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

Sertan Şentürk
Georgia Tech Center for Music Technology
Atlanta, GA, USA
sertansenturk@gatech.edu

Parag Chordia
Georgia Tech Center for Music Technology
Atlanta, GA, USA
ppc@gatech.edu

ABSTRACT

The paper describes a new database, which currently consists of 64 songs encompassing approximately 6600 notes, and a system, which uses Variable-Length Markov Models (VLMM) to predict the melodies in the *uzun hava* (long tune) form, a melodic structure in Turkish folk music. The work shows VLMMs are highly predictive. This suggests that variable-length Markov models (VLMMs) may be applied to *makam*-based and non-metered musical forms, in addition to Western musical traditions. To the best of our knowledge, the work presents the first symbolic, machine readable database of *uzun hava*s and the first application of predictive modeling in Turkish folk music.

1. INTRODUCTION AND MOTIVATIONS

To date, most computational research in music has focused on Western music. In order to further advance the state-of-the-art in MIR, non-Western musics, with their unique challenges should be considered [16]. Such research would expand our knowledge and tools immensely, allowing us to adapt and improve the previous work, and would open up new paths for musical creativity, expressivity and interaction. Computational modeling of distinct musical genres will deepen our knowledge of universal versus genre-specific aspects of music and it will allow us to truly evaluate the generality of various modeling strategies.

Musical improvisation is a complex phenomenon, and there have been many attempts to describe and model it [24]. Moreover, there is a lack of understanding the "music" in the current MIR research with respect to how humans actually perceive the it [27]. Previous work on Western melodies showed that variable-length n-gram models and human judgments of melodic continuation are highly correlated [20]. We hope our research will give clues about how we actually anticipate music [10].

2. BACKGROUND

2.1 Related Work

Computational modeling of musical styles is not a new topic, and is a common tool in algorithmic composition [2, 7]. n-gram modeling have been extensively used in algorithmic composition [19], structure analysis [12], and music cognition [21]. This work is an adaptation of our expressive tabla modeling research [4], which is based on multiple viewpoint modeling [6].

Although information retrieval in world musics has only recently started to attract attention in academia, there has been substantial amount of research in the field [3,8,13,26]. In traditional Turkish music, *n-gram modeling* have been previously used by Alpkoçak and Gedik to classify *makam*s [1].

2.2 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, and has been shaped by the historical events, geographical locations and migrations of the Turkish people. The songs in Turkish folk music are typically anonymous, which have been carried from generation to generation as an oral tradition.

2.2.1 Basic Concepts in Turkish Music Theory

In Western music, an octave is divided into 12 intervals. However, there is no theory that is completely agreed upon in Turkish music due to differences in theory and practice; the suggested number of pitches in an octave ranges from 17 to 79 [28]. Currently, education in *makam* based music is based on the-highly-criticized [25] Arel-Ezgi-Uzdilek theory. According to the theory, a whole tone is divided into intervals named *koma*s, which are used to discretize an octave into 24 consequent tones [18]. However, in Turkish

Fret #	Note	Fret #	Note	Fret #	Note
0	A	6	$C\sharp$	12	$F\sharp^3$
1	$B\flat$	7	D	13	$F\sharp$
2	$B\flat^2$	8	$E\flat$	14	G
3	B	9	$E\flat^2$	15	$A\flat$
4	C	10	E	16	$A\flat^2$
5	$C\sharp^3$	11	F	17	A

Table 1: The notes and the fret numbers in the lowest string group of bağlama in the *bağlama tuning*. $\flat^2$ and $\sharp^3$'s indicate the quarter tones.

folk music, there are typically 17 pitches played in an octave due to selection of the instruments (Section 2.2.2), and the music is indeed explainable by *makam*s [25].

*Makam*s can be depicted as the modes of Turkish music. *Makam*s are progressions (*seyir*) used to generate melodies (*nağme*). *Makam*s obey certain rules such as emphasizing the modal centers, the use of key signatures and maintaining context-specific ascending or descending seyirs [25].

Usul is ”the structure of musical events which are coherent with respect to time.” *Usul* can be roughly translated as ”meter.” An *usul* can be as short as two beats or as long as 128 beats, but it should always have at least one strong and one weak beat. Turkish music also makes a rich use of *usulsüz* (non-metered) progressions [18].

2.2.2 Uzun Hava

Uzun hava (long tune) is a semi-improvisational melodic structure in Turkish folk music. The music is usually sad; the lyrics (if any) are generally about the daily struggles and emotions of the Anatolian people. All *uzun hava*s are *usulsüz* (without any meter), but there can also be *usullü* (with distinct meter) sections in between.

The most common instrument played in *uzun hava*s is bağlama, a traditional Turkish instrument from lute family. It has 17 notes in an octave [25] (Table 1). The strings are grouped into three having 3, 2 and 2 strings from the highest to the lowest. The instrument is a transposing instrument, and the tuning (*düzen*) of these strings may change for different songs. Moreover, the frets are tied to the fretboard (*sap*), so that microtonal adjustments in the temperament can be easily made.

3. UZUN HAVA HUMDRUM DATABASE

For the experiments, the authors, with the help of Prof. Erdal Tuğcular, have built the *Uzun Hava Humdrum Database* [1]. A Humdrum based syntax called **kern format was chosen for its readability and broad search, comparison and editing capabilities [9]. In order to obtain the symbolic data, all of the *uzun hava*s with scores (a total of 123 scores) from The Turkish Radio and Television Corporation's (TRT) Turkish Folk Music Database were chosen [2]. The TRT database consists of the extended Western staff notations saved in .tiff image format.

In the analysis of world musics, there are some intrinsic problems of using symbolic notation such as accepting notation as an adequate means of representing improvisation (especially in oral traditions) and human errors in the transcriptions [17]. Therefore, it might be problematic to make deductions based on symbolic notations, and audio analysis might be more appropriate. Yet, audio analysis is generally not as easy and straightforward as processing symbolic data. Thus, it is more suitable take the initial steps in computational modeling with human annotations, even if they are not perfect.

The scores in the TRT database were read into Finale 2010 by using the built-in SmartScore 5 Lite, exported into MusicXML 2.0 format, and then converted to **kern notation by using xml2hum [23]. After cleaning-up, grace notes, fermatas, quarter tone accidentals and meter changes were added to the **kern files. In order to comply with the standard humdrum notation, instead of creating our own symbols, we have chosen to indicate the quarter tones as deviations in cents in a second spine. 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 the most common instrument played in *uzun havas* is bağlama and it has 17 notes per octave, we have chosen to map all koma values lying between semitones into a single quarter tone with 50 cent deviation from the original note and match the 17-tone scale.

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 and each note is proceeded by the letter ”Q”, which is used to indicate grupettos in **kern format [9]. Finally, the name, region, *makam*, accidentals and *usul* are printed to the start of the file as comments.

Currently 64 songs have been encoded, with a total of 6613 notes, in 8 *makam*s from different regions of Anatolia and Azerbaijan. However, we should note that the *makam*s of the songs are biased towards Hüseyni (82 songs) and Hicaz (17 songs). This is expected as *uzun hava*s are usually played in Hüseyni [11].

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

[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/

4. COMPUTATIONAL MODELING

Parallel to Conklin and Pearce's research [6,20], the computational framework in this work incorporates multiple viewpoints modeling with both long-term and short-term models. Variable-length Markov modeling (VLMM) is used to model the sequences, and the training data is stored as Prediction Suffix Trees. The evaluation of the system is done by entropy-based calculations. The modeling and evaluation framework was implemented in C++ as an external object in Max/MSP along with supporting patches [4]. To the best knowledge, this research is the first attempt to model melodic sequences in traditional Turkish music.

4.1 Markov Modeling

A n^{th} order Markov model is a causal, discrete random process where the probabilities of the next state depends only on the probabilities of the current and the previous states. 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. A $(n-1)^{th}$ Markov model can be represented by n-grams, which are subsequences of length n. n-grams are a commonly used to probabilistically model sequences of elements such as phonemes in speech, letters in a word, or musical notes in a phrase [15].

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, integrating lower order models to the system might also be useful to give 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. In this work, to combine the predictions from different orders, we are using a smoothing method we termed $1/N$. In the $1/N$ smoothing method, weights for the n-th order model are given by $\frac{1}{(maxOrder-n+1)}$, giving greater relative weight to predictions of higher orders. Moreover, the VLMMs are efficiently stored in Prediction Suffix Trees [22] (PSTs) for performance reasons (Figure 1).

While increasing order would allow us to obtain more specific patterns, it also brings the so-called zero frequency problem [5]. 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 very few. This sparsity issue brings a limitation to the order of an n-gram. In order to deal with the zero frequency problem, an escape probability for each level of the trie is reserved. The escape probability of each level is calculated as $e(n) = \frac{T_1(n)}{N(n)}$, where T_1 is the number of symbols that have occurred exactly once and N is the total number of observations so far. When an event, which has never occurred before, is observed, the escape probability is returned instead of 0.

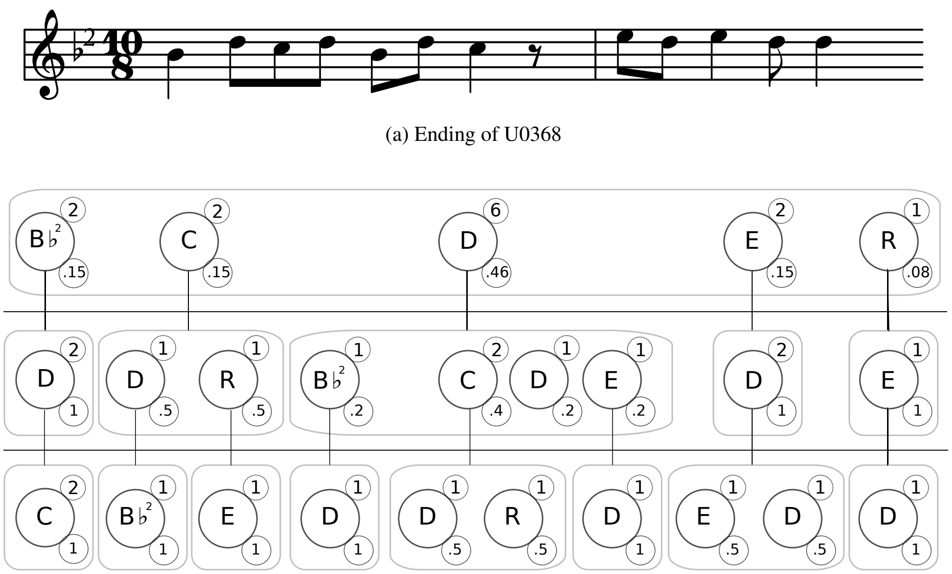
Viewpoint	Explanation
Duration	Duration of the note
Note	Midi number corresponding to the note
NoteWCents	Viewpoint denoting the "true" symbol in Turkish folk music in note and cent deviation, i.e. the floating midi number
Note⊗Dur	Cross type combining note and duration
NoteWCents⊗Dur	Cross type combining note with cent deviation and duration

Table 2: Viewpoints used in the experiment.

4.2 Multiple Viewpoints

A multiple viewpoints system [6,20] separates a musical sequence to independent parallel representations such as pitch, rhythm, instrument, key changes. The next sequence is predicted based on the information incorporated from these viewpoints. Denoting music in multiple representations can be useful to predict the next symbol when one of the representations might be suitable for that particular sequence, whereas another representation is useful in other situations. 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. There can also be cross-type viewpoints, which are generated by mapping the symbols in two or more of the parallel representations into unique tokens: for example for Notes⊗Durations; a quarter C, a quarter D, a eighth C will all be mapped to different symbols. We use 5 viewpoints in our experiment: Durations, Notes, NoteWCents, Note⊗Dur and NoteWCents⊗Dur (Table 2).

A common limitation of training the predictive models over large amount of data is that it renders the model too general to effectively predict patterns specic to the current song: if the song has a peculiar phrasing repeated throughout, due to the phrase having a small probability in the training database, the patterns generated might be irrelevant. In order to obtain predictions which are trained over a particular style and also sounds like a specific song, we use a long-term-model (LTM) built on the entire training set and



(a) Ending of U0368

(b) Prediction Suffix Tree

Figure 1: The ending of U0368 with the repeat sign taken out and the Prediction Suffix Tree representing the Markov models of Notes-with-Cents viewpoint with a maximum order of 2, trained on these two measures. Bubbles on the top right and bottom right of each node denotes the count and the probability of the node respectively.

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 combined and normalized to a single predictive distribution for each of the viewpoints. Given the symbols, $S = \{s_1, s_2, ..., s_N\}$ forming the probability distribution, the probability distribution is weighted inversely proportional to the entropy [6]. The weight of the probability distribution of a model is given as $\omega_m \triangleq \frac{\log_2(N)}{H_m}$, and the entropy of the probability distribution of each model is defined as $H_m \triangleq -\sum_{k=1}^{N} P_m(s_k)\log_2(P_m(s_k))$, where $P_m(s_k)$ is the probability of the symbol, s_k, at the time step, t.

5. EVALUATION

Leave-one-out cross-validation was performed on each of the 64 songs in the *Uzun Hava Humdrum database*. 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_t the probability of the true symbol at t is recorded. From the probabilities, cross-entropy [14] 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, and it is preferable to symbol recognition rate in predictive systems [6, 20]. It is defined as: $H_c = -\sum_{t=1}^{n} p_t \log_2(p_t)$ [3]. If the probability distribution p is unknown, under the assumption of uniform probability distribution ($p_t = \frac{1}{n}$, where n is the number of predictions throughout the experiment), cross-entropy can be approximated by $H_c \approx -\frac{1}{n}\sum_{t=1}^{n}\log_2(p_t)$. Later cross-entropy is converted to average perplexity, which is a measure of the number of choices that the model has picked the true symbol [14]. Perplexity is defined as $P = 2^H$. We also report median perplexity in addition to average perplexity. The prior probabilities of the symbols are used to obtain a baseline for evaluating perplexity results. In other words, perplexity of the 0^{th} order model LTM is used as the baseline.

6. RESULTS

During the experiments, average and median perplexities over the whole dataset and in the song-level are recorded for STM, LTM and combined models with different orders [4]. Table 3 shows that for order 14, STM always gives the most

[3] Notice that the definition of cross-entropy is very similar to the entropy definiton in Section 4.2. However, entropy is calculated from the probabilities of each possible symbol at a given time, whereas cross-entropy is calculated from the chosen predictions at each time step.

[4] The complete set of results and significance tests is available at http://sertansenturk.com/uploads/publications/senturk2011UzunHava

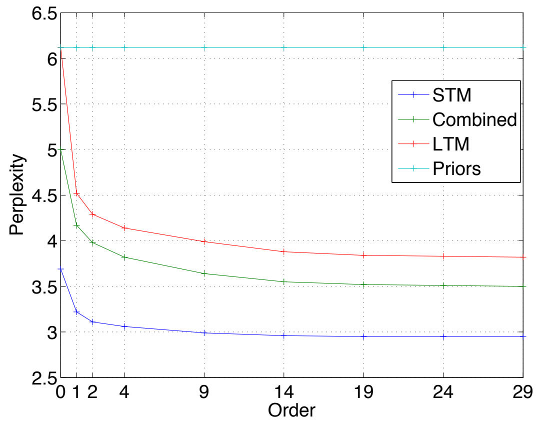


Figure 2: Average perplexity for duration prediction using LTM, STM and combined models for orders 0-29

confident results, while combining STM with LTM does not actually help predictions. STM has an average perplexity of 2.96, 4.13, 4.16, 6.68 and 6.69 for Duration, Note and NoteWCents, Note⊗Dur and NoteWCents⊗Dur respectively. Comparing to the average baseline perplexities (6.12, 11.9, 12.71, 171.76, 148.39), there is a remarkable decrease. The power of STM is even more obvious in the cross types Note⊗Dur and NoteWCents⊗Dur, where the LTM gives perplexities of 30.17 and 31.84.

Another interesting remark is adding the cent information during prediction results in a slight increase in perplexity. For the 14^{th} order model, the average perplexities in the STM are 4.13, 4.16 for Note and NoteWCents, and 6.68, 6.69 for Note⊗Dur and NoteWCents⊗Dur respectively, meaning the system can effectively predict notes with quarter tone accidentals.

Figure 2 shows that the perplexity decreases monotonically with increasing order, as expected. 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 unnecessary. This optimum order is true for all of the viewpoints.

When the average perplexities are checked song by song, it was observed that some songs had exceptionally higher perplexities for LTM and Combined models. Upon inspecting, it was observed that the system was not able to predict the notes properly in the songs with *makam*s which are only represented with a few songs. Similarly the songs which included a lot of triplets, double dotted, 64^{th} notes were harder to predict. On the other hand, the latter problem also affected STM, because the Duration viewpoint in these songs presented a vast symbol space and thus smaller prior values, rendering the next symbol harder to predict.

7. DISCUSSION

The results suggest that *uzun hava* form can be effectively modeled using VLMMs. Between the perplexities obtained from the LTM, STM and Combined models with a maximum order of 14 and viewpoint, there is a significant [5] decrease in confusion, and STM outperforms both LTM and the Combined model. The success of STM over LTM suggests the songs have strong local patterns. Strong patterns are easy to be captured and predicted by the STM; however being a more general model, LTM cannot capture and prioritize song-related patterns as good as STM. This result was expected, because in *uzun hava*s note and the duration repetitions commonly occur during the improvised part and melodies are generally repeated during vocal sections.

One of the most important observations is that extending the possibilities in target pitches from Western music to Turkish music only slightly increases perplexity values. 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, almost all of the counts accumulate on the one note. Additionally, by inspecting perplexities note-by-note, it is easily seen that the quarter tones such as $B\flat^2$ and $F\sharp^3$, are easily distinguished from their neighbor tones, i.e. $B\flat$ and $F\sharp$. This shows that transcriptions strictly obey the key signature of their *makam*s, and multiple viewpoint system is able to model the context-specific pitches in *makam*s, and distinguish the notes from the neighboring notes present in Western music virtually without any penalty. Indeed, the selection of multiple viewpoints might be crucial for the success. For example, the cent deviation information cannot be used without crossing pitch related viewpoints such as absolute note or scale degree. For a generative system, decoupling them might still give good average perplexities, however when the note and the cent deviation are predicted independently from each other, the results might introduce notes with wrong accidentals, disrupting the melodic intervals and the *makam* structure.

In future work, we would like to include more viewpoints incorporating fermata, *usul*, scale degree, melodic interval, contour and try crossing these viewpoints, to obtain better perplexities in prediction. We would also like to generate Medium Term Models (MTM), each one of which will be trained on a single *makam*. Testing will be carried with the MTM of the same *makam*. Using this approach, we hope to find and predict *makam* based patterns with better perplexity. Also, as mentioned in Section 3, extending the framework to variable-length hidden Markov Models (VLHMMs) for audio analysis is a necessary step for a more relevant assessment of *uzun hava*s.

[5] The claim means, it is statistically significant at the 0.01 level as determined by a multiple comparison test using the Tukey-Cramer statistic.

	Duration		Note		NoteWCents		Note⊗Dur		NoteWCents⊗Dur	
	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median
Priors	6.12	3.76	11.9	7.97	12.71	7.98	171.76	171.32	148.39	148.16
LTM	3.88	2.23	5.56	4.13	5.87	4.21	30.17	20.06	31.84	21.21
Combined	3.55	1.93	4.64	3.17	4.70	3.21	15.67	10.40	16.23	10.68
STM	2.96	1.94	4.13	2.96	4.16	3.00	6.68	5.30	6.69	5.30

Table 3: Average and median perplexities for Duration, Note, NoteWCents, Note⊗Dur and NoteWCents⊗Dur for order 14

8. ACKNOWLEDGEMENTS

We would like to thank Turkish Radio and Television Corporation, *Türk Müzik Kültürünün Hafızası Score Archive* and the numerous musicians, transcribers and archivers for their efforts in building the *TRT Turkish folk music database* which made this research possible. We would also like to thank Prof. Erdal Tuğcular for his invaluable contributions to the *Uzun Hava Humdrum Database*, Avinash Sastry for his help in the Max/MSP framework and Prof. Nezihe Şentürk for her assistance in Turkish music theory. This material is based upon work supported by the National Science Foundation under Grant No. IIS-0855758.

9. REFERENCES

[1] A Alpkoçak and A.C Gedik. 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.

[2] C. Ames. The markov process as a compositional model: a survey and tutorial. *Leonardo*, 22(2):175–187, 1989.

[3] W. Chai and B. Vercoe. Folk music classification using hidden markov models. In *Proceedings of International Conference on Artificial Intelligence*, volume 6, 2001.

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

[5] J.G. Cleary and W.J. Teahan. Experiments on the zero frequency problem. In *Proc. Data Compression Conference*, volume 480. Citeseer, 1995.

[6] D. Conklin and I.H. Witten. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1):51–73, 1995.

[7] D. Cope. Experiments in musical intelligence (EMI): non-linear linguistic-based composition. *Journal of New Music Research*, 18(1):117–139, 1989.

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

[9] D. Huron. Music information processing using the humdrum toolkit: Concepts, examples, and lessons. *Computer Music Journal*, 26(2):11–26, 2002.

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

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

[12] K. Lee and M. Slaney. Automatic chord recognition from audio using an hmm with supervised learning. In *Proc. ISMIR*, pages 133–137. Citeseer, 2006.

[13] T. Lidy, C.N. Silla Jr, O. Cornelis, F. Gouyon, A. Rauber, C.A.A. Kaestner, and A.L. Koerich. 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*, 90(4):1032–1048, 2010.

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

[15] C.D. Manning, H. Schütze, and MITCogNet. *Foundations of statistical natural language processing*, volume 59. MIT Press, 1999.

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

[17] H. Myers. *Ethnomusicology: an introduction*, pages 110–164. WW Norton, 1992.

[18] İ.H. Özkan. *Türk mûsikısi nazariyatı ve usûlleri: kudüm velveleleri*. Ötüken Neşriyat, 2006.

[19] F. Pachet. The continuator: Musical interaction with style. *Journal of New Music Research*, 32(3):333–341, 2003.

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

[21] M.T. Pearce, M.H. Ruiz, S. Kapasi, G.A. Wiggins, and J. Bhattacharya. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, 50(1):302–313, 2010.

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

[23] C.S. Sapp. Online database of scores in the humdrum file format. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 664–665, 2005.

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

[25] Yalçın Tura. *Türk Musıkisinin Meseleleri*. Pan Yayıncılık, Istanbul, 1988.

[26] G. Tzanetakis, A. Kapur, W.A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of interdisciplinary music studies*, 1(2):1–24, 2007.

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

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