cleary, J. and Teahan, W., "Experiments on the zero frequency problem," in *Proc. Data Compression Conference*, vol. 480, Citeseer, 1995.

[29] Conklin, D. and Anagnostopoulou, C., "Comparative pattern analysis of Cretan folk songs," in *Proceedings of 3rd international workshop on Machine learning and music*, pp. 33–36, ACM, 2010.

[30] Conklin, D. and Cleary, J. G., "Modeling and generating music using multiple viewpoints," in *Proceedings of the First Workshop on AI and Music*, (Menlo Park, CA), pp. 125–137, AAAI Press, 1988.

[31] Conklin, D. and Witten, I., "Multiple viewpoint systems for music prediction," *Journal of New Music Research*, vol. 24, no. 1, pp. 51–73, 1995.

[32] Conklin, D., "Prediction and entropy of music," Master's thesis, University of Calgary (Canada), 1990.

[33] Conklin, D., "Music generation from statistical models," in *Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, pp. 30–35, 2003.

[34] Conklin, D. and Anagnostopoulou, C., "Representation and discovery of multiple viewpoint patterns," in *Proceedings of the 2001 International Computer Music Conference*, pp. 479–485, International Computer Music Association, 2001.

[35] Cope, D., "Experiments in musical intelligence (EMI): non-linear linguistic-based composition," *Journal of New Music Research*, vol. 18, no. 1, pp. 117–139, 1989.

[36] Cope, D., "Computer modeling of musical intelligence in EMI," *Computer Music Journal*, vol. 16, no. 2, pp. 69–83, 1992.

[37] Downie, J., *Evaluating a simple approach to music information retrieval: Conceiving melodic n-grams as text*. PhD thesis, The University of Western Ontario, 1999.

[38] Dubnov, S., Assayag, G., and Cont, A., "Audio oracle: A new algorithm for fast learning of audio structures," in *Proceedings of International Computer Music Conference (ICMC)*, pp. 224–228, 2007.

[39] Dubnov, S., Assayag, G., Lartillot, O., and Bejerano, G., "Using machine-learning methods for musical style modeling," *Computer*, pp. 73–80, 2003.

[40] Dutta, A. E., *Tabla: Lessons and Practice*. Ali Akbar College, 1995.

[41] Ebcioğlu, K., "An expert system for harmonizing four-part chorales," *Computer Music Journal*, vol. 12, no. 3, pp. 43–51, 1988.

[42] Eerola, T., Järvinen, T., Louhivuori, J., and Toiviainen, P., "Statistical features and perceived similarity of folk melodies," *Music Perception*, vol. 18, no. 3, pp. 275–296, 2001.

[43] Erkut, C. and Valimaki, V., "Model-based sound synthesis of tanbur, a Turkish long-necked lute," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2000.

[44] Ezgi, S., "Ameli ve nazari Türk musikisi," *İstanbul Konservatuvarı Neşriyatı*, 1935.

[45] Feldman, W., "Cultural authority and authenticity in the Turkish repertoire," *Asian music*, vol. 22, no. 1, pp. 73–111, 1990.

[46] Futrelle, J. and Downie, J., "Interdisciplinary communities and research issues in music information retrieval," in *Proceedings of International Society on Music Information Retrieval (ISMIR)*, 2002.

[47] Gedik, A. and Bozkurt, B., "Evaluation of the makam scale theory of Arel for music information retrieval on traditional Turkish art music," *Journal of New Music Research*, vol. 38, no. 2, pp. 103–116, 2009.

[48] Gedik, A. and Bozkurt, B., "Pitch-frequency histogram-based music information retrieval for Turkish music," *Signal Processing*, vol. 90, no. 4, pp. 1049–1063, 2010.

[49] Gillet, O. and Richard, G., "Automatic labeling of tabla signals," in *Proceedings of International Society on Music Information Retrieval (ISMIR)*, Citeseer, 2003.

[50] Gündüz, G. and Gündüz, U., "The mathematical analysis of the structure of some songs," *Physica A: Statistical Mechanics and its Applications*, vol. 357, no. 3-4, pp. 565–592, 2005.

[51] Holzapfel, A., *Similarity methods for computational ethnomusicology*. PhD thesis, University of Crete, 2010.

[52] Huron, D., "Music information processing using the humdrum toolkit: Concepts, examples, and lessons," *Computer Music Journal*, vol. 26, no. 2, pp. 11–26, 2002.

[53] Huron, D., *Sweet anticipation: Music and the psychology of expectation.* The MIT Press, 2006.

[54] İlerici, K., *Bestecilik bakımından Türk müziği ve armonisi.* Milli Eğitim Basımevi, 1981.

[55] Jacob, B., "Algorithmic composition as a model of creativity," *Organised Sound*, vol. 1, no. 3, pp. 157–165, 1996.

[56] Kapur, A., Singer, E., Benning, M., Tzanetakis, G., and Trimpin, "Integrating hyperinstruments, musical robots & machine musicianship for North Indian classical music," in *Proceedings of the 7th international conference on New interfaces for musical expression*, pp. 238–241, ACM, 2007.

[57] Krumhansl, C., Louhivuori, J., Toiviainen, P., Järvinen, T., and Eerola, T., "Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches," *Music Perception*, vol. 17, no. 2, pp. 151–195, 1999.

[58] Lartillot, O., Dubnov, S., Assayag, G., and Bejerano, G., "Automatic modeling of musical style," in *Proceedings of the 2001 International Computer Music Conference*, pp. 447–454, Citeseer, 2001.

[59] Lee, K. and Slaney, M., "Automatic chord recognition from audio using an HMM with supervised learning," in *Proc. ISMIR*, pp. 133–137, Citeseer, 2006.

[60] Leonidas, I., "Estimating the makam of polyphonic music signals: template-matching vs. class-modeling," Master's thesis, Universitat Pompeu Fabra, 2010.

[61] Lidy, T., Silla Jr, C., Cornelis, O., Gouyon, F., Rauber, A., Kaestner, C., and Koerich, A., "On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-Western and ethnic music collections," *Signal Processing*, vol. 90, no. 4, pp. 1032–1048, 2010.

[62] Mallikarjuna, T., "Towards expressive melodic accompaniment using parametric modeling of continuous musical elements in a multi-attribute prediction suffix trie framework," Master's thesis, Georgia Institute of Technology, 2010.

[63] Manning, C. and Schutze, H., *Foundations of Statistical Natural Language Processing*, pp. 60–78. MIT Press, 2002.

[64] Merriam, A., "Ethnomusicology discussion and definition of the field," *Ethnomusicology*, vol. 4, no. 3, pp. 107–114, 1960.

[65] Moelants, D., Cornelis, O., and Leman, M., "Exploring African tone scales," in *Proceedings of the 10th conference of the International Society of Music Information Retrieval*, p. 489, 2009.

[66] Moelants, D., Cornelis, O., Leman, M., Gansemans, J., De Caluwe, R., De Tré, G., Matthé, T., and Hallez, A., "The problems and opportunities of content-based analysis and description of ethnic music," *International Journal of Intangible Heritage*, vol. 2, pp. 57–68, 2007.

[67] Mok, R., "Heterophony in chinese folk music," *Journal of the International Folk Music Council*, vol. 18, pp. 14–23, 1966.

[68] Myers, H., *Ethnomusicology: an introduction.* WW Norton, 1992.

[69] Özkan, I., *Türk mûsikısi nazariyatı ve usûlleri: kudüm velveleleri.* Ötüken Neşriyat, 2006.

[70] Öztürk, O. M., *Zeybek Kültürü ve Müziği.* Pan Yayınları, 2006.

[71] Öztürk, O. M., "Onyedi'den yirmidört'e: Bağlama ailesi çalgılar ve geleneksel perde sistemi," in *Uluslararası Halk Müziğinde Çalgılar Sempozyumu*, Kocaeli Üniversitesi - Motif Vakfı, 2007.

[72] Öztürk, O. M., "Modal structure in Afyon folk songs," in *Dinar 1. Ukuslararası Marsyas Kültür Sanat ve Müzik Festivali Sempozyum/Panel Bildirileri*, pp. 92–113, 2010.

[73] Öztürk, O. M., "Osmanlı musikisinde modernleşme ve başkalaşım," in *Uluslararası Osmanl Dönemi Türk Musikisi Sempozyumu*, pp. 92–113, 2010.

[74] Pachet, F., "The continuator: Musical interaction with style," *Journal of New Music Research*, vol. 32, no. 3, pp. 333–341, 2003.

[75] Papadopoulos, G. and Wiggins, G., "AI methods for algorithmic composition: A survey, a critical view and future prospects," in *AISB Symposium on Musical Creativity*, pp. 110–117, Citeseer, 1999.

[76] Pearce, M., Conklin, D., and Wiggins, G., "Methods for combining statistical models of music," *Computer Music Modeling and Retrieval*, pp. 295–312, 2005.

[77] Pearce, M., *The construction and evaluation of statistical models of melodic structure in music perception and composition.* PhD thesis, City University, 2005.

[78] Pearce, M., Ruiz, M., Kapasi, S., Wiggins, G., and Bhattacharya, J., "Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation," *NeuroImage*, vol. 50, no. 1, pp. 302–313, 2010.

[79] Pearce, M. and Wiggins, G., "Improved methods for statistical modelling of monophonic music," *Journal of New Music Research*, vol. 33, no. 4, pp. 367–385, 2004.

[80] Popescu-Judetz, E., *Meanings in Turkish musical culture*, vol. 1. Pan (Istanbul, Turkey), 1996.

[81] Rabiner, L., "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[82] Rae, A., "Generative rhythmic models," Master's thesis, Georgia Institute of Technology, 2009.

[83] Randel, D., *The Harvard Dictionary of Music.* Harvard University Press Reference Library, Belknap Press of Harvard University Press, 2003.

[84] Ron, D., *Automata Learning and its Applications.* PhD thesis, the Hebrew University, 1995.

[85] Said, E., *Orientalism. 1979.* Vintage Books, 1979.

[86] Sapp, C., "Online database of scores in the humdrum file format," in *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, pp. 664–665, 2005.

[87] Seeger, C., "Prescriptive and descriptive music-writing," *The Musical Quarterly*, vol. 44, no. 2, p. 184, 1958.

[88] Shannon, C., "Prediction and entropy of printed English," *Bell System Technical Journal*, vol. 30, no. 1, pp. 50–64, 1951.

[89] Signell, K., *Makam: modal practice in Turkish art music.* Da Capo Press, 1986.

[90] Sloboda, J., *Generative processes in music.* Clarendon Press, 2000.

[91] Sümbüllü, H., "Gelenekel Türk halk müziği kuramı açısından aşıklık geleneğinde kullanılan ayak kavramı," *Atatürk Üniversitesi Güzel Sanatlar Fakültesi Dergisi*, no. 12, 2007.

[92] Suyoto, I. and Uitdenbogerd, A., "Simple efficient n-gram indexing for effective melody retrieval," *Proceedings of the Annual Music Information Retrieval Evaluation exchange*, 2005.

[93] Tanrikorur, C., "The Ottoman music." http://www.turkmusikisi.com/osmanli_musikisi/the_ottoman_music.htm.

[94] Toiviainen, P., "Modeling the target-note technique of bebop-style jazz improvisation: An artificial neural network approach," *Music Perception*, vol. 12, no. 4, pp. 399–413, 1995.

[95] Triviño-Rodriguez, J. and Morales-Bueno, R., "Using multiattribute graphs to predict and generate music," *Computer Music Journal*, vol. 25, no. 3, pp. 62–79, 2001.

[96] Tsaxalinas, K., Tzedaki, K., Cook, P., Psaroudakes, S., Kamarotos, D., and Rikakis, T., "Physical modeling simulation of the ancient Greek elgin avloi," in *International Computer Music Conference*, 1997.

[97] Tura, Y., *Türk Musıkisinin Meseleleri.* Pan Yayıncılık, Istanbul, 1988.

[98] Tzanetakis, G. and Cook, P., "Musical genre classification of audio signals," *Speech and Audio Processing, IEEE transactions on*, vol. 10, no. 5, pp. 293–302, 2002.

[99] Tzanetakis, G., Kapur, A., Schloss, W., and Wright, M., "Computational ethnomusicology," *Journal of interdisciplinary music studies*, vol. 1, no. 2, pp. 1–24, 2007.

[100] Varèse, E. and Wen-chung, C., "The liberation of sound," *Perspectives of new music*, vol. 5, no. 1, pp. 11–19, 1966.

[101] Wiggins, G., "Semantic gap?? schemantic schmap!! methodological considerations in the scientific study of music," in *2009 11th IEEE International Symposium on Multimedia*, pp. 477–482, IEEE, 2009.

[102] Witten, I. and Bell, T., "The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression," *Information Theory, IEEE Transactions on*, vol. 37, no. 4, pp. 1085–1094, 1991.

[103] Xenakis, I., *Formalized Music: Thought and mathematics in composition.* Pendragon Pr, 1992.

[104] Yarman, O., *79-tone tuning & theory for Turkish maqam music.* PhD thesis, İstanbul Teknik Üniversitesi Sosyal Bilimler Enstitüsü, 2008.