Experiments on the Relationship between Perde and Seyir in Turkish Makam Music

April 5, 2014

Can Akkoç
Institute of Applied Mathematics
The Middle East Technical University (METU)
Ankara, Turkey
can1936@gmail.com

William A. Sethares
Department of Electrical and Computer Engineering
University of Wisconsin
Madison, USA
sethares@ece.wisc.edu
ph: 608-262-5669
http://eceserv0.ece.wisc.edu/~sethares/

M. Kemal Karaosmanoğlu
Istanbul Technical University Turkish Music State Conservatory
Maçka, Istanbul, Turkey
karaosmanoglum@itu.edu.tr

Abstract

"When I was a kid, the elders in the village could tell the makam of a piece just by listening." While interviewing performers, enthusiasts, and experts in traditional Turkish *taksims* (improvisations), variations of this comment were made many times. Some of the respondents claimed to be able to identify the makam of a taksim, but others believed that this ability might now be a lost art. This paper documents a series of experiments (based on caricaturized or skeletonized taksim-like creations) designed to determine if it is possible to identify the makam from purely acoustical features, and, when possible, to determine the relative importance of the various audible features that may be used to establish the makam. Two basic classes of features are investigated: *perde* (the set of pitches used in the performance) and *seyir* (which relates to temporal motion within the piece, for instance, repetitive or common motives or melodic contour). The experiments provide evidence that both kinds of features contribute to the ability to recognize makams. Experiments that randomize the order of events show that pitch cues (perde) are often adequate to allow accurate identification of the makam. In experiments where both pitch and temporal cues are present but conflict (for example, a piece in which the perde is chosen from one makam and the seyir from another), experts often favor the temporal information.

Keywords: Turkish makam recognition, AEU system, recognition of scale and key by experts, motif in Turkish music, taksim (improvisation), seyir (melodic progression, temporal paradigm), perde (pitch paradigm)

Introduction

Makam in Turkish traditional music refers to a style in which each makam-type names a pitch structure (called *perde*) and/or specific patterns of motivic/temporal development (called *seyir*). When hearing a makam performance, a listener will typically have access to more information than just the sound: the performer may be known, the particular piece may be familiar, it may be recognized from a specific place or previous time. Thus the "elder" in the "village" of the introductory quote may indeed correctly recognize the makam, but this identification might stem from extra-musical information and not necessarily directly related from the acoustic structure of the sound itself. As will be shown in Experiment 1, such extra-musical information is not necessary for correct identification; expert listeners can reliably determine the makam from audible features alone. This leads to the second major question that is explored in Experiments 2 through 4: which features of the performance are key to the ability of an expert listener to identify the makam? The experiments are structured so as to investigate the relative importance of the perde (pitches) and the seyir (temporal motion) in the task of identifying the makam of a piece.

Background

Turkish makam music is primarily an oral tradition taught on a single instrument using a *ney* (an end blown cane flute), a *kemençe* (a bowed stringed instrument), the voice (*hanende*), or other traditional instrument. Learning typically occurs in a master-apprentice setting through extended repetition (*meşk*). There are (at least) two major kinds of pieces: composed (*beste*) and improvised (*taksim*). Popular makam music is typically heterophonic, rhythmic, and performed with percussion accompaniment; improvised forms are typically solo performances played with a flowing and relatively unstructured rhythm. A makam can be viewed as a musical setting with (a) a well-defined underlying scale of pitches, and (b) a set of conventions that structure the temporal ordering of the pitches into melodic lines.

The standard theoretical explanation for Turkish pitch sets is given by the "national theory of Turkish Music" called the Arel-Ezgi-Uzdilek System (AEU) (Arel, 1968;

Ezgi, 1933), which can be viewed as a 24-note set constructed from Pythagorean commas, and which can be closely approximated by notes from the 53-tone-equal tempered system (Yarman, 2008). Empirical perde scales (in contrast to scales derived from theoretical considerations), have recently been investigated by taking pitch measurements on performances from renowned masters (Bozkurt, Yarman, Karaosmanoğlu & Akkoç, 2009; Akkoç, 2002). For example, Figure 1 shows the similarities and disparities between the theoretically-derived scale pitches and those measured in performance.

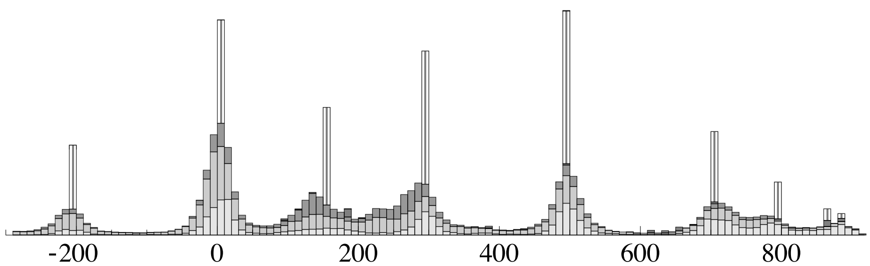

Figure 1: This cumulative histogram displays pitch clusters in eight taksims (improvisations) in the Uşşak makam and contrasts these with the theoretical pitches shown by the tall vertical lines. The horizontal axis is given in cents with respect to the root at zero, the vertical axis is the percentage of time spent on each pitch. Shades of gray correspond to different octaves.

While there remains controversy about the accuracy and applicability of the AEU classification to makam performances, both proponents and critics of the AEU system agree that it is based primarily on pitch relationships. Within the AEU system, more than a hundred makams have been described and labeled as in Figure 2, which are drawn from Karaosmanoğlu et. al. (2009). The music-theoretical information in Figure 2 is complemented by a numerical representation in Appendix A that is used to quantify the interval-set relations between the various makams. Recent work (Bozkurt, Gedik, Savacı, Karaosmanoğlu & Üzbek, 2010) suggests that automated classification of makams can be accomplished with considerable accuracy using only pitch-histogram information. In terms of the introductory quote, it is reasonable to posit that it may be possible to recognize the makam of a piece purely by listening carefully to the pitch relationships within the performance. Since perde is the Turkish word for tones or pitch clusters, we call this the *perde hypothesis for makam recognition*.

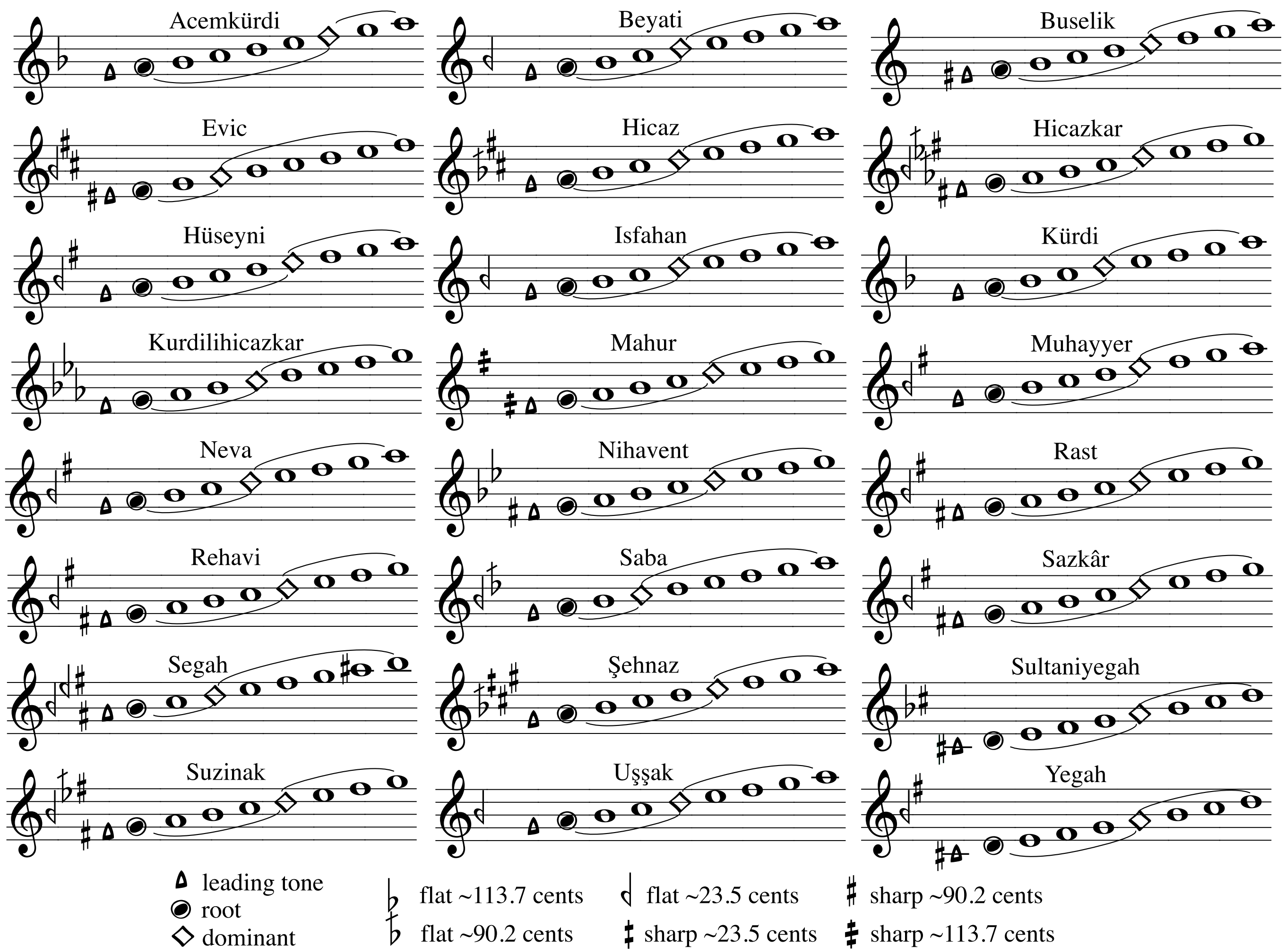

Figure 2: Twenty-four makams that appear in this paper are represented in the AEU system. Special symbols are used to indicate the root, the dominant, and the leading tone. Many of the accidentals represent pitches that fall between the cracks of the Western musical scale. Makams may be characterized as the conjunction of an n-chord and an m-chord (where $n + m = 9$), for example, a tetra-chord and a penta-chord, as indicated by the grouping markers that connect the root and the dominant, and the dominant with the upper octave of the root). These samples are drawn from the software Mus2okur (Karaosmanoğlu et. al., 2009), which lists basic makams as well as hybrid combinations (or transposed forms) in commonly used genus patterns. These interval sets are shown numerically in Table A1.

Others argue that the seyir (a complex of stereotyped motives, melodic signatures, and latent melodic possibilities) associated with each makam is crucial to its identity. Akkoç (2008) views seyir as a "journey between pitch clusters forming the underlying scale of the host makam." Castellano, Bharucha & Krumhansl (1984) observe that this is analogous to an Indian rag, which is characterized both by its scale and by the manner and order in which the scale tones are combined. Thus the rag in Indian music forms a basis on which melodic composition and improvisation may occur. Similarly, the Turkish form exploits sets of scale tones and collections of melodic figures, motives, and patterns that utilize those scale tones. For example, Beken & Signell (2006) state "Every instrumentalist chooses from among confirming, delaying, and deceptive elements to create a taksim in a given makam. These elements may include general melodic direction, certain intervallic relationships, modulations, and most importantly, cadential points, often coming after stereotyped motives." Similarly, the New Grove Dictionary (2013) describes seyir as indicating "prescribed modulations and the general shape of phrases, understood as either predominantly upwards, predominantly downwards, or a combination of both." The New Grove article on mode credits Kantemir (1700) with defining a term for melodic progression, and comments that "the seyir spans a framework of tonal centers" that includes the root, the octave above the root, and other key tones.

Conceiving of seyir as central to the notion of makam subordinates pitch elements to temporal elements such as motives, patterns, and melodic contour. A recent study of related Arabic improvisations (Ayari & McAdams, 2003) shows "the melodic reductions of segments in a given maqām reveal the nature of Arabic modes as involving not just a tuning system, but also essential melodico-rhythmic configurations that are emblematic of the maqām." Confirming this view, our experts were asked (at the conclusion of the experiments) what they listen for when trying to determine the makam. A typical response was that of Expert 2 who said, "I listen for certain sound patterns, like this (singing) in Uşşak or like this (singing) in Hicaz." Tanrıkorur (2005) states that makam encompasses both a pitch structure and (unwritten) rules for melodic progression that must be strictly observed by composers, and memorably states that a makam is 20% pitch and 80% seyir. We call this melodic or motivic notion that emphasizes the temporal aspects of a makam the *seyir hypothesis for makam recognition*.

Related Literature

The domain of music psychology has only recently begun to consider experimental work on the perception of Middle Eastern music, and such studies are limited. Ayari & McAdams (2003) consider a segmentation paradigm that contrasts Arab listeners and European listeners, and investigates the ways in which these two groups partition melodic passages within a makam structure. Thus they are able to investigate cultural differences in listening strategies. Our tests are quite different since they require identification of the makam directly from the audio, a task that those unfamiliar with the genre are unable to accomplish. Indeed, Ayari & McAdams (2003) observed that some of their Arab musicians were able to detect and describe modal structures such as the makam of melodic passages. It is this ability, within the context of Turkish makam music, that is the primary focus of the present paper.

Castellano et. al. (1984) define *tonality* as the "centering of the musical materials around a particular tone," and thus tonality provides one way of structuring and organizing sound materials in music. For example, in Western music, the tonic is a reference tone associated with the musical structure called the *key*. More generally, the tonic provides a reference point whereby a set of musical pitches may be perceived in relationship to that tonic. Bharucha (1984) comments that "considerable exposure to pieces of music that are structurally similar gives rise to the... tonal hierarchy" and suggests that the temporal ordering of tones may be able to activate tonal schemas that are held in long-term memory, a view that is at least partly supported by Boltz (1989). In this general sense, Castellano et. al. (1984) provide evidence for the tonal perception in classical Indian music using a probe-tone technique that compares the responses of Western and Indian listeners. In Turkish makam music, the tonic may be identified with the root tone (as in Figure 2) and the hierarchy would include an n-chord and m-chord (for example, a *tetra-chord* and a *penta-chord*) as indicated for particular makams by the grouping markers in Figure 2. The tonal hierarchy is also evidenced by the most important satellite tones such as the *leading tone* and the *dominant* (in Turkish music, the dominant may be a musical fifth, but it may also be another interval; it is characterized as the pivot tone that is common to the n- and m-chords).

Using the probe-tone technique of Krumhansl & Shepard (1979), Oram & Cuddy (1995) addressed listeners' responses to pitch distributional information in melodic sequences, concluding that more frequent tones (or those with greater total duration) tend to be rated as fitting the melodic context better than less frequent tones. Both musically trained and untrained subjects were responsive to distributional information whether the context was a diatonic subset of 12-tone equal temperament or whether it was a nondiatonic subset, although greater familiarity with tonal material made the task somewhat easier. Similarly, using probe-tone studies inspired by the music of Northern India, Castellano et. al. (1984) found that the most frequent tones were rated as best fitting the melodic context. The most common tones in Turkish makam music tend to be aligned with the most important tones in the pitch hierarchy (the root and the dominant). The melodic motion tends to be characterized by the establishment of a pitch center with many small deviations about that center, followed by a motion to another pitch center with deviations, and so on. Accordingly, the perceptual recognition of a makam may be due to the pitch set itself (and the corresponding frequencies of occurrence of the pitches), although it may also be due to the temporal arrangement of the pitches. This dichotomy nicely parallels the distinction between the perde hypothesis and the seyir hypothesis that the experiments of this paper are designed to untangle.

Overview

Experiment 1 establishes that expert listeners can indeed pinpoint the makam from acoustic clues alone (since the sound examples contain no extra-musical information such as performer or location, and are presented with a uniform synthetic timbre). Experiments 2 through 4 were then designed in an attempt to distinguish the perde hypothesis from the seyir hypothesis. A key technical element in these experiments is the generation of sound examples where the two aspects of the makams (perde and seyir) can be separately controlled, combined, and resynthesized. A series of synthetic taksims (or caricatures) allow precise manipulations of the sound, emphasizing and distorting the perde and seyir until they become unrecognizable, so that potential outer limits of these two interactive core elements can be ascertained, as much as possible. Construction of the sound examples is described in the stimulus section of each experiment with further technical details in the Appendices.

A small number of experts were asked to participate in the experiments. Each is a recognized master of Turkish makam music, and brief biographies are presented in the section on participants. The various experiments were conducted at intervals of about one week (according to the availability of the experts). The experts were not all available simultaneously, so the (roughly) four weeks of the experiments were staggered over the months of October and November 2012. From the expert's point of view, each experiment is simple. They receive an email (or disk) containing a set of six to eight short sound examples. The duration of the examples is between one and three minutes, depending on the particular goals of that example. The listener is asked to state what makam (if any) is represented by the examples. The experts may listen to the examples as many (or as few) times as they wish, and return their evaluations by email. The experts were not given any hints about which makams might be used (so any of the roughly 100 makams conceivably could have appeared in any of the sound examples). After each experiment, the experts were eager to find out more information, and we discussed the sound examples with them in order to encourage their continued participation. We did not disclose details of how the sound examples were constructed, and none of the experts had contact with each other over the duration of the experiments.

Each of the four experiments was designed to test the makam-identification problem in a different way. In many cases, the experts agree in their assessments. In situations where there is disagreement, it is usually possible to look carefully at the way the experiment was constructed to reveal plausible reasons for the ambiguity. The discussion considers the results of all four experiments together and tries to draw conclusions based on all of the experts' responses. Stripping away all appropriate caveats, the major results can be summarized as follows: in the absence of temporal cues, pitch cues are often enough to allow accurate identification of the makam. If both pitch and temporal cues are present but conflict (for example, a piece in which the pitches are chosen from one makam and the temporal information from another), experts often responded most strongly to the temporal cues.

Experiment 1: Establishing a Baseline

As commonly performed, a taksim has three sections: an exposition that establishes the host makam, a middle section that potentially explores other makams, and a recapitulation that restates the host makam, (see Beken & Signell, 2006). Thus it is not realistic to expect that every segment within a given performance lies fully within the host makam. In 1986, while visiting Worcester Polytechnic Institute, the first author asked noted *ney* (an end-blown flute) performer Niyazi Sayın to record a collection of short "skeleton" taksims that remained fully within the host makam. The result was seventeen improvisations in makams Buselik, Eviç, Hicaz, Hicazkar, Hüseyni, Isfahan, Mahur, Muhayyer, Neva, Nihavent, Rast, Saba, Segah, Sultaniyegah, Suzinak, Uşşak, and Yegah. These may be heard in their original form on the Makam Experiment Website (Sethares, 2014) and form the core of the data used throughout these experiments. All of the sound examples from the experiments, as well as intermediate files such as the MIDI data files, are also available. Niyazi Sayın has also contributed to this work by acting as one of the experts.

Procedure

In order to create the synthetic taksims for the experiments, the seventeen improvisations were analyzed and transformed into MIDI files where they could be more easily manipulated. Steps in this transformation are detailed in the stimulus section. The MIDI files were then rendered back into audio using a synthetic "bamboo flute"-like sound that is reminiscent of the ney but without the breathy sounds. Six of these resynthesized taksims were chosen to be the sound examples for Experiment 1: Hicaz, Isfahan, Muhayyer, Neva, Rast, and Uşşak. These six soundfiles were renamed (so as to hide their origins) and the order was randomized (so that the soundfiles, when viewed by computer, would not appear in alphabetical order). They were emailed to the experts who were asked one question:

Q: "What makam, if any, is each piece performed in?"

The experts were given no hints as to the reason for the question (other than a very general "we are studying the makeup of makams and taksims") and no hints as to the origin of the pieces. In particular, they were not told that the "intended" answer would be from among the seventeen taksims recorded by Niyazi Sayın; indeed, there are over 100 different makams catalogued in standard references (Arel, 1968; Ezgi, 1933), so the range of possible answers is large.

Experiment 1 may appear simple, but it immediately confronts the question of whether experts are able to recognize a makam using auditory clues alone. Moreover, it addresses several issues that are crucial for successive experiments. To the extent that the makam remains recognizable, it shows that the process of transforming the original improvisation into MIDI and then back into sound does no harm. For example, it shows that the timbre of the instrumental voice is not crucial (since the original is a breathy end-blown flute while the MIDI rendering is a relatively simple synthesized flute). It also shows that the removal of the pitch glides from the original does not harm the recognizability of the makam. It shows that whatever errors may occur in the pitch tracking algorithms are not crucial. In short, it justifies asking the same question **Q** when more substantial changes are made to the improvisations, justifying the experimental paradigm followed throughout the remaining experiments.

Participants

The listening tests described in the experiments are quite different from tests with "naive" subjects where the aim is to understand the perceptions and abilities of normal listeners. In this case, naive listeners (including most of the authors) cannot correctly name the makam of a piece by listening to it. Accordingly, it was necessary to ask experts. Such experts are not easy to find, and so it was not possible to run the experiments on a large number of subjects (as would be required for a statistical analysis). In some cases, potential experts were deterred from participating out of worries that the experiments were a test of their expertise, or a competition amongst the various masters of the genre.

The authors were fortunate to find three experts who were willing to devote time and energy to listening to the sound examples and answering the (apparently) simple question **Q**. All three have given their permission to be named here as participants. Once engaged, the experts appeared to enjoy the task, and clearly took it seriously. The experts were often confident of their assessments, although some worried in their email responses whether they "got the answers right." One said the experiments of the final test set "were designed to torture. Everything in them was tinkered with, and designed to fool and deceive" indicating that the tasks were not always easy.

This section presents basic biographical information about the experts.

Expert 1: Niyazi Sayın

Neyzen Niyazi Sayın (1932-) is a legendary representative of the 5th generation of master musicians following in the tradition of Hammamizade İsmail Dede (1778-1846). According to Holtzberg (2008), he is regarded as one of the most important living ney players in Turkish classical music. His musical geneology can be traced through his teacher, the painter and ney player Halil Dikmen (1906-1964), who was the student of Ahmet Irsoy (1869-1943), who was the son and student of Zekai Dede (1825-1897), who was the student of Hammamizade İsmail Dede. Sayın's meşk, which Gill-Gürtan (2011) describes as the practice of music transmission, is an oral tradition that places him in a long line of masters who perform makam compositions and improvisations, deal with the culture surrounding the musical performances, engage in arduous training in both listening and performance, and feel a strong sense of a social identity and responsibility for the preservation and continued transmission of the style.

Expert 2: Ruhi Ayangil

Ruhi Ayangil (1953-) started playing Kanun at the age of 10. He graduated from the Faculty of Law at Istanbul University in 1979 and was a student of İhsan Balkır at the Istanbul Municipal Conservatory where he studied harmony and orchestration with Cemal Reşit Rey. Between 1973 and 1981 he trained and conducted the chorus of Robert College, Istanbul and taught courses in Turkish music. Ayangil's book "Learning To Play the Kanun," based on Alnar's technique, provided a basis for lectures while a member of the faculty at the Istanbul University State Conservatory. In 1988 Ayangil's Turkish Music Orchestra and Chorus made the first recording of "Uyan Ey Gözlerim" (Ottoman Sufi Music), compiled by Ali Ufki (1610-1675). Ayangil was awarded the title of "artist of the year" by the Turkish Writers Association for this recording, and he has received several awards for his research into the roots of Turkish music and its evolution. He has now retired as Dean of the Faculty of Art and Design, Istanbul Yıldız Technical University.

Expert 3: Necdet Yaşar

Necdet Yaşar (1930-) is a tanbur (lute) player, music teacher, and a founding member of the Istanbul State Turkish Music Ensemble. In 1991 he was awarded the title "National Artist" by the Turkish government, and Signell (2011) notes that he is a leading tanbur player who has performed classical Turkish music throughout the world. Yaşar was the pupil of Mesut Cemil, son of the legendary Tanburî Cemil Bey. According to Aksoy (2005), Yaşar is a master at avoiding stereotyped musical phrases in his original improvisations, a "composer of improvisations, a poet of the tanbur, who recites makamic verses." Yaşar was also a primary source in Signell (1986), and an illustrated biography has recently been published (Tokuz, 2009).

Stimuli

The sound examples in Experiment 1 (and throughout later experiments) rely on a kind of analysis-resynthesis method; the audio .wav files in a collection of taksims are analyzed and transformed into MIDI note-level representations where the pitch and timing can be straightforwardly manipulated. The process begins with the pitch detector of de Cheveigné & Kawahara (2002), which is used to estimate the instantaneous pitch of the original performances at a rate of 100 times per second. The raw pitch data are converted into estimates of the pitch centers and transition probabilities using the

Makam	Expert 1	Expert 2	Expert 3
Hicaz	Hicaz	Hicaz	Hicaz
Isfahan	Hüseyni	Isfahan/Uşşak	Isfahan/Uşşak/Neva
Muhayyer	Muhayyer	Muhayyer	Muhayyer
Neva	Uşşak/Neva	Bayati/Neva	Neva
Rast	Rast	Rast	Rast
Uşşak	Uşşak	Uşşak	Uşşak

Table 1: Results for Experiment 1 "Establishing a Baseline"

Expectation Maximization (EM) algorithm of Welch (2003). These are then used to estimate the note-start and note-end times via the Viterbi (1967) algorithm. As these pitch extraction steps are somewhat involved, the steps are detailed in Appendix B.

The output of the above processes is a set of note-level data that are translated into a standard MIDI file. The sound examples of Experiment 1 are a resynthesis of the analyzed performances using a flute-like instrumental sound generated using the Alchemy additive synthesizer by Camel Audio (2013). The sound patch is amplitude-modulated at a slow rate, imitating (somewhat) the vibrato and timbre of the ney, although without the breathy effect common with the ney. The performance is quantized to eight notes (per octave) and the specific pitches are determined by the pitches present in the original performances. While this leaves much of the melodic motion intact, it removes pitch glides and microtonal ornamentations. The original sound files, the extracted MIDI files, and the resynthesized versions can all be heard at the Makam Experiment Website (Sethares, 2014).

Results

Experiment 1 consists of six sound examples that are essentially resynthesized versions of the original taksims from the 1986 performances. The taksims chosen for resynthesis are listed in Table 1, along with the responses of the experts. The specific taksims chosen maintain a balance between those which are more and less common.

All three experts correctly identified four of the synthetic taksims precisely (Hicaz, Muhayyer, Rast, and Uşşak). Isfahan and Neva are comparatively uncommon, and it makes sense that rarer makams would be more difficult to recognize. Even so, Expert

3 identified Neva correctly while Experts 1 and 2 mentioned Neva in conjunction with another more common makam (Uşşak and Beyati). To understand this, recall that in "normal" performances, a piece will begin on a host makam, modulate through other makams, and then resolve back to the host. In Table 1, there are four places where the experts identified more than one makam (those with a slash /). When we asked about this, Expert 2 said that he "heard echoes of both" in the piece, and could not decide which is the most prominent (and so listed both). Figure 2 shows that Neva and Beyati have the same dominant and tonic, and otherwise differ by a single sharp in the key signature of the AEU representation. Similarly, Neva and Uşşak differ by the same sharp and also have identical dominants and tonics. Moreover, Beyati and Uşşak are indistinguishable from interval content alone in the classic AEU perspective, as shown in Figure A1. What this shows is that makams that are most closely related according to standard theory may be among the most readily confused in listening tests. Similarly, Isfahan and Uşşak are identical from the AEU perspective and are often considered to be among the makams that are "most alike." Ederer (2013) observes there are two kinds of Isfahan: the one used here that might be mistaken for Uşşak or Beyati is formally known as "Basit Isfahan" while younger musicians might be more familiar with "Mürekkeb Isfahan." The asymmetric confusion between Uşşak and Isfahan is likely due to the (relative) uncommonness of Isfahan.

These results do not mean that the synthetic taksims and the original improvisations sound in any sense "the same"; rather, it means that the transformations into the synthetic versions retain the essence of the makam structure (whatever that may be). This experiment shows that it is possible to recognize the makam from a synthetic version, using only auditory clues. There is no real possibility that such a string of matches could have resulted from chance: with over 100 possibilities for each of the sound examples, even getting one correct in six tries would be highly unlikely. Although we only asked the single question (which could in principle be answered by a single word) the expert's responses were rarely so succinct. For the most part (in this and in subsequent experiments), the experts indicated that they listened to the sound examples several times and in several cases, provided second-by-second analyses of the pieces. These would typically end with "and therefore I conclude it is in makam X."

κ	Agreement
< 0	poor
$0 - 0.2$	slight
$0.2 - 0.4$	fair
$0.4 - 0.6$	moderate
$0.6 - 0.8$	substantial
$0.8 - 1$	almost perfect

Table 2: Interpretation of κ values according to Landis & Koch (1977)

One way of quantifying the amount of agreement among subjects is the kappa coefficient (Fleiss, 1971), which can be applied to multiple raters on categorical data (Sim & Wright, 2005). The kappa value $\kappa = \frac{\bar{P}-R}{1-R}$ indicates the "proportion of agreement beyond that expected by chance," where $\bar{P}$ is the observed agreement and R is the agreement expected by chance alone. κ lies on a scale between -1 (complete disagreement) through 0 (chance agreement only) to $+1$ (complete agreement). A full discussion of the kappa calculation is presented in Appendix C. For Experiment 1, applied to the three experts, κ lies in the range $(0.671, 0.735)$, which can be interpreted as in Table 2 to reflect substantial agreement among the experts. It should be noted that there is no universally accepted interpretation of κ values. In our experiments, κ may be understated because it only takes into account the categories (the makams) that actually appear in one or more responses; the actual difficulty of the task is also dependent on the universe of possible answers (the 100 or so makams); this would tend to make κ a conservative estimate of the true agreement. Observe that the kappa coefficient does not have a well-accepted notion of statistical significance, and since our sample sizes are quite small, it is probably best to view the numbers as suggestions for interpretation rather than precise yardsticks.

We report the response of Expert 1 to Experiment 1 here for completeness; it should be noted that this expert was the original source of the seventeen taksims from 1986. Although he did not consciously recognize any of the sound examples in this experiment (after a lapse of more than 25 years), the possibility of the use of extra-acoustic information in the responses cannot be ruled out. Accordingly, it may be best to discount Expert 1's responses in this experiment. This caveat does not apply to Experiments 2-4 where the source material was manipulated and disguised. (Somewhat paradoxically, recalculating the kappa values with Expert 1 removed raises the value of κ slightly.)

Experiment 2: Scrambling Time

Experiment 1 establishes that the host makam can indeed be recognized from auditory features alone; it is possible to ask which features are crucial to this ability. The two primary candidates are the perde hypothesis and the seyir hypothesis, and Experiments 2 through 4 are aimed at narrowing the possibilities and uncovering the kernel or invariants that lead to recognizability of a host makam.

The perde hypothesis posits that makam identification is crucially dependent on pitch relationships while the seyir hypothesis posits that the identification is crucially dependent on time ordering. Motifs, sound patterns, and melodic contours are inherently ordered; sound X followed by sound Y is fundamentally different from sound Y followed by sound X. This is well studied in the case of melodies, where rearranging the temporal order of a melody can leave even a familiar melody unrecognizable (Deutsch, 1982). In contrast, the set of pitches present in a piece is invariant with respect to temporal rearrangements. Accordingly, the sound examples of Experiment 2 rearrange the order of the notes of the improvisations, leaving the pitches unchanged. This is done in two different ways: by scrambling segments, and by time reversal.

Procedure and Participants

Roughly one week elapsed between Experiments 1 and 2, and the same expert listeners were again asked question **Q**. The timing of each of the experiments was somewhat different for each expert due to scheduling constraints. Because the sound files were sent by email (and in one case on CD), the sound examples could not be "returned" after the experiment was over, and so the experiments must be considered to be cumulative. When we asked (at the end of all four experiments) if the experts had referred back to sound examples from previous experiments, all said 'no.' Because our subjects are experts who were donating their time and energy, we did not feel it was appropriate to answer direct questions with dissembling responses. Accordingly, we supplied feedback when asked. Expert 1 was the most persistent: asking questions about who was playing, what instrument was being played, what we had done to create the sound examples, and what the "right" answers were. Expert 2 was considerably more circumspect, more interested in "why" we were making the sound examples than in "how," and Expert 3 asked no questions. Our belief is that our responses supplied no useful information to influence future decisions, but we cannot say this with complete certainty. In Appendix C, we try to address this quantitatively.

Stimuli

The scrambling-by-segments method relies on the observation that the original improvisations are built from a number of small segments. These segments correspond to the points at which the performer breathes (these were performed on a solo ney) and so provide natural stopping and starting points for the temporal rearrangement. For example, the Hicazkar improvisation was performed in nine small segments. Numbering the segments sequentially, two resynthesized versions performed the segments in the orders 135724689 and 183754629. Thus both start and end the same, but move through the piece in different orders. This technique is analogous to the scrambling of melodies presented in (Rabinovitz, 2011).

The second method of rearrangement is time reversal. In these examples, the notes of the improvisation were performed backwards: first the final note, then the penultimate note, then the 3rd to last, and so on, all the way back to the first note. Although there is no musical score, the effect is the same as if the performer played the piece note-by-note from end to start. (This is not the same as reversing the soundfile, which drastically changes the timbre of the instrument.) Synthesized versions of Hicaz, Mahur, and Uşşak were time reversed in this fashion.

Results

With a single exception, the sound examples of Experiment 2 are not simple copies of an original performance; rather, the sounds are scrambled in time in a nontrivial

Sound Example	Expert 1	Expert 2	Expert 3
hicazkar183754629	Hicazkar/Şehnaz	Şehnaz/Hicaz	Hicaz/Şehnaz
hicazRev	Hicaz	Hicaz	Hicaz transposed
mahurLinear	Mahur	Mahur	Mahur
mahurRev	Mahur/Rast	Rast/Mahur/Rast	Mahur
hicazkar135724689	Nihavent	Şehnaz/Hicaz	Şehnaz
ussakRev	Acemkürdi	Uşşak	Rast

Table 3: Results for Experiment 2 "Scrambling Time"

manner. The two Hicazkar examples are scrambled by section while the three "Rev" examples (short for "time reversed") invert the temporal motion of the piece. The example labeled mahurLinear does not fit this pattern and is, instead, a relatively faithful rendering of the original Mahur taksim where the pitches were fit with a linear slope (as described in the stimuli of Experiment 1). This example conceptually belongs with Experiment 1, and all three experts correctly reported the makam. (We did not want to "give away" what we were testing in each set by having all the sound examples created in the same way, so we split the examples among the sound sets: mahurLinear logically belongs in Experiment 1, while the two RastXXX examples from Experiment 3 logically belong to Experiment 2. We report them here as the experiments were conducted.)

There is, as might be expected, more variation in the responses to the scrambled sound examples. For instance, all the experts correctly identified the time-reversed Hicaz (although one perceived it as a transposition of this makam), and all identified Mahur as a component of the time-reversed Mahur. Two experts also mentioned Rast in this example: the pitch content of Mahur and Rast differ by one comma flat and one comma sharp. The final time-reversed example was correctly identified by Expert 2 as Uşşak, but was heard by Expert 1 as Acemkürdi and by Expert 3 as a transposition of Rast. The scales of Acemkürdi and Uşşak differ by only one half-flat and in the particular performance used this note does not occur frequently. While the above explanation is probably clear to a Western reader, it should be noted that Turkish practitioners may conceive the similarity between Acemkürdi and Uşşak in terms of the root-position *cinses* of the two makams, (as noted by Ederer, 2011). In this case, Acemkürdi requires a Kürdi tetrachord, while the Uşşak makam requires an Uşşak tetrachord. Alternatively, Uşşak is often thought of as having an ascending character while Acemkürdi has a descending character. Thus these might naturally be confounded after a time reversal.

Of the two scramblings of Hicazkar, only Expert 1 pinpointed Hicazkar, and he also heard excursions into Şehnaz. Experts 2 and 3 also heard Şehnaz, but alternating with Hicaz. As the names imply, these two are closely related since Hicazkar is formed, according to classic AEU theory, by adding a Hicaz tetrachord (built on perde neva) to a Hicaz tetrachord (built on perde rast). Perhaps the most straightforward interpretation of this rests on the observation that the scales of Şehnaz and Hicazkar (in the AEU representation) are transpositions of each other. Ederer (2013) notes that "it is possible to perform very simple renditions of Şehnaz and Hicazkar in such a way that it would be hard to tell which was which without also knowing what the tonic is." Signell (1986) writes about the relationships between Hicazkar and Şehnaz and gives two extended examples (numbers 113 and 114) that demonstrate the extensive similarities and subtle differences. In terms of the interval sets, Şehnaz and Hicazkar are rotations of each other, as shown in Figure A1. In the exit interview, Expert 2 said that he had recognized the similarity between the first and fifth sound examples when answering, and he thought we had used the same twice, "perhaps to try and trick him." Thus he ensured the answer was the same for both. The outlier here is the response of Nihavent, for which we find no obvious explanation.

Calculating the kappa coefficient for this experiment as in Appendix C gives an agreement rating among the three experts of κ in the range $(0.263, 0.345)$. According to Table 2, this is a "fair" agreement among the experts. Because of the tight relationship between some of the makams used (as mentioned above) and because of the large number of possible answers that did not appear, this may understate the agreement.

Experiment 3: Randomizing Events

Another way of removing temporal relationships is via randomization. In the simplest situation, a histogram can be used to count how many times each perde occurs. A "new piece" can be built by picking notes at random with probabilities based on the histogram. In the output, notes occur with roughly the same frequency as in the original, but in different order. A less drastic randomization can be made by considering all pairs of notes and then choosing notes for the new piece based on the probabilities of the pairs. This would tend to replicate the original time ordering somewhat more faithfully. Continuing in this fashion, it is possible to consider triplets ($n = 3$), quadruplets ($n = 4$), etc. As n increases, the randomized output tends to more faithfully replicate the original. Said another way, the events in the input pieces are randomized in such a way so as to destroy long-term temporal structure but to preserve short-term temporal structure, where "long" and "short" are determined by n. A large database is needed in order to generate these probabilities; we used the source material from the "Turkish Makam Music Symbolic Database for Music Information Retrieval" (Karaosmanoğlu, 2012). Details of the procedure used to create the sound examples with this nth order Markov Chain method are presented below.

The sound examples of Experiment 3 include n=1 and n=3 randomizations of the Hicaz and Uşşak makams. The n=1 sound examples completely destroy all temporal relationships between the notes of the taksims while leaving the perdes and the histogram (approximately) intact. If the makams from these examples can still be recognized, this can be considered strong evidence for the perde hypothesis. The n=3 sound examples retain some of the temporal motion of the original (in particular, all length-three sequences in the output must occur somewhere in the input) so these retain more of the temporal character of the original taksims. Since these manipulations are made at the symbolic level (i.e., on the MIDI file) the randomization does not include fast features of the performance such as ornaments, pitch glides, and other intermittent note-level phenomena.

Procedure and Participants

The procedures and participants were the same as in Experiments 1-2.

Stimuli

Successive events in musical performances are not independent. Shannon (1948) suggests a way to model redundancies in text:

> . . . one opens a book at random and selects a letter on the page. This letter is recorded. The book is then opened to another page, and one reads until this letter is encountered. The succeeding letter is then recorded. Turning to another page, this second letter is searched for, and the succeeding letter recorded, etc.

There is nothing about Shannon's technique that is inherently limited to dealing with text sequences, and nothing that limits the technique to single letters. An implementation called "Poem Maker" that allows any number of sequential letters using text sources drawn from the Wolfram library of curated data has been written by Sethares (2011). With $n = 1$, the letters are effectively chosen randomly from the distribution of letters within the text. With $n = 2$, the letters are chosen from successive pairs; with $n = 3$, they are chosen from successive triplets, etc. The probabilities of clusters of letters are defined implicitly by the choice of the source text.

By considering a piece of music as a sequence of symbols, Shannon's book can be replaced by a suitable corpus of music. The Turkish makam database SymbTr (Karaosmanoğlu, 2012) provides a suitable collection of pieces classified by makam. Accordingly, the text-based "Poem Maker" was translated into a MIDI-based sequence generator. Instead of generating text based on n-term probabilities, the MIDI generator creates sequences of notes where n-note patterns occur with probabilities specified by the source collection. Thus for $n = 1$, individual notes occur with the same probabilities as in the makams of the SymbTr database. For $n = 2$, pairs of notes occur with the same probabilities as in the database, etc. The randomized synthetic taksims of Experiment 3 were generated in this manner, and then realized using the same simulated-ney sound as in Experiment 1.

Sound Example	Expert 1	Expert 2	Expert 3
rast13572468	Rast	Sazkâr	Rast
hicazRand3	Hicaz	Hicaz	Hicaz
ussakRand1	Beyati	Beyati/Acemkürdi	None
rast17654328	Rast	Rast	Rast
hicazRand1	Hicaz	Hicaz	Hicaz
ussakRand3	Uşşak	Hüseyni	None

Table 4: Results for Experiment 3 “Randomizing Events”

Results

Experiment 3 again consists of two “kinds” of sound examples. Logically, the two rastXXX examples belong with the scrambling examples from Experiment 2. The new technique is embodied in the examples with the “Rand” suffix, which indicates that these were created using the Markov chain randomization technique described above.

Tura (1988) comments that Rast is the root of all makams, so it might be anticipated that it would be among the easiest to recognize, even in scrambled form. Indeed, all three of the experts identified the scramblings of the Rast makam, although Expert 2 identified a more complex structure that lies “on a Rast scale with plenty of Segah, most like Sazkâr” (we have abbreviated this in the table). In terms of interval sets, Sazkâr and Rehavi are both the same as Rast, as shown in Figure A1. Hence this answer is quite reasonable.

The Hicaz and Uşşak makams were randomized according to the Markov chain method with n=1 and n=3, where n is the length (or memory) of the chain. The n=1 examples have no memory (pedantically, a memory of one note) and are effectively the same as if one generated notes at random from probabilities dictated by the histogram. The n=3 examples have a memory of triplets, that is, one-, two-, and three-note sequences will occur with the same probabilities as in the original database. Both Hicaz randomizations were correctly identified by all three experts. The Uşşak randomization with n=1 was identified as Beyati by two of the experts. This is easy to understand since the Uşşak and Beyati makams have the same set of pitches; indeed, some authors such as (Arel, 1968) do not consider these to be distinct makams. In terms of the interval sets of Figure A1, the distance between Uşşak and Beyati is zero. Signell (1986) makes the argument that they differ primarily in melodic direction: that Uşşak is an ascending form (tonic-dominant-tonic) while Beyati is an ascending-descending form (dominant, lingers, then to tonic). Such directional motions are annihilated by the randomization. Expert 2 heard the n=3 Uşşak randomization as Hüseyni, which is again a closely related makam. This is the same confusion found in Experiment 1 and may be understood by observing that the single sharp difference occurs on a relatively rare tone.

Perhaps the most interesting responses in Experiment 3 were provided by Expert 3 to the two randomized Uşşak makams, who wrote "piece wanders over pitches with no makam structure detected." This is the only case where any of the experts took advantage of the "if any" clause in the question **Q**. Although the amount of randomization was the same, apparently the "wandering" in the Uşşak examples was more pronounced than the wandering in Hicaz. For at least one of the experts, the sound manipulations had annihilated the makam structure.

Expert 2 commented that the n=3 randomized examples "sound like an overture." An overture typically contains many small snippets of the pieces that are to come; the n=3 randomizations contain many small (3-note) snippets from the database from which the parameters of the Markov chain are drawn. Thus Expert 2 was likely hearing many of the small motifs inherited from the SymbTr database (Karaosmanoğlu, 2012).

To calculate the kappa value for the agreement between the three experts requires handling the "no makam" response of Expert 3. If this is interpreted as another category of response, the kappa value lies in the range $(0.447, 0.622)$ as shown in Table C1. These values may be interpreted as in Table 2 as "substantial agreement" for the sound examples of this experiment.

Experiment 4: Cross-Makam Generation

The sound examples for Experiment 4 are constructed to help determine which hypothesis (perde or seyir) is stronger. The examples are formed by merging two of the taksims, grafting the perdes of one makam onto the seyir of a second (and vice versa). For instance, the sound example labeled perHusSeyMah uses the perdes of the

Hüseyni makam along with the seyir of the Mahur makam. Similarly, the sound example labeled perMahSeyHus uses the perdes of the Mahur makam along with the seyir of the Hüseyni makam.

Such cross-generated sound examples do not have a single "correct" answer. Listeners may choose the makam represented by the pitch structure (providing support for the perde hypotheses), they may choose the makam represented by the temporal structure (providing support for the seyir hypotheses), or they may respond with some other makam. The latter case may indicate that the crossing procedure has destroyed the nature of the makam, that the sound example was inherently ambiguous, or perhaps that crossing of certain pairs of makams may imply a third. Such situations may be challenging to interpret.

Procedure and Participants

The procedures and participants were the same as in Experiments 1-3.

Stimuli

The sound stimuli for Experiment 4 are constructed by grafting the pitch profile of one makam onto the temporal profile of another. The basic source material is the original corpus of 17 makams, each of which is subjected to the analysis of Appendix B. Two makams A and B are chosen for each example, and a one-to-one mapping is constructed that replaces each note of a makam with the corresponding note from the other. For instance, in the first sound example, the pitches/perdes from Muhayyer are mapped onto the temporal motion of the Uşşak makam and the result is called "perMuhSeyUss" in the first line of Table 5. The complete procedure is described in detail in Appendix D.

Results

Experts 1 and 2 heard sound examples 1, 4, and 5 as dictated by the temporal motion, in support of the seyir hypothesis. Both also heard sound examples 7 and 8 as dictated by the pitch content, in support of the perde hypothesis. Expert 3 agreed on

	Sound Example	Expert 1	Expert 2	Expert 3
1	perMuhSeyUss	Uşşak	Uşşak	Rast
2	perHusSeyMah	Similar to Neva	Uşşak	Rast
3	perNihSeyYeg	Kürdilihicazkar	Kürdi	Muhayyer
4	perMahSeyHus	Hüseyni	Hüseyni	Hüseyni
5	perUssSeyMuh	Muhayyer	Muhayyer	Muhayyer
6	perYegSeyNih	Rast	Rast	None
7	perNihSeySuz	Nihavent	Nihavent	ends in Nihavent
8	perSuzSeyNih	Suzinak	Suzinak/Rast	Rast

Table 5: Results for Experiment 4: "Cross-Makam Generation"

the importance of the seyir in sound examples 4 and 5 and replied that sound example 7 "goes through a bunch of mixed (confusing) melodies and concludes in Nihavent makam," thus agreeing (in the end) with the others in support of the perde hypothesis. In contrast, the responses to sound examples 2, 3, and 6 apparently show little agreement with either of our prior expectations. There was no overlap in the answers of the three experts about the perceived makam of sound examples 2 and 3. While Experts 1 and 2 perceived Rast in example 6, Expert 3 declared sound example 6 to be in no recognizable makam.

To shed light on the results of this experiment, we focus on three issues. First, what distinguishes examples 1, 4, and 5 (where the experts agree with the seyir hypothesis) from examples 7 and 8 (where the experts agree with the perde hypothesis)? Second, what distinguishes examples 1, 4, 5, 7 and 8 (where the experts mostly agree) from sound examples 2, 3, and 6 (where the experts mostly disagree)? Finally, what distinguishes sound examples 2 and 3 (where the experts completely disagree) from sound example 6 (where two agree and one hears no makam at all)?

One way to understand the difference between the sound examples where the seyir hypothesis dominates (1,4,5) and the sound examples where the perde hypothesis dominates (7,8) is that in the former the pitch changes are mild while the tonic-dominant relationship is different, while in the latter the pitch differences are large and the tonic-dominant relationships are the same. For example, from the point of view of interval content, Uşşak and Muhayyer (of examples 1 and 5) are quite similar (Figure 2 shows the key signature[1] of Uşşak and Muhayyer as differing by just one sharp, and Figure A1 displays this difference in the interval sets in the lightest shade of grey). On the other hand, the tonic-dominant relation in Uşşak is a fourth while that in Muhayyer is a fifth. In contrast, Nihavent and Suzinak (of examples 7 and 8) have almost no relationship in terms of perdes (the dark grey coloring in Figure A1 indicates a large difference in interval set) while the dominant-tonic relationship is identical (a fourth). Moreover, agreement among the experts in example 8 may be greater than is obvious from a glance at Table 5 since Rast (the outlier response of Expert 3) is closely related to Suzinak (Figure 2 displays the key signatures as differing by just one flat). Example 4 fits a similar pattern, since both Hüseyni and Mahur differ greatly in interval content but have the same dominant-tonic relationship (a fifth). There are two common threads that run through these cases. First, the seyir hypothesis tends to dominate when the interval sets are close; the perde hypothesis tends to dominate when the pitch sets are significantly different. Second, when the dominant-tonic interval is the same, the responses tend to support the perde hypothesis; when the dominant-tonic interval differs, the responses tend to support the seyir hypothesis.

While the experts are in considerable agreement in the above five sound examples, they appear to be in considerable disagreement in the other three. For example, there are six different makams cited by the three experts in examples 2 and 3. As we will argue, this disagreement is more apparent than real. Expert 1 cites Neva for example 2 and Kürdilihicazkar for example 3. Figure A1 shows that the interval sets of Neva and Hüseyni are identical, and that the interval sets for Kürdilihicazkar and Nihavent are identical. Since Hüseyni and Nihavent are the expected answer under the perde hypothesis, Expert 1's answers both support the perde hypothesis, under the assumption that rotations of the interval set of a makam are identified. Similarly, Expert 2 chose Kürdi for example 3 (which has the same interval content as Nihavent) and Uşşak for example 2 (which has a key signature that is exactly one sharp different from Hüseyni, as shown in Figure 2). Thus Expert 2's responses also support the perde hypothesis

[1] This is not the way a Turkish practitioner would describe these relationships. Ederer (2013) comments that the scalar material of the Uşşak makam consists of an Uşşak tetrachord on dügâh conjoined with a Nihavent pentachord on neva, while the Muhayyer makam consists of an Uşşak pentachord on dügâh conjoined with a Buselik tetrachord on hüseyni.

under the assumption that nearby makams (in the sense of key signatures) may be identified. Similarly, Expert 3's response of Muhayyer can be understood as further support for the perde hypothesis because its key signature is one sharp different from Kürdi. Of these six responses, the only one that is not interpretable in this manner is Expert 3's choice of Rast. Thus five of the six responses in examples 2 and 3 can be viewed as support for the perde hypothesis, under the assumptions that makams with nearby key signatures (no more than one accidental difference) and those which are rotations of the interval set are identified.

In example 6, two of the experts responded with Rast and one responded that the example had no discernible makam structures. There are three possible explanations. It may be a result of the strong leading-tone: Yegah, Nihavent and Rast all have a leading tone that is close to the root (Nihavent and Rast have the F# to G relationship while Yegah and Rast have the same pitch set, differing only by a modal transposition). On the other hand, Yegah and Rast have almost identical pitch sets that differ by only one note (the low C# in Yegah). This C# does not occur frequently in the particular Nihavent taksim. Meanwhile, Nihavent and Rast have the same root-dominant structure (which may again be interpreted as an aspect of temporal motion and seyir), which is distinct from the root-dominant structure in Yegah. Thus Rast may be viewed as a combination or hybrid of Yegah and Nihavent with the pitch set drawn from Yegah and the seyir drawn from Nihavent. Appendix G of Ederer (2011) offers a third alternative based on the complex historical relationships between Yegah and Rast. Overall, sound example 6 remains somewhat enigmatic.

Interpreting the "no makam" response as a category of response, the kappa value for this experiment lies in the range $(0.411, 0.424)$ as shown in Table C1. This may be interpreted as in Table 2 as "moderate agreement" among the experts for the sound examples of this experiment.

Discussions and Conclusions

When we first contemplated these experiments, we feared that even experts might not be able to name the intended makams, leaving us with a dilemma; would this mean that the experts failed to identify the makam, or would it mean that the process of creating the resynthesized sound examples had destroyed the essence of the makam? Fortunately, in the majority of examples, the experts concurred with our intended makams and with each other. This gave us confidence that the sound resynthesis techniques were transparent (at least from the point-of-view of makam recognition). The answer to the question as to the identifiability of the makam from purely acoustical data is unequivocal: yes, expert listeners can accomplish this task.

The bulk of the experiments were then designed to uncover the acoustic cues that the experts might use to achieve this feat of cognition, centering on the two central hypotheses of a pitch/perde-based recognition and a temporal/seyir-based recognition. The results of Experiments 2 and 3 show that pitch relationships alone (i.e., the scale) can account for the recognition of makams in many cases. Whether the notes of the improvisation are scrambled or randomized, the experts were often able to identify the intended makam. When they "missed," it was often easy to see why: makams with similar interval sets are easy to confound. Had we stopped the experiments at this point, we would have concluded (in agreement with a simple interpretation of the AEU theory) that recognition of makams is primarily a pitch-based activity; we would have been pleased to report that the auditory pitch acuity of the experts was fine enough to distinguish many makams from their pitch content alone.

But the testimony of performers and authorities on Turkish music (including our experts) suggests that temporal information ought to be significant. We were not able to design an experiment that isolates the seyir hypothesis (as Experiments 2 and 3 isolate the perde hypothesis). But we were able to conduct Experiment 4, which tests the relative importance of the two hypotheses. In many cases, the seyir-based recognition dominates the perde-based recognition. Thus, although pitch relationships alone can be used to identify the makam, when both pitch information and temporal information are present and conflict, listeners tend to choose the makam represented by the temporal information (e.g., sound examples 1, 4, and 5 of Experiment 4), especially when the interval sets of the makams are close. In cases where the pitches are very different (sound examples 7 and 8 of Experiment 4), the perdes may still dominate. While one or the other of these phenomena may dominate in any given experiment, it should be understood that in normal listening, perde and seyir work together to define the makam.

Makam recognition is a highly complex cognitive process. When there is only pitch information available (as in the contrived sound examples of Experiments 2 an 3), it is often enough to identify the makam. However, when temporal information is present, and the perde structures happen to be "close," the temporal information may be the preferred vehicle for makam recognition. Some of the cases where the expert's answers differed from our prior expectations may provide clues to relationships between makams. For example, sound examples 2 and 6 of Experiment 4 suggest that when makams A and B are combined (choosing the perdes of A and the seyir of B), the proper response may be neither A nor B, but a third makam C which is related to the two input makams by a balance of both pitch and temporal similarities. Further experiments, designed specifically to test such relationships, could be conducted with the aim of uncovering a "distance" function that might measure the similarities and differences between makams based on pitch and temporal structures.

Castellano et. al. (1984) consider the features of tonal organization that may become internalized through experience in the context of Indian rags. We did not extend our experiments to Western listeners because it would be difficult for those without experience to distinguish one makam from another, and clearly impossible to name them. It would be interesting to conduct probe-tone experiments analogous to those of Krumhansl & Shepard (1979) using the tonal material of Turkish makams. These would require an expanded palette of pitches to include at least the 24 tones of the AEU system, although they might also benefit from the full 53-tone set of commas and/or the inclusion of 12-tone equal tempered pitches, which might be significant for listeners who also have significant exposure to Western idioms. In order to bypass the need to choose what pitches to use (and indeed, to bypass the need to choose a music-theoretic system on which to base the experiments) we have analyzed specific musical performances and derived the pitch sets used in the experiments directly from those performances. The accuracy of the pitch extraction is on the order of one cent, and so is finer than any common theoretical system.

There are many examples throughout a variety of musical cultures where small pitch changes are used as expressive elements in performance; these are often considered to be ornamental inflections about some set of nominal pitches. Ayari & McAdams (2003) report that for many Arab listeners, small comma-sized variations may signal a change in the identity of the makam. The experiments in this paper confirm that a comma change (such as those that distinguish the various makams) may carry important information about the form and organization of the piece, at least in realm of the taksims of Turkish makam music. Such differences in form (i.e., the various makams) can often be perceived and identified by experts.

Oram and Cuddy (1995) observe that a listener's sensitivity to pitch-distributional information may be important in developing listening strategies for atonal music. This may equally hold when listening to an unfamiliar musical style (such as a Westerner listening to makam music) where the form of tonality is different. Extending this one step further, it is also plausible that an expert may use pitch distributional information when other, more familiar information (such as temporal order) is unavailable. For example, in the scrambling experiments, where the order of pitches were randomized and temporal cues destroyed, the experts may have adjusted their listening strategies. When those temporal cues were returned in Experiment 4, it is plausible that their listening strategies readjusted to focus on the most pertinent information. (Bharucha, 1984) comments that the tonal hierarchy may be evoked either by the relative durations of tones in a piece or by activation of long term memory.

Deutsch (1984) comments on Castellano et. al. (1984) and asks, in the context of Western music, if key assignments tend to be made on the basis of pitch collections alone, or if the order of the notes is also significant. This provides a simple experimental paradigm where the interactions between tonal perceptions and temporal ordering of events can be studied. Deutsch constructs an example where identical sets of notes imply different keys depending the order in which they are played, and concludes that "we are dealing with an elaborate bootstrapping operation... so that ultimately both a key and a sequential representation are arrived at by the listener." This is consistent with our conclusion that a makam may be identified from pitch information alone but that when sequential information is present, it may also exert a significant influence. Indeed, in certain cases (such as the makams along the diagonals in Figure A1) the pitch sets of two makams are identical and the sequential presentation is crucial. The results of Experiment 4 attempt to address the relative importance of the sequential and the pitch information.

In linguistics, the saying that "native speakers do not make grammatical errors" (Andersson and Trudgill, 1990) can be interpreted to mean that language is a social construct where the limits of usage are governed by the speakers of that language. In musical discourse, while there is no "native speaker," experts do spend years learning, training, and performing in a style that is governed by their practice. Our intention is not to idolize such experts (although we do have great esteem for their abilities) but to use their responses to understand the limits of the makam style and the limits of human perception. In the experiments, the simplest situation is when the experts agree with our intended makam. When the experts agree with each other (but disagree with our intended makam), this indicates that we designed the experiment poorly or misunderstood some aspect of the sound example. When the experts disagree among themselves, it is possible that one has "made a mistake" perhaps through inattentive listening or happenstance, but our first presumption is that they disagree because they are attending to different aspects of the experimental stimulus. In such cases, we have tried to pinpoint plausible explanations for such disagreements. It is also possible that certain combinations are fundamentally ambiguous.

Seyir and perde are core elements of makam music, and they function as central features in makam identification. The experiments presented above for the purpose of exploring the interactions between the dynamic elements of seyir (temporal information) and the static elements of perde (pitch information) in Turkish makam music reveal some of the intricate acoustic features needed for makam recognition, pointing to a combined seyir-perde architecture underlying the makams. The experiments rely on a back-door approach that explores the inner workings of the seyir-perde mechanisms by investigating what does and does not work, by finding the limits of what expert listeners do (and do not) hear as proper makam structure. This was achieved by the creation of synthetic taksims which deliberately distort and caricaturize elements of seyir and perde in the hopes of approaching the essential "kernel" of makam-ness. This same kind of approach (of designing experiments to concretely and unambiguously uncover the abilities of expert listeners) may be applied to related issues. For instance, interviews suggest that the performance of a makam ought to be more than a mere collection of stereotypical motives; if so, what are these audible and measurable quantities, and how can they be demonstrated or falsified?

Acknowledgements

The authors would like to thank the experts Niyazi Sayın, Ruhi Ayangil, and Necdet Yaşar, without whom this study would not have been possible. Others who were helpful or supportive of this work include Ozan Yarman, Gülçin Yahya Kaçar, Uğur Keçecioğlu, Eric Ederer, and the participants of the workshop *Türk Musıkısinde İcraya Dayalı Nazariyat Modeli Arayışları* in Nov. 2012 at ODTÜ. Ali Yaşar was key in our interactions with Necdet Yaşar throughout the experiments and Uğur Keçecioğlu helped with the logistics in Istanbul. Mehmet Polat helped decipher some of the musical riddles caused by the experts' responses. The authors are grateful to the Institute for Applied Mathematics at the Middle East Technical University in Ankara for support throughout a sabbatical visit by the second author. Preliminary presentations of the work in this paper were made at Başkent University and at ODTÜ in Nov. 2012, and at McGill University in Jan. 2013.

References

Akkoç, C. (2002). "Non-deterministic scales used in traditional Turkish music," *Journal of New Music Research*, Vol.31, No.4, pp. 285-293.

Akkoç, C. (2008). "Constructing a theory of Turkish music based on practice by traditional masters," Conference on Issues Surrounding the Conflict between Theory and Practice in Turkish Music and Potential Resolutions, Istanbul, Turkey.

Aksoy, B. (2005). "Necdet Yaşar (record label biography)" Kalan Müzik, Kalan.com.

Andersson, L. G. & Trudgill, P. (1990). *Bad Language*, John Wiley & Sons, Limited.

Ayari, M., & McAdams, S. (2003). "Aural Analysis of Arabic Improvised Instrumental Music (Taqsım)," *Music Perception*, Winter, Vol. 21, No. 2, 159-216.

Arel, H. S. (1968). Türk Musikisi Nazariyatı Dersleri. Istanbul, Hüsnütabiat Matbaası.

Bharucha, J. J. (1984). "Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling," *Journal of Experimental Psychology: General*, 113(3), 421-425.

Beken, M. & Signell, K. (2006). "Confirming, delaying, and deceptive elements in Turkish improvisations," Maqam Traditions of Turkic Peoples, Berlin.

Boltz, M. (1989). "Rhythm and good endings - effects of temporal structure on tonality judgments." *Perception & Psychophysics*, 46(1), 9-17.

Bozkurt, B., Yarman, O., Karaosmanoğlu, K. & Akkoç, C. (2009). "Weighing diverse theoretical models on Turkish makam music against pitch measurements: a comparison of peaks automatically derived from frequency histograms with proposed scale tones," *Journal of New Music Research*, Vol. 38, Issue 1, pp. 45-70.

Bozkurt, B., Gedik, A. C., Savacı, F. A., Karaosmanoğlu, M. K. & Üzbek, M. E. (2010). "Klasik Türk müziği kayıtlarının otomatik olarak notaya dökülmesi ve otomatik makam tanıma," TÜBİTAK project 107E024 final report, Izmir.

Camel Audio (2013). "Alchemy," `http://www.camelaudio.com/Alchemy.php` (last accessed on April 3, 2014).

Castellano, M. A., Bharucha, J. J. & Krumhansl, C. L. (1984). "Tonal hierarchies in the music of North India," *Journal of Experimental Psychology: General*, 113(3), 394-412.

de Cheveigné, A. and Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.* 111, 1917.

Deutsch, D. (1982). Ed., *The Psychology of Music*, Academic Press Inc., San Diego, CA.

Deutsch, D. (1984). "Two issues concerning tonal hierarchies: Comment on Castellano, Bharucha, and Krumhansl," *Journal of Experimental Psychology: General*, 113(3), 413-416.

Ederer, E. B. (2011). *The Theory and Praxis of Makam in Classical Turkish Music 1910-2010*, Ph.D. Dissertation, UC Santa Barbara.

Ederer, E. B. (2013). private communication.

Ezgi, S. Z. (1933). Nazarî ve Amelî Türk Mûsıkîsi. Vol. I. Istanbul, Milli Mecmua Matbaası. pp. 8-29.

Fleiss, J. L. (1971). “Measuring nominal scale agreement among many raters,” *Psychological Bulletin,* 76(5), 378-382.

Gill-Gürtan, D. (2011). “Performing meşk, narrating history: legacies of transmission in contemporary Turkish musical practices” *Comparative Studies of South Asia, Africa and the Middle East* 31(3): 615-630.

Holtzberg, M. (2008). *Keepers of Tradition: Art and Folk Heritage in Massachusetts*, Univ of Massachusetts Press.

Kantemir, D. (1700). “Kitāb-i ilmül mūsiki ala vechil-hurūfāt,” (Istanbul, 1700; Turkish trans., 1976)

Karaosmanoğlu, M. K., Yılmaz, S. M., Tören, Ü., Ceran, S., Uzmen, U., Cihan, G. & Başaran, E. (2009). “Mus2okur: Multimedia Encyclopedia of Turkish Music,” `http://www.musiki.org/Mus2okur_en.aspx` (last accessed on April 3, 2014).

Karaosmanoğlu, M. K. (2012). “A Turkish Makam Music Symbolic Database for Music Information Retrieval: SymbTr,” *Proceedings of ISMIR*. `http://compmusic.upf.edu/node/140` (last accessed on April 3, 2014).

Krumhansl, C. L. & Shepard, R. N. (1979). “Quantification of the hierarchy of tonal functions within a diatonic context,” *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594.

Landis J. R. & Koch, G. G. (1977). “The measurement of observer agreement for categorical data,” *Biometrics*, 33: 159-174.

New Grove Dictionary (2013). “Turkey: Makam”

Oram, N. & Cuddy, L. L. (1995). “Responsiveness of Western adults to pitch-distributional information in melodic sequences,” *Psychological Research*, 57(2), 103-118.

Rabinovitz, B. E. (2011). “Importance of Temporal Organization in Memory for Music and Language,” Ph.D. Dissertation, The American University, Dept. Psychology.

Sethares, W. (2011). “Poem Maker,” Wolfram Demonstrations Site, `http://demonstrations.wolfram.com/PoemMaker` (last accessed on April 3, 2014).

Sethares, W., (2014). “Makam Experiment Website,” `http://sethares.engr.wisc.edu/makam/makam.html` (last accessed on April 3, 2014).

Shannon, C. E. (1948). “A Mathematical Theory of Communication,” *The Bell System Technical Journal*, vol. 27, nos. 7 & 10, pp. 379-423 and 623-656, July & October.

Signell, K. L. (1986). *Makam: Modal Practice in Turkish Art Music*, DaCapo Press. Second Edition: Usul Editions 2008.

Signell, K. L. (2011). “The art of master musician Necdet Yaşar as a key to the subtleties of classical Turkish music,” in *Ethnomusicological Encounters with Music and Musicians: Essays in Honor of Robert Garfias*, Ed. T. Rice, UCLA, Ashgate.

Sim, J. & Wright, C. C. (2005). “The kappa statistic in reliability studies: use, interpretation, and sample size requirements,” in *Physical Therapy*, Vol. 85, No. 3, pp. 257-268.

Six, J. & Cornelis, O. (2011). “Tarsos - a platform to explore pitch scales in non-western and western music,” *Proc. of the 12th Int. Soc. for Music Information Retrieval*, ISMIR.

Tanrıkorur, C. (2005). Osmanlı dönemi Türk musikisi, Dergah yayınları, 2nci baskı, Istanbul, p. 166. (Compilation of Tanrıkorur’s writings including “(the art of) Taksim in Classical Turkish Music,” presented at the Univ. of Maryland, Nov, 1989).

Touma, H. H. (1996). *The Music of the Arabs*, trans. Laurie Schwartz. Portland, Oregon: Amadeus Press.

Tokuz, G. (2009). *Taburı Necdet Yaşar*, Brainstorm, BS Reklam Ltd. Şti, Istanbul.

Tura, Y. (1988). *Türk Musikisinin meseleleri*, Pan Yayıncilik, Istanbul.

Viterbi A. J. (1967). “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE Trans. on Information Theory* 13 (2): 260269.

Welch, L. R. (2003). “Hidden Markov models and the BaumWelch algorithm,” *IEEE Information Theory Society Newsletter*, Vol. 53, No.4.

Yarman, O. (2008). *79-Tone tuning & theory for Turkish maqam music as a solution to the non-conformance between current model and practice,* Istanbul Technical University, Ph.D. thesis.

Acemkürdi	4	9	9	9	4	9	9
Beyati	7	6	9	9	4	9	9
Buselik	9	4	9	9	4	9	9
Eviç	5	9	8	9	5	9	8
Hicaz	5	12	5	9	8	5	9
Hicazkar	5	12	5	9	5	12	5
Hüseyni	7	6	9	9	7	6	9
Isfahan	7	6	9	9	4	9	9
Kürdi	4	9	9	9	4	9	9
Kürdilihicazkar	4	9	9	9	4	9	9
Mahur	9	9	4	9	9	9	4
Muhayyer	7	6	9	9	7	6	9
Neva	7	6	9	9	7	6	9
Nihavent	9	4	9	9	4	9	9
Rast	9	8	5	9	9	8	5
Rehavi	9	8	5	9	9	8	5
Saba	7	6	6	12	4	9	9
Sazkâr	9	8	5	9	9	8	5
Segah	5	9	8	9	5	13	4
Sultaniyegah	9	4	9	9	4	13	5
Suzinak	9	8	5	9	5	12	5
Şehnaz	5	12	5	9	5	12	5
Uşşak	7	6	9	9	4	9	9
Yegah	9	8	5	9	7	6	9

Table A1: Interval sets of the 24 makams of Figure 2 can be represented as integer multiples of the Holdrian comma. Data are drawn from (Karaosmanoğlu et. al., 2009). These scales are shown in musical notation in Figure 2.

A Measuring the Distance Between Makams in Pitch Space

One way of characterizing the pitch content of makams is via the set of successive intervals that occur in the scale. Table A1 shows the interval sets for the 24 makams of Figure 2. Intervals are measured in terms of the Holdrian comma, an interval of 1/53 of an octave (22.6 cents) (Touma, 1996). For example, in the Uşşak makam, the interval between the root and the second tone is 7 commas, between the second and third tones is 6 commas, etc. Because the scales repeat at the octave, each row of the table sums to 53.

In the standard AEU theory, commas are restricted to multiples of 4, 5, 8, 9, 12 and 13. Bozkurt, et. al. (2009) and Akkoç (2002) show that in practice, commas of size 6 and 7 also occur, and Table A1 adopts these values. For example, in the standard AEU theory, the interval set for the Saba makam would be 8, 5, 5, 13, 4, 9, 9. Such differences arise from inconsistencies between theory and practice and may be subject to controversy.

The interval sets can be used to describe a metric in many ways. Perhaps the simplest is to calculate the sum of the absolute values of the L_1-distance $||x - y||_1$ where x and y are interval sets. Somewhat more meaningful from a musical perspective is to consider all rotations (circular shifts) of the interval sets. Let $R_i(x)$ be a circular shift of the interval vector x by i positions to the right. Then

$$d(x, y) = \min_i ||R_i(x) - y||_1 \tag{A1}$$

is the value of the smallest of the L_1-differences between the interval set y and all possible rotations of the interval set x. This effectively identifies those scales which are identical but for transposition. In a Western context, Eq. A1 would identify scales such as C-Major, D-Dorian, and E-Phrygian (etc.) that contain the same interval-set but start on a different note. From a mathematical perspective, observe that $d(x, y) = d(y, x)$ for all x and y. Identifying all scales x and y for which $d(x, y) = 0$ into an equivalence class makes Eq. A1 a metric on the space of interval sets.

Figure A1 shows the distances $d(x, y)$ between all the makams of Figure 2 as measured by Eq. A1. In the figure, white represents zero distance. For example, Uşşak, Beyati, and Isfahan contain the same set of intervals, indicating the close relationship between these scales. Black represents the largest distance; for the makams of Table A1, this is between the pair Şehnaz/Hicazkar and the triplet Uşşak/Beyati/Isfahan, which has a numerical value of 18. Gray values represent intermediate distances. In order to more clearly display makams with similar interval sets, the order of presentation has been rearranged according to a k-means clustering algorithm. The effect of this reordering can be seen in the white squares that sit along the main diagonal; the Şehnaz/Hicazkar group and the Uşşak/Beyati/Isfahan groups are clearly delineated, as

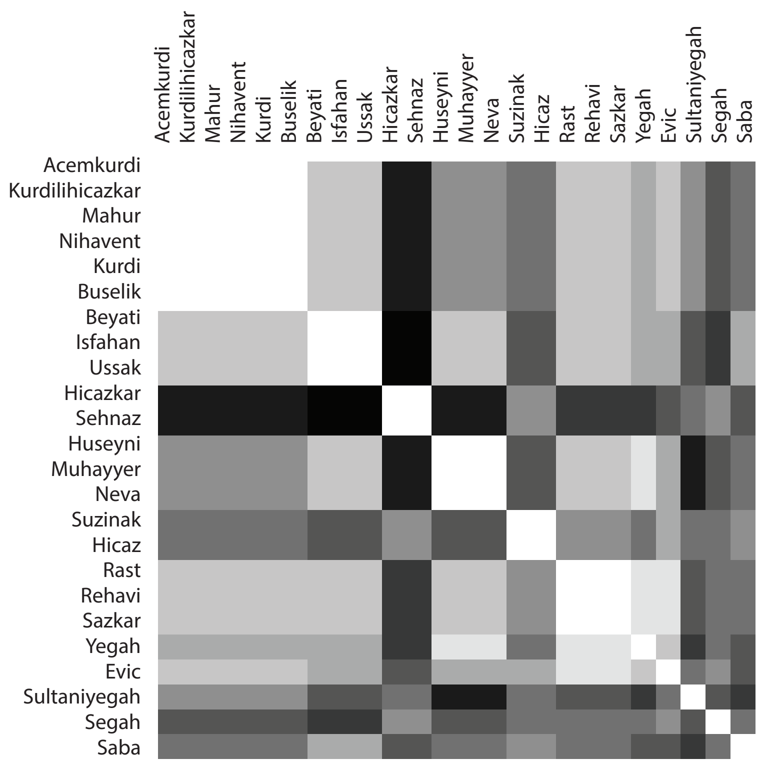

Figure A1: The distance $d(x, y)$ as calculated by Eq. A1 for all the makam pairs in Table A1. White represents zero distance, black represents the largest distance in the set, and shades of gray indicate intermediate values. White blocks along the diagonal show sets of makams with identical interval sets, and include Beyati/Isfahan/Uşşak, Hicazkar/Şehnaz, Hüseyni/Muhayyer/Neva, Suzinak/Hicaz, and Rast/Rahavi/Sazkâr. The largest group, at the top left, contains Acemkürdi/Mahur/Kürdilihicazkar/Nihavent/Kürdi/Buselik, all of which are constructed from the same interval set (allowing for rotation). Makams with closely related (but not identical) interval sets are indicated in light grey.

are several other sets of makams with identical interval sets under the metric given by Eq. A1.

B Analysis-Resynthesis Method of Generating Sound Stimuli

This appendix details the common steps that underly the synthetic taksims used in the experiments. The overall approach is a method that relies on a computer-based analysis of the original corpus of 17 taksims. For each taksim, a set of instantaneous pitch measurements is made using the open source program *Tarsos* (Six & Cornelis, 2011). This implements the YIN pitch detection algorithm (de Cheveigné & Kawahara, 2002) and is used to estimate the instantaneous pitch of the original performance at a rate of 100 times per second. The raw pitch data are converted into estimates of the pitch centers and transition probabilities using the Expectation Maximization (EM) algorithm (Welch, 2003). Effectively, this transforms the 100-times-per-second data to a small collection of eight pitches (per octave) and corresponding transition probabilities, which indicate the likelihood that any given pitch will transition to any other given pitch throughout the analyzed piece. These probabilities, together with the raw pitch measurements, are then used to estimate the note start and end times via the Viterbi (1967) algorithm. This reduces the raw pitch data to a set of "note-level" data that can be transformed into a MIDI file. Since the pitches of the notes do not all lie on the pitches of the Western 12-tone equal tempered scale (as is the default in a MIDI representation), each note is coded as a pitch value along with a pitch-bend value. Together, these allow the sounded MIDI notes to have a repeatable pitch accuracy of better than one cent.

The model presumes eight pitches (or pitch clusters) per octave, as is common in Turkish makams. The centers of these clusters are denoted $s_1, s_2, \ldots, s_m$. These are unknown, and the EM algorithm is used to estimate the states of the underlying Markov chain $\{X_i\}$ and the transition probabilities

$$\alpha_{ij} = P(X_{l+1} = s_j | X_l = s_i), \quad i, j = 1, 2, \ldots, m. \tag{B1}$$

Suppose there are n observations $y_1, y_2, \ldots, y_n$ where $y_i = X_i + n_i$ and where n_i conditioned on $X_i = s_l$ is independent of both $\{X_1, X_2, \ldots, X_{i-1}\}$ and $\{n_1, n_2, \ldots, n_{i-1}\}$, and is Gaussian with mean μ_l and variance σ_l^2. In Figure B1, the points y_i form the clouds of small dots; they tend to cluster around certain perde centers that are the pitches s_i used in the performance. For notational convenience, define $\lambda_l = (\mu_l, \sigma_l^2)$, $l = 1, 2, \ldots, m$. Conditioned on $X_i = s_l$, y_i is Gaussian with mean $s_l + \mu_l$ and variance σ_l^2, and the probability density is

$$f(y : \lambda_l) = \frac{\exp(-(y - \mu_l - s_l)^2 / 2\sigma_l^2)}{\sqrt{2\pi\sigma_l^2}}. \tag{B2}$$

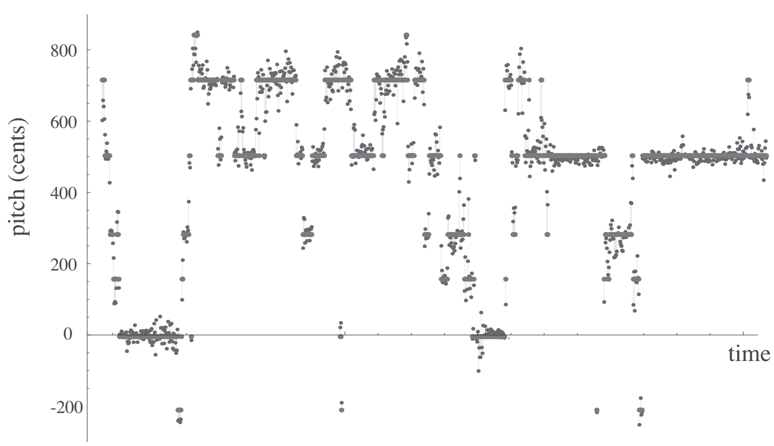

Figure B1: A six-second segment from the Uşşak taksim. The cloud of small points are the instantaneous pitches as detected by the Tarsos-YIN procedure. The solid horizontal lines are the perde centers as detected by the EM-Viterbi method, which can be exported directly into MIDI. The EM step detects the scale values (in this case, eight pitches at -212, -3.5, 158, 276, 495.5, 713, and 880 cents as well the probabilities of transition, which are not shown). The Viterbi step finds the most probable path (the best set of horizontal lines) to maximize the likelihood throughout the complete performance.

The EM algorithm updates the estimates of the parameter values α_{ij} and λ_i for

$i, j = 1, 2, \ldots m$ by computing the "forward" probabilities

$$a_j(1) = \hat{u}_j(1) f(y_1 : \lambda_j),\ j = 1, 2, \ldots, m \tag{B3}$$
$$a_j(i) = \sum_{k=1}^{m} a_k(i-1) \alpha_{kj} f(y_i : \lambda_j),\ j = 1, 2, \ldots, m \quad i = 2, \ldots, n$$

(which can be initialized to $u_j(1) = 1/m$, $j = 1, 2, \ldots, m$) and the "backward" probabilities

$$b_j(n) = 1,\ j = 1, 2, \ldots, m \tag{B4}$$
$$b_j(i) = \sum_{k=1}^{m} \alpha_{jk} f(y_{i+1} : \lambda_k) b_k(i+1),\ j = 1, 2, \ldots, m \text{ and } i = n-1, n-2, \ldots, 1.$$

The likelihood $L = \sum_{j=1}^{m} a_j(n)$ increases at each iteration. Updated estimates of the parameters are

$$\begin{aligned} \alpha^*_{jk} &= \frac{\sum_{i=2}^{n} \hat{v}_{jk}(i)}{\sum_{i=2}^{n} \sum_{l=1}^{m} \hat{v}_{jl}(i)} \quad j, k = 1, 2, \ldots, m \\ \mu^*_j &= \frac{\sum_{i=1}^{n} \hat{u}_j(i) y_i}{\sum_{i=1}^{n} \hat{u}_j(i)} \quad j = 1, 2, \ldots, m \\ \sigma^{2*}_j &= \frac{\sum_{i=1}^{n} \hat{u}_j(i)(y_i - \mu^*_j)^2}{\sum_{i=1}^{n} \hat{u}_j(i)} \quad j = 1, 2, \ldots, m \end{aligned} \tag{B5}$$

where

$$\hat{u}_j(i) = \frac{a_j(i) b_j(i)}{L} \quad j = 1, 2, \ldots, m \quad i = 1, 2, \ldots, n \tag{B6}$$
$$\hat{v}_{jk}(i) = \frac{\alpha_{jk} f(y_i : \lambda_k) a_j(i-1) b_k(i)}{L} \quad j, k = 1, 2, \ldots, m \quad i = 2, \ldots, n.$$

The iterations may be initialized with $\alpha_{ij} = 1/m$ for all $i, j = 1, 2, \ldots, m$ and $\mu_i = 0$. The most probable path is the sequence of states that maximizes the likelihood. Let

$$\delta(n) = \max_{x_1, \ldots, x_n} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, y_1, y_2, \ldots, y_n | \text{parameters}),$$

$$\delta_j(k) = \max_{x_1, \ldots, x_{k-1}} P(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}, X_k = s_j, y_1, \ldots, y_k | \text{parameters}),$$

and let $m_j(k)$ be the argument $x_1, x_2, \ldots, x_{k-1}$ at which the maximum of $\delta_j(k)$ occurs. This is the most probable path to be in state s_j at time k, given the observations up to time k. Initializing $\delta_j(1) = \frac{1}{m} f(y_1 : \lambda_j)$ for $j = 1, 2, \ldots, m$ and $p_j(1) = m_j(1) = \emptyset$, this can be computed by the iteration:

$$\begin{aligned} \delta_j(k+1) &= (\max_i \delta_i(k)\alpha_{ij}) f(y_{k+1} : \lambda_j) \quad j = 1, 2 \ldots, m \;\; k = 1, 2, \ldots, n-1 \\ p_j(k+1) &= \arg\max_i \delta_i(k)\alpha_{ij} \quad j = 1, 2 \ldots, m \;\; k = 1, 2, \ldots, n-1 \\ m_j(k+1) &= [m_{p_j(k+1)}(k), s_{p_j(k+1)}]. \end{aligned} \tag{B7}$$

The most probable path is then constructed from the Viterbi backtracking procedure. Since $\delta_j(n)$ is known for $j = 1, 2, \ldots, n$, let $p_n = \arg\max_j \delta_j(n)$. The most probable path

$$\{m_{p_n}(n), s_{p_n}\} \tag{B8}$$

represents an approximation to the perde centers of the performance. This is the “best” (in the maximum likelihood sense) set of “notes” (the horizontal lines in Figure B1) to approximate the instantaneous pitch measurements (the small dots). While this method of pitch extraction may seem complicated, it does not require a large number of free parameters (such as thresholds and filter-lengths) that are dependent on the details of the timbre of the sounds being analyzed (Bozkurt et. al., 2010). The only parameters that must be chosen are initial values for the scale steps s_i; these can be conveniently chosen from the peaks of the histogram of the observations y_i.

C Calculation of Fleiss' Kappa

The kappa coefficient (Fleiss, 1971) can be applied to multiple raters on categorical data. The kappa value

$$\kappa = \frac{\bar{P} - R}{1 - R} \tag{C1}$$

indicates the amount of agreement beyond that expected by chance, where $\bar{P}$ is the observed agreement and R is the agreement expected by chance alone. These are calculated as shown in Equations C2 and C3 below.

Let N be the number of sound examples in the experiment, M the number of raters (experts), K the number of categories (makams), and $m_{n,k}$ the number of experts who chose the kth makam in response to the nth sound example. The proportion of all assignments to the kth makam is

$$p_k = \frac{1}{NM} \sum_{n=1}^{N} m_{n,k}.$$

Since $\sum_k m_{n,k} = M$, the p_k's sum to unity, i.e., $\sum_k p_k = 1$. If the experts chose makams at random, the average agreement would be

$$R = \sum_{k=1}^{K} p_k^2. \tag{C2}$$

The amount of agreement observed among the M experts in the nth sound example is calculated from the proportion of agreeing pairs out of all the $M(M-1)$ possible pairs. This is

$$P_n = \frac{1}{M(M-1)} \sum_{k=1}^{K} m_{n,k}(m_{n,k} - 1),$$

which are averaged to give

$$\bar{P} = \frac{1}{N} \sum_{n=1}^{N} P_n. \tag{C3}$$

Equations C2 and C3 are then combined to give the κ of C1. κ values may be interpreted as in Table 2, although it is important to understand that with the small number of sound examples, there is no plausible way to measure the statistical significance of these values.

In calculating the κ values for the makam tests, some decisions are required in the interpretation of the responses. For example, some of the experts gave more than one answer: should the κ value be calculated based on the first/primary answer alone or should the response be "split into two" with each being weighted (and if so, what weighting should be applied)? Another issue occurs when two makams have the same interval content; for example, whether Beyati and Uşşak, which have identical interval sets, should be considered agreement or disagreement. Since each such variation may

Experiment	N	K	κ
1	6	8	$\kappa_1 = 0.671$
1	6	8	$\kappa_2 = 0.714$
1	6	7	$\kappa_3 = 0.735$
2	6	7	$\kappa_1 = 0.345$
2	6	7	$\kappa_2 = 0.263$
2	6	5	$\kappa_3 = 0.319$
3	6	7	$\kappa_1 = 0.500$
3	6	8	$\kappa_2 = 0.447$
3	6	5	$\kappa_3 = 0.622$
4	8	10	$\kappa_1 = 0.424$
4	8	10	$\kappa_2 = 0.424$
4	8	6	$\kappa_3 = 0.411$
Pseudo-experiment			
1	7	9	$\kappa_1 = 0.725$
1	7	7	$\kappa_2 = 0.755$
1	7	9	$\kappa_3 = 0.762$
2	7	9	$\kappa_1 = 0.298$
2	7	8	$\kappa_2 = 0.295$
2	7	5	$\kappa_3 = 0.359$
3	4	5	$\kappa_1 = 0.387$
3	4	5	$\kappa_2 = 0.304$
3	4	4	$\kappa_3 = 0.361$

Table C1: Details of the calculation of the κ parameters for the four experiments and the three pseudo-experiments (see text). The number of experts M is three in all cases.

give a slightly different value, we report a range of values. In particular, κ_1 uses only the first/primary answer of each expert, κ_2 weights all multiple responses equally, and κ_3 considers makams with identical interval sets to be "the same response." For example, in Experiment 1, $N = 6$ sound examples and $M = 3$ experts. With $K = 8$ different makams in the responses, $\kappa_1 = 0.671$; with $K = 8$, $\kappa_2 = 0.714$; with $K = 7$, $\kappa_3 = 0.735$. Accordingly, we report the range of values $\kappa \in (0.671, 0.735)$. A list of the parameters used in the κ calculations is given in Table C1.

Because the sound examples in each experiment were not all based on the same kinds of sound manipulations, we also calculate the κ-values for a set of *pseudo-experiments* which analyze the results of the sound examples with all equivalent sound modifications analyzed together (instead of analyzed in the groupings in which they were presented to the experts). These are:

1. Pseudo-experiment 1: all the data from Experiment 1 plus the mahurLinear sound example from Experiment 2

2. Pseudo-experiment 2: data from Experiment 2 (with mahurLinear removed), plus the RastXXXX examples from Experiment 3.

3. Pseudo-experiment 3: data from Experiment 3 (with RastXXXX examples removed).

The recalculated kappa values for these revised experiments are shown in the bottom half of Table C1. The ranges change somewhat: from (0.671, 0.735) for Experiment 1 to (0.725, 0.762) for Pseudo-experiment 1, from (0.263, 0.345) for Experiment 2 to (0.295, 0.359) for Pseudo-experiment 2, and from (0.447, 0.622) for Experiment 3 to (0.304, 0.387) for Pseudo-experiment 3. Experiment 4 is unaffected by this regrouping strategy. In terms of the agreement shown (as in Table 2), the only change is that Pseudo-experiment 3 has "fair" agreement while Experiment 3 has "moderate" agreement. This may indicate some influence of learning or influence from side information gleaned from conversations between the experiments. On the other hand, it should be intuitively clear that the randomized scramblings of Pseudo-experiment 3 pose at least as difficult a task as the more modest scramblings of Pseudo-experiment 2. This intuition is more consistent with the overlapping κ-ranges of Pseudo-experiments 2 and 3 than with the increase in κ-ranges from Experiments 2 to 3.

D Generation of Cross-Makam Sound Stimuli

This appendix details the technique used to create the sound stimuli for Experiment 4 in which each sound example is constructed using the pitch profile of one makam and the temporal profile of another. The basic source material is the original corpus of 17 makams, each of which is subjected to the analysis of Appendix B. For each sound example, two makams are chosen, which are labeled A and B.

The most probable path for makam A is given by Eq. B8 as a sequence of frequency/time triples

$$(f_1, t_1, e_1), (f_2, t_2, e_2), \ldots, (f_{n_A}, t_{n_A}, e_{n_A}) \tag{D1}$$

where n_A is the total number of pitch events (notes) in the performance, the start and end times t_i and e_i are in seconds (accurate to about 0.01 s), and the fundamental frequencies f_i are in Hz. The Viterbi procedure leading to Eq. B8 ensures that the f_i are quantized to a small number of values (eight per octave) which are the scale steps $s_1^j, s_2^j, \ldots, s_8^j$ where

$$j = \begin{cases} 1 & \text{octave above root} \\ 0 & \text{octave of the root} \\ -1 & \text{octave below root} \end{cases}.$$

Similarly, the most probable path for makam B is

$$(\hat{f}_1, \hat{t}_1, \hat{e}_1), (\hat{f}_2, \hat{t}_2, \hat{e}_2), \ldots, (\hat{f}_{n_A}, \hat{t}_{n_B}, \hat{e}_{n_B}) \tag{D2}$$

where n_B is the total number of notes in performance B and frequency is again quantized by Eq. B8 to the scale steps $\hat{s}_1^j, \hat{s}_2^j, \ldots, \hat{s}_8^j$.

In any given performance, not all scale steps may appear in all octaves. For example, there may be no occurrences of scale pitch s_2^1 in makam A even though there are occurrences of the corresponding pitch in makam B. Since the goal is to create a mapping between s and $\hat{s}$, missing terms can be "filled-in" using occurrences of the same scale step in other octaves. Thus the missing s_2^1 would be set equal to $2s_2^0$. The mapping between the two makams assigns the root note s_1^0 of makam A to the root note $\hat{s}_1^0$ of makam B, the second scale step s_2^0 of makam A to the second scale step $\hat{s}_2^0$, etc, until all have been assigned and there is a one-to-one correspondence between the pitch sets used in the two makams.

Somewhat more formally, let m be the number of (possibly filled-in) scale steps and relabel the s to remove the octave notation so that the s_i^j are relabeled with a single subscript $\sigma_1, \sigma_2, \ldots, \sigma_m$. Similarly, relabel the $\hat{s}_i^j$ as $\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_m$. Let g be the map which takes σ_k to $\hat{\sigma}_k$ and g^{-1} be its inverse, that is, $\hat{\sigma}_k = g(\sigma_k)$ and $\sigma_k = g^{-1}(\hat{\sigma}_k)$ for

$k = 1, 2, \ldots, m$.

The output sequences are created by applying the two mappings g and g^{-1} to the performances as extracted in Equations D1 and D2. For instance, each of the elements f_i in D1 corresponds to one of the σ_k, which is mapped by g to $\hat{\sigma}_k$. With a slight abuse of notation, denote this element $g(f_i) \equiv g(\sigma_k) = \hat{\sigma}_k$, and the sequence becomes

$$(g(f_1), t_1, e_1), (g(f_2), t_2, e_2), \ldots, (g(f_{n_A}), t_{n_A}, e_{n_A}), \tag{D3}$$

which consists of n_A pitches from makam B, each of which is associated with a start and end time specified by makam A. This sequence is then translated into a MIDI representation (where the frequencies are specified by MIDI note-number and pitch-bend) and then synthesized into audio using the same sounds as in the previous sound stimuli. The resulting sound example contains the perdes/pitches from makam B (i.e., the pitches $g(\sigma_k)$ performed with the timing/temporal information from makam A (i.e., at times $t_1, t_2, \ldots, t_{n_A}$). Similarly, the sequence

$$(g^{-1}(\hat{f}_1), \hat{t}_1, \hat{e}_1), (g^{-1}(\hat{f}_2), \hat{t}_2, \hat{e}_2), \ldots, (g^{-1}(\hat{f}_{n_B}), \hat{t}_{n_B}, \hat{e}_{n_B}) \tag{D4}$$

consists of n_B pitches from makam A which are associated with starting and ending times $\hat{t}_i$ and $\hat{e}_i$ from makam B. When translated into MIDI and then into sound, this contains the perdes/pitches from makam A with the timing/temporal information from makam B.

To show the procedure concretely, consider the first sound example which combines elements of Uşşak and Muhayyer. The procedure of Appendix B applied to the Uşşak makam results in a sequence defined by Eq. D1 containing $n_A = 105$ note events, each with a specified pitch, start time, and end time. There are 10 distinct pitches (those without asterisks in first column) with fundamental frequencies that appear in the second column of Table D1. The same procedure applied to the Muhayyer makam results in a sequence given by Eq. D2 containing $n_B = 183$ note events with the 14 different fundamental frequencies listed in the seventh column of Table D1.

Corresponding to the frequencies are the scale steps s_i^j and $\hat{s}_i^j$ and the relabeling

Comment	Uşşak Fundamental Frequencies	Scale Step	Label	Label	Scale Step	Muhayyer Fundamental Frequencies	Comment
lead	199.14	s_7^{-1}	σ_1	$\hat{\sigma}_1$	$\hat{s}_7^{-1}$	202.30	lead
root	225.97	s_1^0	σ_2	$\hat{\sigma}_2$	$\hat{s}_1^0$	226.71	root
	242.32	s_2^0	σ_3	$\hat{\sigma}_3$	$\hat{s}_2^0$	243.74	
	263.78	s_3^0	σ_4	$\hat{\sigma}_4$	$\hat{s}_3^0$	264.06	
dom	297.76	s_4^0	σ_5	$\hat{\sigma}_5$	$\hat{s}_4^0$	298.71	
	342.26	s_5^0	σ_6	$\hat{\sigma}_6$	$\hat{s}_5^0$	343.95	dom
	352.02	s_6^0	σ_7	$\hat{\sigma}_7$	$\hat{s}_6^0$	367.01	
lead	398.28	s_7^0	σ_8	$\hat{\sigma}_8$	$\hat{s}_7^0$	404.59	lead
octave	451.94	s_1^1	σ_9	$\hat{\sigma}_9$	$\hat{s}_1^1$	453.43	octave
*	484.64	s_2^1	σ_{10}	$\hat{\sigma}_{10}$	$\hat{s}_2^1$	487.48	
	527.56	s_3^1	σ_{11}	$\hat{\sigma}_{11}$	$\hat{s}_3^1$	528.13	
dom*	595.52	s_4^1	σ_{12}	$\hat{\sigma}_{12}$	$\hat{s}_4^1$	597.43	
*	684.52	s_5^1	σ_{13}	$\hat{\sigma}_{13}$	$\hat{s}_5^1$	687.91	dom
*	704.04	s_6^1	σ_{14}	$\hat{\sigma}_{14}$	$\hat{s}_6^1$	734.02	

Table D1: Example of cross-makam generation of Uşşak and Muhayyer. All frequencies are given in Hz.

of the corresponding σ_k and $\hat{\sigma}_k$. The comments show important notes in the scales: the leading tone, root, and dominant. For Uşşak and Muhayyer, these align closely; the major difference is the location of the dominant. (These labels play no role in the construction of the sound examples, they are intended to aid in the interpretation of the results.) The four asterisks in the leftmost column indicate the four notes that appear in the performance of the Muhayyer makam that are missing from the Uşşak performance. In order to have a one-to-one mapping between the scale steps of the two makams, it was necessary to fill-in these values as described above. This is why (for instance) s_4^1 is exactly twice s_4^0. The mappings g and g^{-1} can be read directly from the rows of the table, and constructing the sound examples D3 and D4 is now a matter of substituting the desired values into the sequences. The other sound examples are constructed similarly.

Thus each of the synthesized sound examples contains certain aspects derived from makam A and other aspects derived from makam B. In the example, the pitch structure of makam A (Uşşak) is grafted onto the temporal structure of makam B (Muhayyer); the sequence of pitches and their timings is determined by the Muhayyer, but the exact pitches that occur are determined by the Uşşak. We state this concisely (although somewhat inaccurately) by saying that the example contains “perdes from Uşşak” and the “seyir from Muhayyer.”