A TURKISH MAKAM MUSIC SYMBOLIC DATABASE FOR MUSIC INFORMATION RETRIEVAL: SymbTr

M. Kemal Karaosmanoğlu

Yıldız Technical University kkara@yildiz.edu.tr

ABSTRACT

Turkish makam music needs a comprehensive database for public consumption, to be used in MIR. This article introduces SymbTr, a Turkish Makam Music Symbolic Representation Database, aimed at filling this void. SymbTr consists of musical information in text, PDF, and MIDI formats. Raw data, drawn from reliable sources, and consisting of 1,700 musical pieces in Turkish art and folk music was processed featuring distinct examples in 155 diverse makams, 100 usuls and 48 forms. Special care was devoted to selection of works that scatter across a broad historical time span and were among those still performed today. Total number of musical notes in these pieces was 630,000, corresponding to a nominal playback time of 72 hours. Synthesized sounds particular to Turkish makam music were used in MIDI playback, and transcription/playback errors were corrected by input from experts. Symbolic representation data, open to the public, is output from a computer program developed exclusively for Turkish makam music. SymbTr was designed as a wholesome representation of aforementioned distinct auditory and visual features that distinguish Turkish makam music from other music genres. This article explains the database format in detail, and also provides, through examples, statistical information on pitch/interval allocation and distribution.

1. INTRODUCTION

Turkish makam music is a genre drawing roots from a thousand year old tradition, featuring distinct melodic patterns called *makam* and rich rhythmic structures called *usul*. Since the number of tones per octave is greater in Turkish makam music, compared to Western music, several sharp and flat accidentals appear in printed scores. Additionally, one must take into consideration a multitude of idiosyncratic rhythmic structures. Although there exists only one version of the score, independent of the instrument or key, musicians perform improvised transpositions during performance, as permitted by the ranges of their instruments and the vocalist on hand. Probably the most prominent feature of Turkish makam music is its

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2012 International Society for Music Information Retrieval

monophonic —and incidentally heterophonic— structure. Another characteristic is the number of notes in an octave: 17, 24, and, according to some musicologists, even a greater number of tones to the octave make up the pitch palette of Turkish makam music [12], [16]. Although displaying a higher pitch count compared to Western music, there is no one-to-one correlation between the fixed frequency values, music theory, implied in engraved scores and what is actually performed in practice [3].

Everything mentioned up to this point was to differentiate Turkish makam music from many other world music genres. It then follows; data structures and algorithms developed for other musical traditions are not directly applicable to Turkish makam music. On the other hand, there are only a handful of researchers working on computational models for Turkish makam music. There remains much to be done in areas related to data collection/compilation, algorithm development, and research. *SymbTr* is hopefully a likely candidate to be a pioneer in the field, since it is capable of accommodating and expressing information specific to makam music. Secondly, early studies ([9], [18]) have returned encouraging results. It is anticipated that *SymbTr* might provide a setting for scholars interested in makam music, potentially



Figure 1. A Turkish folksong's scoring in (a) KTM, (b) THM, and (c) mixed format

stimulating further research at a global level.

2. STYLES OF TURKISH MAKAM MUSIC

Turkish makam music is viewed under two major headings:

- 1. Classical Turkish music (KTM),
- 2. Turkish folk music (THM).

Since both styles originate from the same cultural roots, their modal motifs and rhythmic structures are very similar in character [17]. Owing to political movements emerging at the turn of the 20th Century, a superficial bifurcation took place, which led to a divergence between the two styles resulting in two separate traditions. Today, these two traditions differ considerably when it comes to their respective theoretical models, notation systems, and terminology. Fig. 1 shows the first two measures from a folksong score, which is part of both the THM and KTM repertoires of TRT ¹. Scores are shown in three different notational systems.

As can be detected in the scores, accidental symbols, in particular, are different even though the melody is essentially the same: In the KTM version reverse and hooked flat signs (Fig. 1-a) represent the accidentals, while in the THM version superscripts over ordinary flat signs are used (Fig. 1-b). The mixed notation in (Fig. 1-c) displays a combination of the two. Reverse and hooked flats, by definition, lower a note by 1 and 4 Holdrian commas (Hc) ² respectively. However, many of the measurements ([1], [4], [5]) evince that, printed scores for works in the Saba makam should carry a 2 comma flat sign for B and a 3 comma flat sign for D as the key signature. Indeed, these values are substituted in THM and mixed notations.

The difference between KTM and THM notation lies not only in the symbols representing the accidentals. In KTM notation there are almost no ornamentation symbols on the score. In THM, on the other hand, ornamentation is achieved by repeated use of notes with smaller rhythmic values, as shown in the trill in Fig. 1-b [17]. When such passages are converted into *SymbTr*, note clusters representing ornamentation are indicated by a single *core* note, with its type shown in the *Code* field.

Because of the fact that THM and KTM have mixed and intertwined traditionally, the *SymbTr* database naturally accommodates pieces from both categories. Format in the database was, therefore, designed to reconcile the artificial disparity between the two traditions. The most important design element for the database was the fundamental tuning selected. The Arel - Ezgi (AE) tonesystem, which has been recognized and widely adopted as the official KTM system since the 1950s, has 24 notes in an octave. In contrast, THM has adopted a notation with 17 notes to the octave. Twelve of the said 17 notes are common with the AE system. Moreover, both tonal scales

present a near-perfect subset of 53 tone equal temperament (53TET), with deviations less than 1 cent [19] (Fig. 2). Possibly due to this structural connection, Turkish makam music education has been built around 53TET, whether acknowledged by name (Ayomak, Sarısözen) or

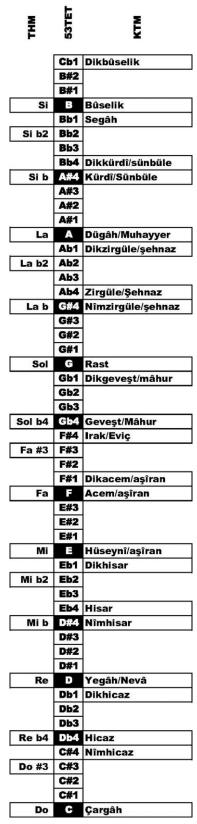


Figure 2. 17 tones in THM (left column), 24 tones in KTM (right column), and 53TET in between

¹ Turkish Radio/TV Broadcasting Corporation.

² Interval unit obtained by the division of the octave into logarithmically equal 53 parts: $Hc = 1200 / 53 \approx 22.5$ cent. In this article comma signifies the Holdrian comma.

not [13]. Hence, the term "comma", when describing *makams* and preparing printed scores, refers to the Holdrian comma as the basic intervallic unit, obtained by equally dividing the octave into 53 equal parts. Selecting 53TET as the master underlying tuning in *SymbTr* also facilitates transpositions across *ahenks* (pitch-levels). *Ahenks* can be defined as 7 principal and 5 minor categories corresponding to 12 chromatic pitch-levels akin to what key transposing instruments of Western music accomplish. Detailed information about *ahenks* and Turkish makam music in general can be found in [10] and [14].

3. MAKAM MUSIC AND SYMBOLIC DATA

SymbTr database is generated by using the output from a computer program Mus2-Alpha, developed by the author of this article. This software is the first notation and playback application for Turkish makam music to the best of our knowledge. All pieces in the database were entered manually using the said software. Printed scores and MIDI files were, then, prepared for every piece in the database. Initially, before the introduction of Mus2-Alpha and its sister applications (Nota 2.2 1, Notist 2), scores were engraved either manually or using programs such as Finale or Sibelius, that were developed solely to transcribe Western music. Since these programs were not designed to notate flats and sharps specific to Turkish makam music, their standard output formats such as MusicXML and MIDI have not been useful in research on Turkish makam music [7].

The format for *SymbTr* described in this article was derived from *Mus2-Alpha*'s original format that was used initially to transcribe printable sheet music for pieces in Turkish makam music. Since this format includes reprise markings such as *segno* and *coda*, some modifications for scientific research are necessary. In *SymbTr*, notes are linearized just as they are performed. An advantage associated with *Mus2-Alpha* originating data is that pieces can be amended through consultation with experts, using listening tests based on synthesized sound output. An entry level version of this program, *Mus2okur* ³, has reached thousands of users, thereby resulting in a wide scale screening of possible errors in the database.

The main source of data in *SymbTr* is TRT and other trustworthy archives (Recollection of Turkish Music Culture ⁴), where almost all of them were entered using the AE notation. To synthesize realistic intonations, however, it was necessary to use pitches not included in the AE tone-system. Five notes in the THM scale lie outside the AE scale (Fig. 2). As a courtesy for Turkish musicians, a composite system was adopted in the printout scores of *SymbTr*: Symbols for flats and sharps were taken directly from AE, and numerical superscripts were inserted to express comma-alterations for notes that were not available

Code	Note53	Comma53	NoteAE	CommaAE	Num.	Denom.	ms	TNS	VelOn	Syllable
9	Do5	318	C5	318	1	4	667	95	96	Bir
9	Re5b3	324	D5b4	325	1	8	333	99	108	dal
9	Re5b3	324	D5b4	325	1	16	167	99	96	
9	Do5	318	C5	318	1	16	167	95	84	
9	Si4b2	312	B4b1	313	1	4	667	95	72	da
12	Do5	318	C5	318	1	8	333	99	96	i
9	Do5	318	C5	318	1	16	167	99	96	
9	Si4b2	312	B4b1	313	1	16	167	95	96	
9	La4	305	A4	305	1	8	333	99	96	ki
8	Si4b2	312	B4b1	313	1	8	42	99	96	
9	La4	305	A4	305	1	16	167	99	96	
9	Sol4	296	G4	296	1	16	167	95	96	
9	La4	305	A4	305	1	8	333	99	96	ki
9	Si4b2	312	B4b1	313	1	8	333	95	96	
9	Do5	318	C5	318	1	4	1334	45	84	raz

Table 1. *SymbTr* representation of the score in Fig. 1.c

in the tone-system (Fig. 1-c).

4. SymbTr FORMAT

Basic information such as *makam*, *form* and *usul* related to each piece in *SymbTr* is indicated in the filename. In this manner, any piece can be accessed directly from the file system:

Some fields in the *SymbTr* format consist of different representations of the same information. Therefore, one field can be easily converted into the other with the help of the relevant computer code. However, since this additional information requires very little extra storage space, it is provided separately for the convenience of researchers. These basic and readily derived fields are described under common headings below.

Code: Signifies a normal note (#9) or ornamentation. The most commonly used ornamentation codes are as follows: #7 for tremolos, #8 for acciaccatura, #12 for trills, and #23 for mordent.

NoteAE / **CommaAE**: A kind of scientific pitch notation [20]: Indicates note letter, its octave (for example, G5 for gerdaniye), and its comma equivalent (349) (Fig. 3). Notes in THM sheets that do not exist in the AE system are represented by their closest equivalent AE note, e.g. $Mi \ b2 = Dikhisar \ (Eb1)$ (Fig. 2). C4 is the

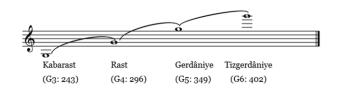


Figure 3. Triple octave operational range of *SymbTr* database and some of the comma numbers

http://www.tulgan.com/Nota22/

² http://notist.org/

³ www.musiki.org

www.sanatmuziginotalari.com/ under http://devletkorosu.com
Please go to the second link 'http'first to reach the main site. Then,
look for and click on the first address 'www'.

note with the frequency of about 262 Hz and numbered as 60 in the MIDI standard. All notes excluding *C*'s have a fractional *MIDI Nr*. The *MIDI Nr* corresponding to *CommaAE* can be computed by the following formula:

$$MIDI_Nr = \frac{CommaAE \cdot Hc}{100}$$
 (1)

In order to represent flats and sharps in *Hc* units, the "b" and "#" prefixes were used respectively. For example, the *segah* note in AE tone-system is represented as *B4b1*; since, according to *AE* theory, it should sound one comma lower than the natural *B* (*Si* - *buselik*). Its comma equivalent is 313, and *MIDI Nr* is 70.87.

Note53 / **Comma53**: Indicates the code and the value of the note in 53TET. If there is no difference between the performance and the sheet music, *CommaAE* and *Comma53* values are the same. However, in some *makam* sequences such as *Uṣṣak*, *Hüzzam*, *Saba* and *Karciĕar*, these two values often vary. For example, in some makams the pitch that corresponds to *B4b1* in AE is *Si4b2* in 53TET, since, in practice, this note should sound 2 commas lower then *Si* (*B*). Its comma equivalent is 312, and *MIDI Nr* is 70.64.

Numerator / **Denominator** and **ms**: Stands for the rhythmic value of the note, with its duration measured in milliseconds. When the tempo (quarter note beats per minute) of the piece is known, these two values can be converted to each other by the following formula:

$$ms = \frac{240\ 000}{Tempo} \cdot \frac{\text{Numerator}}{\text{Denominator}}$$
 (2)

In Turkish makam music, changes in the tempo of a piece is a run-of-the-mill situation (e.g., the 4th section of sazsemaisi pieces are performed faster than other sections), and since the database can be used for rhythmic analysis purposes [9], it was found useful to enter these two strands of information in the same record.

LNS (Legato / Normal / Staccato): Indicates how tied or detached the notes are to be played. This information is extracted by listening to performances in synch with verses and syllables in the lyrics. The default value is 95; that is, the last 5% of the duration time for normal notes is completed with silence. 50 means playback should be of *staccato*. Rest signs are determined using this value.

VelOn: Indicates the volume or strike of the note, making nuanced performance possible. Turkish makam music scores ordinarily do not contain dynamics markings like *piano* or *crescendo*. In *SymbTr* an attempt has

been made to compensate, as much as possible, for this deficiency.

Syllable1: Indicates the syllable corresponding to a note. There is one space character at the end of the syllables that occur at word endings and two space characters at the end of the verses. This information was added to facilitate the tracking of the melody, as well as

for its utility in studies of lyrics-based analyses [8]. In instrumental pieces, it is used to represent the beginning of sections such as "TESLİM", etc... In other places this field is left blank. Instrumental parts of vocal pieces contain a series of dots in this field. In the original *Mus2-Alpha* database, repetitive passages have a separate field for the second syllable. However, due to copyright considerations there is only one field in *SymbTr*.

The representation of the score of Fig. 1-c in *SymbTr* is listed in Table 1. The data starts immediately after the column headings. Fields are tab-delimited.

5. MAKAM MUSIC AND MIDI

It is impossible to produce makam music intonations using ordinary MIDI messages. Therefore, it becomes necessary to use pitch-bend techniques. To generate the needed feature, a pitch-bend message must be sent with the same delta-time value as the note, just before the 'Note on' message. The pseudo-MIDI messages for the first 5 notes in Fig. 1-c are as follows:

Delta	Pitch	Note
Time	Bend	On
0	7 960	C4
4	9 429	D5b
2	9 429	D5b
1	7 960	C4
1	6 492	В

The anchor note is A (La). Therefore, pitch-bend is unnecessary for any A in all octaves. Bend is required for all other pitches. For example, the A – C interval is 13 Hc wide. This value is up to 5.7 cents narrower than the 12TET minor third. Taking into account that 100 cents = 4096 pitch bend units, bending for C is calculated as follows: $8\ 192 - 5.7 \cdot 40.96 \approx 7\ 960$.

MIDI files in *SymbTr* database are not for listening to music. They are included, so that the researchers may find it useful to hear the tune in its simplest *raw* form. To this end, even the instrument information has not been added. Voicing is done with the default MIDI instrument.

6. SOME STATISTICS

SymbTr has been created mainly for the purpose of education and scientific research, and hence, endeavored to be as rich as possible in the diversity of *makams*, *forms*,

Nr.	Makams	# of Pieces	Usuls	# of Pieces	Forms	# of Pieces
1	Hicaz	118	Sofyan	251	Şarkı	677
2	Rast	88	Aksak	246	Türkü	285
3	Nihavent	85	Düyek	143	Seyir	169
4	Uşşak	85	Aksaksemai	101	Küpe	120
5	Segah	74	Curcuna	91	Peşrev	74
6	Hüseyni	72	Ağıraksak	83	Aranağme	72
7	Hüzzam	65	Yürüksemai	75	Sazsemaisi	66
8	Mahur	54	Nimsofyan	74	İlahi	32
9	Kürdilihicazkar	51	Semai	69	Yürüksemai	27
10	Muhayyer	51	Senginsemai	54	Beste	23

Table 2. The most used 10 makams, usuls, and forms in SymbTr

Nr	AE Name	AE Code	Quantity %	Duration %
1	Neva	D5	16.1%	16.1%
2	Çargah	C5	11.0%	10.7%
3	Hüseyni	E5	9.4%	9.7%
4	Gerdaniye	G5	8.5%	9.1%
5	Dügah	A4	8.5%	7.9%
6	Segah	B4b1	6.9%	6.6%
7	Acem	F5	5.4%	5.6%
8	Muhayyer	A5	4.8%	5.3%
9	Eviç	F5#4	4.7%	5.0%
10	Rast	G4	4.0%	3.6%
11	Nimhicaz	C4#4	2.9%	2.8%
12	Hisar	E5b4	1.9%	1.9%
13	Kürdi	B4b5	1.8%	1.7%
14	Dikkürdi	B4b4	1.6%	1.5%
15	Buselik	B4	1.4%	1.4%
16	Nimhisar	D5#4	1.2%	1.2%
17	Dikhisar	E5b1	1.0%	1.0%

Table 3. The most commonly used 17 pitches

usuls, and so on. There are many examples such as seyir composed for educational purposes. One criterion in the selection of pieces has been music lovers' familiarity with them, as to whether a piece be average or above-average. We did not adopt random sampling (as in [2], [11], and [15]) as proper methodology when one considers 80% of the twenty five thousand pieces in the TRT repertoire have hardly ever been performed or have become obsolete. A musical piece, composed but almost never performed cannot be held equivalent to one widely known and frequently performed.

Some statistics about SymbTr as follows:

Total number of pieces: 1 700 Number of notes: ~ 630 000

Classical: 1 400 Folk: 300

Vocal pieces: 1 295 Instrumental pieces: 405

Religious: 49

The number of distinct *makams*: 155 The number of distinct *usuls*: 100 The number of distinct *forms*: 48.

Interval (Hc)	Name, Direction	%
-9	Whole Tone (Tanini), descending	18.1
0	Unison	15.4
-5	Apotome (Küçük Mücennep), desc.	12.5
9	Whole Tone (Tanini)	11.4
5	Apotome (Küçük Mücennep)	8.6
-4	Limma (<i>Bakıyye</i>), desc.	6.3
-8	Minor Whole Tone (B. Mücennep),	4.2
4	Limma (Bakıyye)	4.1
8	Minor Whole Tone (Büyük Mücennep)	2.6
13	Augmented Second	2.1
-12	Augmented Second, desc.	2.0
-13	Augmented Second, desc.	1.8
22	Perfect Fourth	1.6

Table 4. The most commonly used 13 AE intervals

Highest ranking 10 makams, usuls, and forms are shown in Table 2.

7. PITCHES AND INTERVALS

Of all the pitches in the database, 17 that are used over 1 per cent in quantity and duration are listed in descending order in Table 3. Percentages in quantity and duration exhibit slight variances but these do not affect the ranking.

Figure 4 shows a histogram of these pitches in the two octave range between *yegah* (D4: 274) - *tizneva* (D6: 380) using the note codes as given in Table 3.

It is interesting to note that 9 pitches in the 3 octave range (Fig. 3) have never been used (for example, *kabahicaz*, and *kabadikhicaz*). When we excluded the notes that were heard for less than one thousandth of the time, only 33 pitches remained, whereas there were 72 pitches defined in this range in the AE tone-system. These observations seem to support Can's results [6].

The most commonly used 13 AE intervals and their usage as quantity in percentages are listed in Table 4.

The SymbTr database can be accessed at the following address, open for public consumption:

http://compmusic.upf.edu

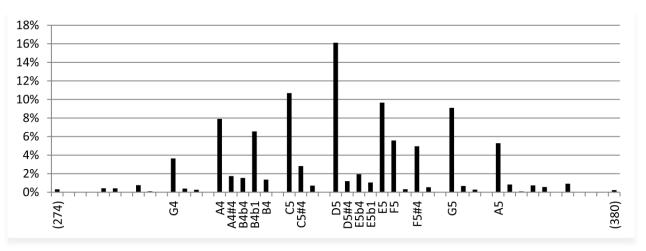


Figure 4. Usage of the notes in *SymbTr* as durations in percentages

8. SIMILAR DATASETS

In this article, we announce the availability of a new database called *SymbTr*, the most extensive machine readable database for Turkish makam music currently available. There is only one other compilation that would qualify to be called a database: the recently launched *TSM Corpus* [2] (TÜBİTAK ¹ ref. is PN: 110K040) consisting of symbolic data that relate to 600 pieces. These two databases are far from adequately representing Turkish makam music. New data, however, is being continually added to the *SymbTr* database through various projects. In addition, *Mus2* (Turkish makam and microtonal music notation program) ², which is still being marketed commercially, can produce output in the *SymbTr* format.

TSM Corpus project, supported by TÜBITAK, can be quite useful. However, the following deficiencies in database design need to be resolved:

- Presence of data belonging to various pieces in a single Excel format file makes usage difficult,
- Syllabized lyrics are not included in the database,
- Tempo information for musical pieces is not provided. Only one quantization information is included concerning durations: 1/4 meter note = 100 units. This is a serious drawback for musical pieces that require, in particular, the inclusion of tempo and / or usul modulations throughout,
- It is not specified which engraved score variant is employed when entering symbolic data.

9. DISCUSSION

If MIR community members at large run their applications on the *SymbTr* database, making necessary small changes, it may lead to two-way improvements: Mysteries of makam music may be unraveled on a grand scale, at a global setting while scholars keep tapping into new structures and patterns, thus moving into uncharted territories of human cognition.

10. ACKNOWLEDGEMENT

This research was partly funded by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

11. REFERENCES

- [1] C. Akkoç, "Non-deterministic scales used in traditional Turkish music", Journal of New Music Research, vol. 31, no. 4, 2002.
- [2] N.B. Atalay, S. Yöre: Türk Sanat Müziği derlemi [Data File]. Retrieved from <u>www.tsmderlemi.com</u>, Konya, 2011.
- [3] G. Ay, L.B. Akkal: İTÜ Türk Musikisi Devlet Konservatuarı Türk müziğinde uygulama - Kuram

sorunları ve çözümleri - Uluslararası çağrılı kongre bildiriler kitabı, İstanbul, 2009.

- [4] R. Ayangil: Türk makam müziği perdelerini çalabilen piyano imali projesi. Yıldız Teknik Üniversitesi, İstanbul, 1998.
- [5] B. Bozkurt, O. Yarman, M.K. Karaosmanoğlu, C. Akkoç: "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, 2009.
- [6] M.C. Can: "Geleneksel Türk Sanat Müziğinde Arel-Ezgi-Uzdilek Ses Sistemi ve Uygulamada Kullanılmayan Bazı Perdeler", Gazi Üniversitesi, Gazi Eğitim Fakültesi Dergisi, Ankara, 2001.
- [7] M.C. Can: Müzikte bilgisayar destekli istatistiksel analiz, Ankara, 2004.
- [8] H.T. Cheng et al. "Multimodal Structure Segmentation and Analysis of Music Using Audio and Textual Information". EEE International Symposium, 2009.
- [9] A. Holzapfel, Y. Stylianou: "Rhytmic Similarity in Traditional Turkish Music". *ISMIR 2010* http://www.csd.uoc.gr/~hannover/MMILab-Andre_files/HolzapfelIsmir09.pdf.
- [10] Y.F. Kutluğ: *Türk musikisinde makamlar*. YKY, İstanbul, 2000.
- [11] N.O. Levendoğlu: "Türk Müziğinde Rast Makamının İstatistiksel Değerlendirmesi". Selçuk Üniversitesi Fen Bilimleri Enstitüsü basılmamış yüksek lisans tezi, Konya, 1996.
- [12] G. Oransay: "Das Tonsystem Der Türkei Türkischen Kunstmusik", Musicforschung, Germany, 1959.
- [13] İ.H. Özkan: Türk musikisi nazariyatı ve usulleri Kudüm velveleleri. Ötüken Neşriyat, İstanbul, 1998.
- [14] K.L. Signell: *Makam Modal practice in Turkish art music.* Da Capo Press, 1986.
- [15] H.T. Sümbüllü, A. Albuz: "Türk Sanat Müziği dizilerinin, bilgisayar destekli makamsal analizi", *Uluslararası İnsan Bilimleri Dergisi,* cilt 8, sayı 1, 2011.
- [16] A. Töre ve E. Karadeniz: Türk musikisi nazariye ve esasları, T. İş Bankası yayınları, İstanbul, 1965.
- [17] Y. Tura: Türk musikisinin meseleleri. Pan Yayıncılık, İstanbul, 1988.
- [18] E. Ünal, B. Bozkurt, M.K. Karaosmanoğlu, "N-gram based Statistical Makam Detection on Makam Music in Turkey using Symbolic Data", submitted to ISMIR, 2012.
- [19] O. Yarman: "79-tone Tuning & Theory for Turkish Maqam Music", *PhD Thesis, Istanbul Technical University, Social Sciences Inst.*, İstanbul, 2007.
- [20] R. W. Young: "Terminology for Logarithmic Frequency Units". The Journal of the Acoustical Society of America 11 (1), 1939.

¹ The Scientific and Technological Research Council of Turkey

² www.mus2.com.tr