A Quantitative Comparison of Chrysanthine Theory and Performance Practice of Scale Tuning, Steps, and Prominence of the Octoechos in Byzantine Chant *

Maria Panteli[1], Hendrik Purwins[2,3]
[1]Department of Computer Science University of Cyprus
1 University Avenue, 2109 Aglantzia, Cyprus
[2]Neurotechnology Group, Berlin Institute of Technology
Marchstr. 23, 10587 Berlin, Germany
[3]Sound & Music Computing Group Aalborg University Copenhagen
A.C. Meyers Vænge 15, DK-2450 Copenhagen SV, Denmark

Abstract

Byzantine Chant performance practice is computationally compared to the Chrysanthine theory of the eight Byzantine Tones (octoechos). Intonation, steps, and prominence of scale degrees are quantified, based on pitch class profiles. The novel procedure introduced here comprises the following analysis steps: 1) The pitch trajectory is extracted and post processed with music-specific filters. 2) Pitch class histograms are calculated by kernel smoothing. 3) Histogram peaks are detected. 4) Phrase ending analysis aids the finding of the tonic to align pitch histograms. 5) The theoretical scale degrees are mapped to the empirical ones. 6) A schema of statistical tests detects significant deviations of theoretical scale tuning and steps from the estimated ones in performance practice. 7) The ranked histogram peak amplitudes are compared to the theoretic prominence of particular scale degrees. The analysis of 94 Byzantine Chants performed by 4 singers shows a tendency of the singers to level theoretic particularities of the echos that stand out of the general norm in the octoechos: theoretically extremely large steps are diminished in performance. The empirical intonation of the IV. scale degree as the frame of the first tetrachord is more consistent with the theory than the VI. and the VII. scale degree. In practice, smaller scale degree steps (67-133 cents) appear to be increased and the highest scale step of 333 cents appears to be decreased compared to theory. In practice, the first four scale degrees in decreasing order of prominence I, III, II, IV are more prominent than the V., VI., and the VII..

Keywords: Byzantine Chant, modes, Byzantine Tones, Chrysanthine theory, Octoechos, pitch class profile, echos, computational ethnomusicol-

* *Preprint* of: Maria Panteli and Hendrik Purwins: *A Quantitative Comparison of Chrysanthine Theory and Performance Practice of Scale Tuning, Steps, and Prominence of the Octoechos in Byzantine Chant*, Journal of New Music Research, 42 (3), 2013, pp. 205-221.

ogy, kernel smoothing, non-parametric density estimation, peak picking, tonic detection

1 Introduction

Byzantine Chant is the Christian liturgical song of the Eastern Roman Empire (*Byzantium*) that gradually emerged from the Roman Empire from the 4^{th} century on. Since then up to now, Byzantine Chant has been the dominant liturgy of the Eastern orthodox Christianity. Since the earliest surviving manuscripts with notation of Byzantine Chant dating to the 9^{th} century, and the earliest theoretic accounts of Byzantine Chant dating to the 10^{th} century, several notation systems and theories have been introduced, including the *Chrysanthine* notation and theory, still used in the official chant books of several Christian orthodox churches practicing Byzantine Chant (Levy & Troelsgård, 2011). However, Levy and Troelsgård (2011) refer to medieval treaties on Byzantine Chant: *'most of them were intended for those already proficient in the performance of chant, and they are often imprecise with regard to basic questions of rhythm, ornamentation, the exact tuning of scales'*. Referring to various theoretic accounts on Byzantine Chant, Zannos in (Zannos, 1990) argues that *'none of them can be said to correspond with contemporary empirical study'*. In this paper, it will be investigated to what extent performance practice is in accordance with Chrysanthine theory.

Music theory and performance practice can be compared using Music Information Retrieval (MIR), a relatively young multidisciplinary research field at the cross section of musicology, music psychology, signal processing, and statistics. With these methods, large collections of recorded music can be analysed. The field of computational ethnomusicology (Tzanetakis, Kapur, Schloss Andrew, & Wright, 2007) comprises the computational analysis of ethnic or traditional music styles. This approach is particularly useful for the study of predominantly orally transmitted music traditions such as Byzantine Chant, since manual transcriptions might be subjective or erroneous if created by people with different criteria and musical backgrounds (Toiviainen & Eerola, 2006), in particular with respect to subtle nuances in intonation or temperament. In addition, manual analysis of scores and recordings is a time consuming task, thus restricting studies on relatively small and statistically insignificant music collections.

The present research studies a theoretical model of Byzantine music with developed MIR techniques. Precisely, a theoretical model of Byzantine Chant, namely the Chrysanthine theory, is computationally compared to empirical data extracted from Byzantine recordings. The main analysis tool used is a pitch class profile (Tzanetakis, Ermolinskyi, & Cook, 2002) with fine pitch resolution, extracted from audio recordings with the aid of specifically designed algorithms. These are applied on a large music collection of Byzantine Chant . The overall behavior and consistency of empirical scale tuning, of the steps between consecutive scale degrees, and of the prominence of scale degrees is contrasted to theory through a series of tests and experiments.

This paper is organized as follows. First the Byzantine Tones, the Chrysanthine theory, and the used music corpus are introduced. Then the computational analysis procedure is explained, consisting of pitch detection, pitch histogram smoothing, scale degree alignment, and statistical tests. Results on theory-practice comparison are given with respect to the tuning, scale steps, and prominence of scale degrees. The paper concludes with a discussion of the analysis.

2 The Octoechos and the Chrysanthine Theory

According to Winnington-Ingrams Mode in *Ancient Greek Music* (Winnington-Ingram, 1936) a *'mode is essentially a question of the internal relationships of notes within a scale, especially of the predominance of one of them over the others as a tonic, its predominance being established in any or all of a number of ways: e.g., frequent recurrence, its appearance in a prominent position as the first note or the last, the delaying of its expected occurrence by some kind of embellishment'*.

The tone (singular: *echos*, plural *echoi*) in Byzantine Chant is based on the concept of the mode. According to Mavroeidis (1999) and Thoukididi (2003), an echos is defined by the following five characteristics: 1) the scale degree steps between consecutive scale degrees, 2) the most prominent scale degrees (two or three scale degrees out of which I - III and I - IV are the most reoccurring scale degree pairs), 3) a short introductory phrase that marks the reference tone, 4) the cadences in the middle and at the end of a phrase, and 5) the modulations (alterations) applied to particular scale notes depending on whether they are reached by an ascending or a descending melody.

In Byzantine Chant, the number of echoi and the tuning of their scale degree steps have been discussed during the long history of Byzantine Chant notation and theory that developed through the following three stages: 1) the Palaeo-Byzantine method (10^{th} – 12^{th} century), 2) the middle Byzantine notation (mid-12^{th}– about 1815), and 3) the new (Chrysanthine) notation (from the 1820s) (Levy & Troelsgård, 2011). The latter is attributed to the three teachers Chrysanthos of Madytos, Chourmouzios the Archivist and Gregorios the Protopsaltes. Subject to a reform in the 1880s concerning the tuning of the scale degrees in particular, the Chrysanthine notation method is used in the official chant books of the Greek Orthodox Church up to now (Levy & Troelsgård, 2011). In the 20^{th} century, Karas (1970) suggested an alternative Byzantine notation by re-introducing some old (palaeographic) qualitative signs, reconstructing the interval structure and revising the classification of the modes. However, his theory is controversial among scholars and performers of Byzantine Chant (Angelopoulos, 1986). In this paper, we use the Chrysanthine theory as a reference.

In Chrysanthine theory, the octave is divided into 72 equal partitions, each of 16.67 cents[1] (*morio*, plural: *moria*). The scale degree steps are measured in multiples of a morio. Scale degree steps can be of the size of a semitone (100 cents = 6 moria) and a whole tone (200 cents = 12 moria). But also step sizes between the halftone and the whole tone are frequently used: the *minor* tone (166.67 cents = 10 moria) and the *minimal* tone (133.33 cents = 8 moria) (Thoukididi, 2003).

[1]According to Mavroeidis (1999) the morio of Chrysanthine theory is 17 cents as the 12^{th} part of a Pythagorean whole tone of 204 cents. However, we assume 72-tone equal temperament resulting in a morio of 16.67 cents.

The Chrysanthine theory defines in total eight basic echoi, a system also referred to as *octoechos* ('eight' + 'mode'). These eight modes occur in pairs of *authentic* and corresponding *plagal* modes[2]: *First Authentic, Second Authentic, Third Authentic, Fourth Authentic, First Plagal, Second Plagal, Grave, Fourth Plagal.* The First Authentic/Plagal and Third Authentic/Grave echoi pairs each share the same sequence of scale steps (cf. Table 1). The plagal mode has a different reference tone (tonic) than its authentic counterpart, usually a perfect fifth lower than the one of the authentic mode[3]. Furthermore, both differ in melodic characteristics.

The seven scale degree steps of an echos are constructed in the following way. The octave is divided into a fourth, whole tone, and another fourth. The first fourth (*first tetrachord*) is further subdivided by 2 tones yielding three scale steps. The subdivision proportions are identically repeated for the second fourth (*second tetrachord*). To give an example, the tetrachord of the First echos is defined by the scale step sequence minor - minimal - whole tone ($10 - 8 - 12$ moria) which is repeated after the whole tone step (12 moria) that lies in the middle of the scale (cf. Table 1). Only the Fourth echos deviates from this tetrachord structure.

The scale degree steps may vary according to the chant genre: *Heirmoi* (singular: *Heirmos*) are chants in a relatively fast tempo with each note corresponding to one syllable. *Stichera* (singular: *Sticheron*) are in medium tempo with more than one note corresponding to the same syllable (melisma). *Papadika* (singular: *Papadikon*) are sung in a slow tempo with a phrase of notes corresponding to one syllable (Mavroeidis, 1999; Thoukididi, 2003).

To give an example, some of the echoi such as the Grave consist of different scale degree steps depending on whether the chant type is Heirmos/Sticheron or Papadikon. Similarly, the Fourth authentic, consists of different scale degree steps for chants of Heirmos type and chants of Sticheron/Papadikon type.

Our study is limited to the basic and simplest echos scales[4] (cf. Table 1). We will not consider the fact that scale degree steps of an echos can be modulated (altered) based on the melodic characteristics of a chant or other criteria (cf. Mavroeidis (1999), Thoukididi (2003)).

3 Music Corpus

In the long history of Byzantine music, choirs and various musical instruments have been used in the liturgical ceremonies (Braun, 1980). However, musical instruments are forbidden in Orthodox ecclesiastic music to this day (Thoukididi, 2003). The singing voice, as the main instrument is used solo or in choirs. The corpus of music analysed in this study consists of recorded monophonic chants

[2]The notion of the authentic and plagal modes in Byzantine music should not be confused with the notion of church modes in Roman theory.

[3]The *Grave* shares the same scale degree steps with the Third Authentic as well as the same reference tone.

[4]For the Grave echos the scale of Heirmos/Sticheron chant type is considered, and for the Fourth authentic the scale of Heirmos chant type, since these are the most basic and commonly used scales for the corresponding echoi (Mavroeidis, 1999). Chants of the Heirmoi genre of the Second, Heirmoi of Second Plagal, Papadika of Grave, Stichera and Papadika of Fourth, and Papadika of Fourth Plagal use different scales than the eight basic echoi. Thus they are omitted in this study.

Echos	Chant Type	Step (in Moria) between Scale Degrees						
		I II	II III	III IV	IV V	V VI	VI VII	VII I
First	All chants	10	8	12	12	10	8	12
First Plagal	All chants	10	8	12	12	10	8	12
Second	Stichera/Papadika	8	14	8	12	8	14	8
Second Plagal	Stichera/Papadika	6	20	4	12	6	20	4
Third	All chants	12	12	6	12	12	12	6
Grave	Heirmoi/Stichera	12	12	6	12	12	12	6
Fourth	Heirmoi	8	12	12	10	8	12	10
Fourth Plagal	Heirmoi/Stichera	12	10	8	12	12	10	8

Table 1: The scale structure of the eight modes (octoechos) measured in multiples of a morio. The two tetrachords are indicated.

from the album series of Protopsaltes Georgios Kakoulidis (Kakoulidis, 1999), Protopsaltes Ioannis Damarlakis (Damarlaki, 1999), Protopsaltes Panteleimon Kartsonas (Kartsonas, n.d.), and Protopsaltes Dimitrios Ioannidis (Ioannidis, 2005). From the total of 94 recordings, 13 are in the First Authentic echos, 15 in First Plagal, 6 in Second Authentic, 6 in Second Plagal, 18 in Third Authentic, 9 in Grave, 10 in Fourth Authentic and 17 in Fourth Plagal. Only in a few chants, a drone appears at low voice in the accompaniment. The type of chants collected for each echos follow the specifications in Table 1. The duration of chants in this music corpus ranges between 30 to 220 seconds with mean duration 70 and standard deviation 35 seconds.

4 Computational Analysis Process

The objective of the study is the theory-practice comparison of features of Byzantine scales. To study scale degree pitch, prominence, and steps empirically, pitch modulo octave histograms are investigated. Built on pitch histograms, pitch class profiles have been applied to detect key and tone centres in classical Western music (Purwins, Blankertz, & Obermayer, 2000; Gómez, 2006). Chordia and Rae (2007) adapted the latter approach to Raag recognition. Bozkurt (2008) proposed a method to extract the tuning of a scale, applied to Turkish maqam. Moelants, Cornelis, and Leman (2009) introduced a peak picking heuristics to extract the scale tuning from a modulo octave pitch histogram of African scales. Serrà, Koduri, Miron, and Serra (2011) used pitch class histograms to investigate scale tuning in Hindustani and Carnatic music. In particular, they investigated whether this music follows equal temperament rather than just intonation.

The procedure proposed in this article is summarized in Figure 1. First, the pitch (f_0) trajectory is extracted from each recording. A pitch histogram is computed, compressed into one octave and smoothed. Then peaks are detected in the histogram. With the reference to the tonic, the pitch trajectories and pitch histograms are aligned. From pitch trajectories of recordings of the same echos, the *echos histogram* is computed. Pitch distributions around the peak locations of this histogram are used to determine the empirical scale degree pitches. From the aligned pitch histogram of each recording, peak locations are mapped to theoretical scale degree pitches. The interval sizes between consecutive peak locations are employed to determine the usage of scale steps between consecutive scale degrees. Finally, a sequence of statistical tests is used to compare the estimated empirical scale tuning and scale step sizes with the theoretical ones. The ranked peak amplitudes are compared with theoretic prominence of particular scale degrees.

4.1 Pitch Trajectory via F0 Detection

Algorithmic pitch estimation is usually done assuming a close relationship of the perceived pitch and the fundamental frequency (f_0) of the signal. In the current study, we use the f_0 estimation Yin algorithm (de Cheveigne & Kawahara, 2002). The algorithm is based on the autocorrelation method (Rabiner, 1977) with a number of modifications that improve its performance. The error rate of the Yin frequency estimates depends on the acoustic characteristics of the signal. Considering the melodic characteristics of the analysed music as well as the particularities of the singing voice, the following post processing filters are designed:

1. **Noise:** An aperiodicity threshold θ_n is applied to eliminate erroneous frequency estimates at the noisy parts of the recording. Assuming min max normalization of the aperiodicity variable in the range $[0, 1]$, the aperiodicity threshold is set to $\theta_n = 0.8$.
2. **Silent Gaps:** A loudness threshold θ_s is applied to remove frequencies corresponding to silent and/or quiet parts. Assuming min max normalization of the instantaneous power variable in the range $[0, 1]$, the loudness threshold is set to $\theta_s = 0.05$.
3. **Octave/fifth errors:** To avoid false estimates due to confusion with other harmonics of the fundamental frequency (octave/fifth errors), an octave/fifth correction algorithm is designed. Based on some initial assumptions, the algorithm recalculates an instantaneous frequency value whenever an interval greater than a fifth is found between consecutive time-ordered pitches.

As a result, a trajectory of estimated pitches $p = (p_1, \ldots, p_N)(1 \leq n \leq N)$ is generated for each recording, where n denotes indexes through time-ordered pitches.

4.1.1 Behavioral F0 Trajectory Evaluation Experiment

The efficiency of the f_0 trajectory extraction algorithm as well as the tonic detection Algorithm 2 was evaluated with a behavioral experiment. We decided to perform the evaluation with few expert singers of Byzantine Chant rather than with a large number of non-experts, since, in general, the subtleties of detecting the tonic in Byzantine Chant make the tonic-detection task too difficult for non-experts. As subjects, we chose three accomplished singers of Byzantine Chant. For the stimuli, 20 chant endings were selected, stemming from all 8 main echos types and all singers of the music corpus. The experiment consisted of two parts: the first one for evaluating the estimated pitch trajectory, and the second one for evaluating the estimated tonic (cf. Section 4.4.2). For pitch

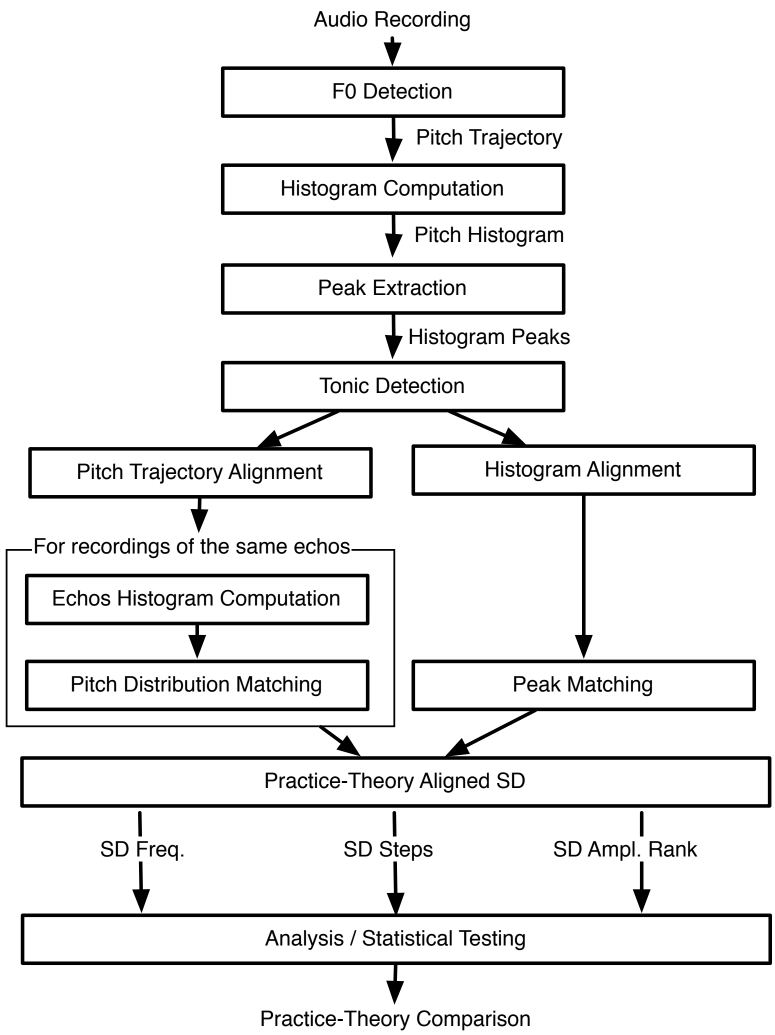

Figure 1: Flow diagram of the analysis process of a recording yielding a theory-practice comparison of the scale tuning, the scale step sizes, and the scale tone prominence (SD=Scale Degree).

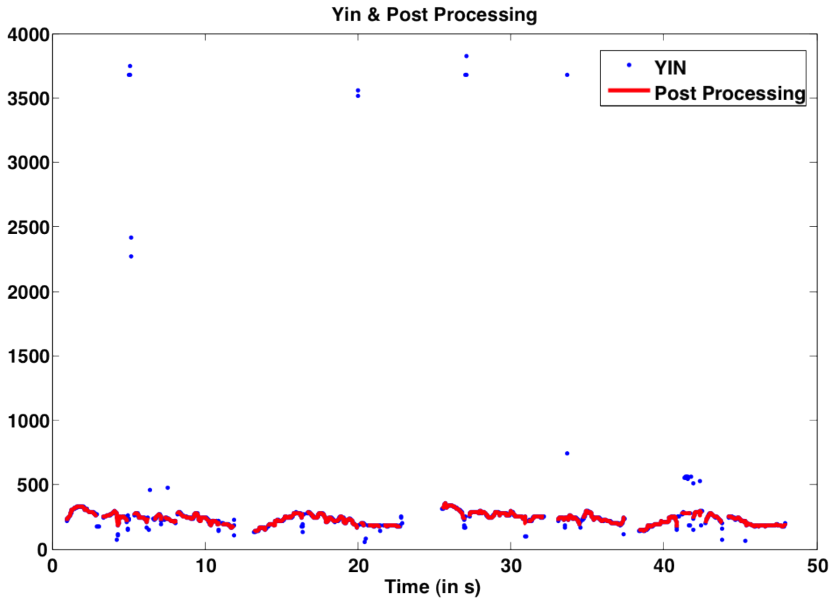

Figure 2: Yin f_0 estimations before and after post processing.

trajectory evaluation, the subjects were asked to first listen to an excerpt of a Byzantine chant and then listen to a synthesized version of the estimated pitch trajectory. Choosing from a scale of 1 to 5, where 1 corresponds to "no match at all" and 5 corresponds to "perfect match", the subjects were asked to evaluate how well the melody of the original matches the melody of the synthesized version. The three experts evaluated the estimated pitches for 20 Byzantine excerpts with an average value of 4.2/5. One melody was evaluated, on average, below 3.5, and this corresponds to a recording with a relatively noisy background. In general, our f_0 detection algorithm works sufficiently well.

4.2 Pitch Histogram via Kernel Smoothing

Pitch histograms are used as an analytic tool for scale estimation (Akkoç, 2002; Bozkurt, 2008; Moelants et al., 2009; Chordia & Rae, 2007). Ideally, the histogram distribution peaks at the most frequently appearing notes of the melody. By definition, a pitch histogram $c = c^{p,b} = (c_1^{p,b}, \ldots, c_k^{p,b})$ partitions the data p into K distinct bins of width h and then counts the number of p observations falling in each bin. This can be expressed as

$$c_k^{p,b} = \sum_{n=1}^{N} q_r\left(\frac{p_n - b_k}{h}\right) \tag{1}$$

where b_k is the centre of the bin k for $k = 1, \ldots, K$, and $q_r(u)$ the *rectangular kernel function* defined as

$$q_r(u) = \begin{cases} 1 & \text{if } |u| \le \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

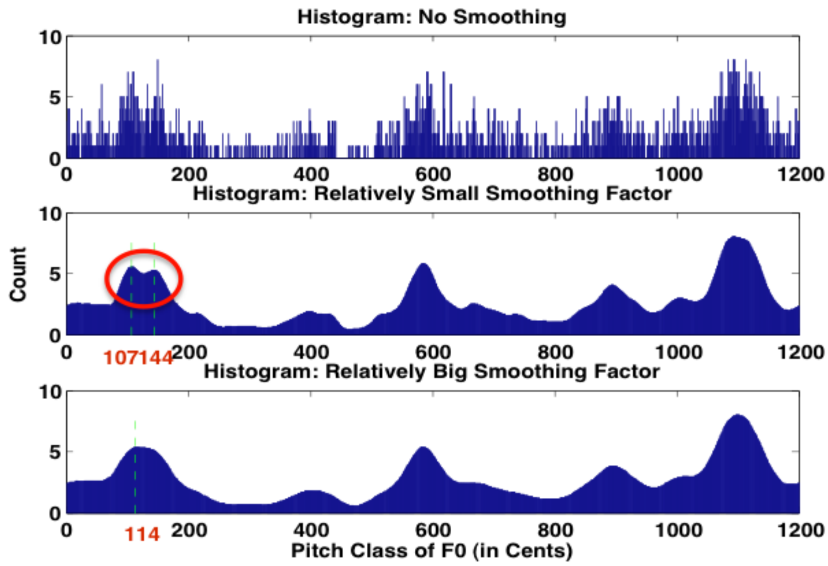

Figure 3: Histogram with no smoothing (rectangular kernel q_r, $h = 6$ cents, *top*), with relatively small smoothing factor (Gaussian kernel, , $h = 12$ cents, *middle*), with relatively large smoothing factor (Gaussian kernel, $h = 18$ cents, *bottom*). The circled peaks of the middle graph differ by less than 4 moria (67 cents) and are merged in the bottom graph thus avoiding an - according to theory - false scale-tone peak.

The choice of h is critical for the subsequent stage of peak picking, since a too high h can eliminate relevant peaks whereas a too small h can create spurious peaks in the pitch histogram. Used in Turkish music theory, the Holdrian comma provides a $K = 53$ equal temperament division of the octave. Based on that division, Bozkurt (2008) yields $h = \frac{1200}{53 \cdot 3}$. For Byzantine Chant, the Chrysanthine theory divides the octave into 72 equal partitions and we multiply this division by 3 arriving at $h = \frac{1200}{72 \cdot 3}$, thereby yielding sufficient bin resolution and robustness. We consider the equally spaced modulo octave *pitch class histogram* defined by

$$c_k^{p,b} = \sum_{n=1}^{N} q_r(u) \tag{2}$$

where

$$u = \frac{(b_k - (p_n - p_0) \bmod 1200) + p_0}{h} \tag{3}$$

and p_0 is the pitch offset.

Although a large h increases the smoothness of the pitch histogram, discontinuities in the histogram remain. These discontinuities are artefacts due to the partitioning of the pitches in a discrete set of predefined bins. A smoothing of the histogram has the effect of simplifying its shape and making it less dependent on the variance of the data from which it is generated.

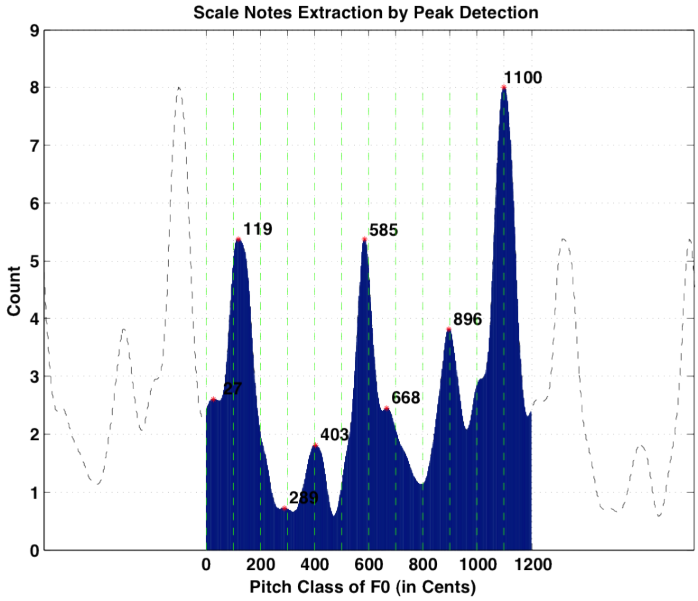

Figure 4: Peak detection applied to the three-histogram copies, where the middle part avoids the peak discontinuities at the edges of the original (single) histogram.

The sharp-edged rectangular kernel function $q_r(u)$ in Equation (2) can be replaced by a smooth *Gaussian kernel function* (Bishop, 2006) yielding the histogram

$$c_k^{p,b} = \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}h} e^{-\frac{(p_n - b_k)^2}{2h^2}} \tag{4}$$

for $k = 1, \ldots, K$.

In the smoothed histogram, the Gaussian kernel replaces each single pitch point by a smooth Gaussian and then adds up all Gaussians across all pitch points. As with the rectangular kernel, the *bandwidth* h determines the smoothness of the histogram. Again there is a trade-off between noise sensitivity at small h and over-smoothing at large h values. Tests with data-driven determination of h (Sheather & Jones, 1991) gave overly smoothed curves eliminating too many peaks. The selection of the appropriate smoothing parameter h has to be guided by the task the histogram is used for. Whereas e.g. for characterizing a singer's ornamentation, a small h, i.e. a detailed histogram may be appropriate, for estimation of scale degree tuning, as one of the objectives of this study, a relatively high smoothing factor is employed to avoid spurious peaks in a too detailed histogram. To determine an adequate h, the following assumption is made: Byzantine theory recognizing 4 moria (67 cents) as the smallest scale degree step and choosing the quartertone ($\delta_{min} = 50$ cents) as the smallest acceptable distance between two histogram peak positions allows for a margin for investigating the deviations between theory and practice. Experimenting with h and looking at the resulting histograms (cf. Figure 3), the assumption is satisfied when h is set to 18 cents.

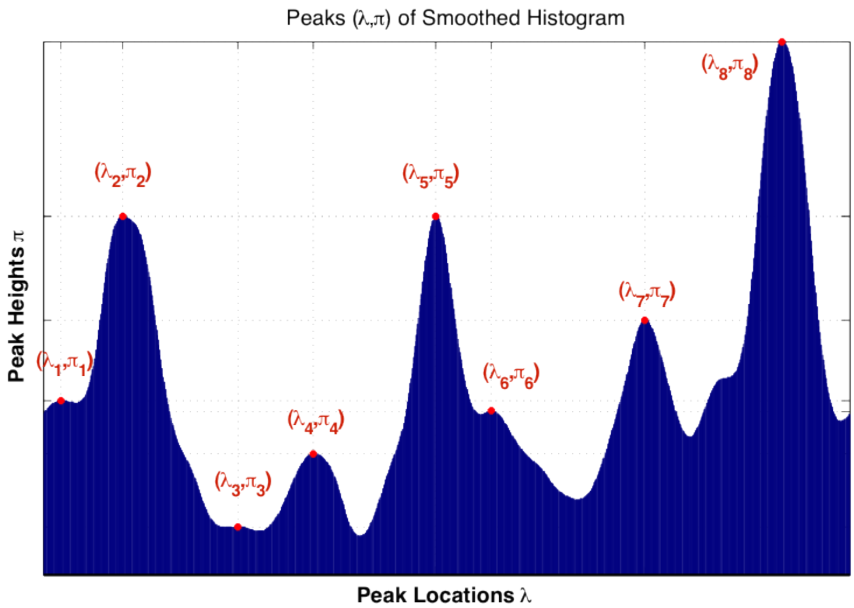

Figure 5: The histogram peaks (λ, π) can be interpreted as the pitches (peak locations λ) and the prominences (peak heights π) of the scale degrees.

4.3 Histogram Peak Extraction

Adjusting the smoothing factor of the histogram is an important step in our approach. The smoothed histogram distribution is expected to peak at exactly seven maxima (the seven scale notes of the melody), thus, eliminating spurious peaks. Peaks that lie closer than the theoretical boundary predicts, are merged (with appropriate amount of smoothing) to a single peak with centre frequency adjusted according to the properties of the histogram distribution. Alternatively, the centre frequencies of the seven scale notes could be estimated via the k-means method applied to the non-smoothed histogram distribution. The 7-means method was tested but failed to identify correctly the seven scale notes due to the significant deemphasis of the scale degrees VI and VII (cf. Figure 9).

To overcome the smoothing artifacts at both borders of the histogram, we copy the histogram three times, pasting it next to each other, yielding the *three-histogram* (cf. Figure 4). From the three-histogram copies, the middle one has accurately smoothed edges due to the continuation beyond the histogram borders, and therefore is the only one considered for further processing.

Using a peak extraction algorithm, from the smoothed pitch class histogram, c, U peaks (λ, π) can be detected consisting of *peak locations* $\lambda = (\lambda_1, \ldots, \lambda_U)$ and *peak heights* $\pi = (\pi_1, \ldots, \pi_U)$ (cf. Figure 5). Moelants et al. (2009) proposed a couple of heuristics to be used in peak picking, such as the consideration of size and height of peaks and intervals between peaks. Here Algorithm 1 is proposed. The number U of peaks to be detected and the minimum peak distance δ_{min} have to be determined beforehand.

From the histogram $c= (c_1, \ldots, c_K)$, the peak picking algorithm (cf. Algorithm 1) iteratively chooses the peak location $\lambda_u = b_{k_u}$ with maximum peak height $\pi_u = c_{k_u}$, then removing the potential location candidates in a $\pm\delta_{min}$ neighborhood around the selected peak position, to chose the next peak until U peaks are picked. U is determined as follows: According to Byzantine theory each echos has at least 7 scale notes. In addition, Byzantine theory knows of note alterations. To account for further intonation variants in practice, we set $U = 12$. $\delta_{min} = 50$ cents is chosen as the minimum neighborhood.

Algorithm 1 Peak picking algorithm.

Variables
Minimum peak distance δ_{min}
Number of peaks U
Pitch class histogram $c_1, \ldots, c_k$
Bins $b = (b_1, \ldots, b_K)$
Initialization
$\Gamma_1 = (1, 2, \ldots, K)$
For $u = 1$ **to** U
$k_u = \text{argmax}_{k \in \Gamma_u} c_k$
$\Gamma_{u+1} = \Gamma_u \setminus \{\nu \colon |\nu - b_{k_u}| < \delta_{min}\}$
$\lambda_u = b_k, \pi_u = c_k$
End For

4.4 Practice-Theory Aligned Scale Degrees

4.4.1 Tonic Detection

It would be advantageous to compare pitch histograms in a way that is invariant to pitch transpositions. Scale degrees could then be compared to their position relative to the tonic rather than to their absolute pitch (Purwins, Blankertz, Dornhege, & Obermayer, 2004). This way, two instances of the same echos could be compared, even if their tonics are different. To make the first bin b_1 correspond to the tonic, the histogram has to be circularly (modulo K) shifted by $-p_0$, with p_0 being the bin of the tonic. Gedik, Ali C and Bozkurt (2008) applied kernel smoothing to each of the makam scale prototypes. Then they calculated the cross correlation between all scale prototypes and all circularly shifted versions of a histogram. The pitch shift $-p_0$ that gives the maximum cross-correlation corresponds to the estimated tonic and is used to circularly shift the latter histogram. Exploiting the scale structure of the octoechos for tonic detection would be likely to improve the tonic estimate. However, in our approach we will not use this information at this point, because we want to minimize the used musical knowledge and the resulting bias in the subsequent testing.

In order to compare empirical profiles with the theoretical tonic and scale degrees, we aim at increasing the reliability of detecting tonic p_0 by incorporating more musical knowledge into the tonic (p_0) detection algorithm, following an idea in (Bozkurt, 2008). According to theory, the tonic of a Byzantine echos is stated at the end of the phrase in most of the cases. Although there are exceptions to this general rule in the majority of the chants this rule applies and we will build our analysis on this assumption. We present a novel tonic detection algorithm that computes the pitch of the last phrase note from the onset and frequency information. A couple of heuristics are implemented to increase the noise robustness of the method. The algorithm makes the following assumptions:

1. The last note in the recording is the last note of the melodic phrase.
2. The final phrase note lasts for at least half a second.

Assumption 1 is fulfilled if the complete song is recorded. Presegmentation however is employed to standardise the number of similar musical phrases in the excerpts. Assumption 2 is generally met in religious singing in Byzantine Chant.

Onsets in the estimated pitch $p = (p_1, \ldots, p_N)$ are detected based on the onset detection function

$$d_n = \frac{1}{r}\left(\sum_{m=n+\frac{r}{2}}^{n+\frac{3}{2}r-1} p_m - \sum_{m=n-\frac{3}{2}r+1}^{n-\frac{r}{2}} p_m\right), (\frac{3}{2}r \leq n \leq N - \frac{3}{2}r + 1), \qquad (5)$$

which takes the difference d_n of sums of an even number of r instantaneous pitches left and right from pitch p_n.

To give a short explanatory example, apply Equation 5 to $r = 10$ and $n = 15$ and

$$p_m = \begin{cases} 10 & \text{for} \quad m < 15 \\ 20 & \text{otherwise} \end{cases} . \qquad (6)$$

Then $d_n = \frac{1}{10}(\sum_{m=20}^{29} p_m - \sum_{m=1}^{10} p_m) = 10$.

Applying Algorithm 1 for extracting $U = 100$ peaks gives onset candidate locations $o'' = \{o''_1, \ldots, o''_U\}$ from which the ones with an onset detection function d_q above pitch threshold $\theta_f \cdot \max(o'')$ are considered: $o' = \{o_q \in o'' \colon d_q > \theta_f \cdot \max(o'')\}$. Analogously, onsets o^* are detected based on the log energy, computed from the output of YIN algorithm, and energy threshold θ_e, instead of pitches p. Onsets due to pitch as well as energy are united to $o = o' \cup o^*$. From these, the ones at least a minimal inter onset interval Δ_{min} apart from their predecessor are extracted: $o = \{o_q \in o \colon o_q - o_{q-1} > \Delta_{min}\}$. In the sequel, the following parameter settings are used: $r = 32, \theta_f = 0.1$, energy threshold $\theta_e = 0.005$ and $\Delta_{min} = 0.1$s.

The accuracy of detecting the tonic from the musical phrase ending is limited by the characteristics of the analysed music and particularly the frequent use of vibrato on the last note. For example, in extreme cases vibrato ranges up to ± 2 semitones and false onsets due to pitch variation are detected on frames with instantaneous frequency that differs more than a semitone from the actual pitch of the last note (cf. Algorithm 2). Additionally, onsets due to energy variation fail to detect the attack of the last note if melismatic ornamentation is employed at the phrase end, e.g. if the singer softly introduces the final note without changing the syllable and without a noticeable energy variation.

Empirical tests showed that for our music collection the onset detection algorithm usually detects up to three onsets within the last note, as a result of energy and pitch variations. The algorithm implemented takes into account this fact, and compensates pitch inaccuracies with a final alignment of the tonic to the closest histogram peak (cf. Figure 6).

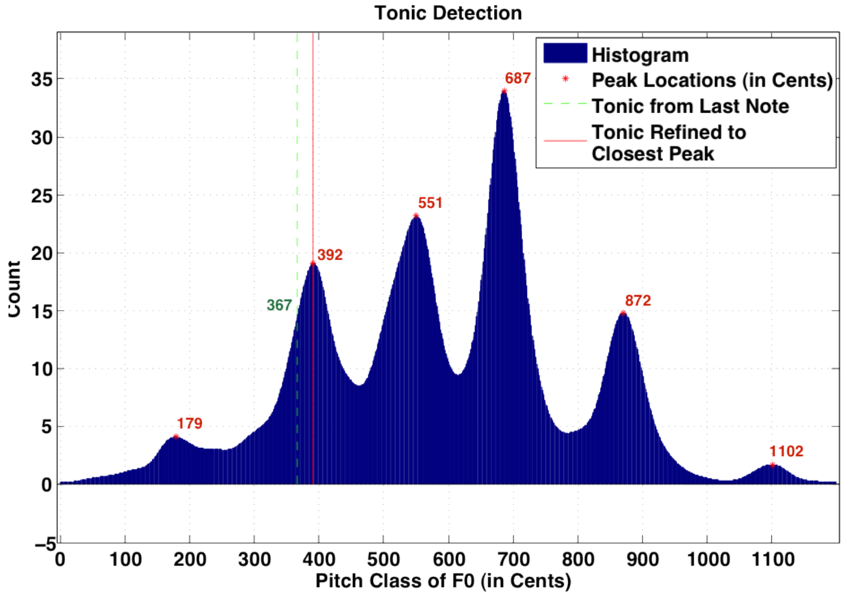

Figure 6: The tonic detected from the last phrase note may deviate from the empirical scale degree pitches (locations of histogram peaks) due to vibrato. With the post processing step of tonic refinement to the closest histogram peak this inaccuracy is resolved.

Algorithm 2 Tonic refinement algorithm.

Preprocessing

Detect the last three onsets o_{o-2}, o_{o-1}, o_o

Pitch trajectory p

Get the three index sets $\Pi_{o-2}, \Pi_{o-1}, \Pi_o$ of the pitches p corresponding to the previous inter-onset-intervals.

Last Note Detection

IF $\max_{0 \leq i \leq j \leq 2 |p_{o-i} - p_{o-j}|} < 100$ cents

(Small vibrato range or no vibrato)

Calculate tonic as average of instantaneous frequencies of frames of three inter-onset-intervals $p_o = \frac{1}{|\Pi|} \sum_{n \in \Pi} p_n, \Pi = \cup_{0 \leq i \leq 2} \Pi_{o-i}$

ELSEIF the last two frequencies lie within a semitone $|p_o - p_{o-1}| < 100$ cents

(Vibrato range larger than a semitone or first of the three onsets is false)

Calculate tonic as average of instantaneous frequencies of frames of last two inter-onset-intervals $p_o = \frac{1}{|\Pi|} \sum_{n \in \Pi} p_n, \Pi = \cup_{0 \leq i \leq 1} \Pi_{o-i}$

ELSE

(Vibrato range larger than a semitone or first two onsets are false)

Calculate tonic p_o as average of instantaneous frequencies of frames of the last half-second of the melody

END

Postprocessing

Reset tonic estimate to highest/closest peak location in histogram:

Get right (λ_1, π_1) and left (λ_2, π_2) location/height peak pair closest to p_o

IF $|\lambda_1 - \lambda_2| < 100$ cents **THEN**

$p^* := \text{argmax}_{1 \leq i \leq 2}(\pi_i)$

ELSE

$p^* := \text{argmin}_{1 \leq i \leq 2}(|p_o - \lambda_i|)$

END

$p_o := \lambda_{n^*}$

4.4.2 Behavioral Tonic Detection Evaluation Experiment

The efficiency of the tonic detection algorithm was evaluated with a behavioral tonic-finding experiment jointly with the f_0 trajectory evaluation experiment. For further details cf. Section 4.1.1. Programmed in a Matlab interface, the subject heard a chant ending and then was asked to move the slider "so that the tonic produced represents the tonic of the original song". The slider controlled the pitch of a sine tone and was moved until the subject agreed that the sine tone would represent the tonic of the excerpt. The subjects could repeat listening to the stimulus and adjusting the tonic pitch as often as they wanted before moving on to the next excerpt.

For 18 of the 20 stimuli, the subjects' responses' standard deviation was below 42 cents. Only for 2 excerpts, one expert disagreed with the other experts, resulting in a larger standard deviation of 189 and 262 cents among the experts. For two out of the 20 excerpts the computed tonic differed greatly from the mean of the subject's response distribution. One recording of the First Plagal echos and one recording of the Third echos are detected with a wrong tonic. These are exactly two of the few exceptions in which the final note of the chant does not end on the tonic. This occurs for instance when the current chant anticipates the following chant in the course of the Byzantine liturgy. However, since these cases are relatively rare, we will see that they do not distort the results of the analysis to an important extent. Cf. the discussion of these results in Section 6.

4.5 Practice-Theory Comparison

4.5.1 Scale Degree Assignment

The objective is to assess the deviation between scale degree tuning of a particular echos according to theory and scale degree tuning in musical practice. The normalized *theoretical scale degree pitches* are defined as a set of scale degree pitches (locations) $\nu^\theta = (\nu_1^\theta, \ldots, \nu_L^\theta)$, with normalization $\nu_0^\theta = 0$ (in cents) and $L = 7$.[5] For all recordings $1 \leq m \leq M_i$ of echos i, the pitch histogram $c^i = c^{p,b}$ is built from the pitches p of these recordings and bins

$$b = (0, \frac{1200}{216}, \frac{2 \cdot 1200}{216}, \ldots, \frac{215 \cdot 1200}{216}). \tag{7}$$

From histogram c^i, peaks (λ, π) with pitches $\lambda = (\lambda_1, \ldots, \lambda_U)$ and amplitudes $\pi = (\pi_1, \ldots, \pi_U)$ are selected via Algorithm 1 (peak picking). From (λ, π), the empirical scale degree pitches are estimated. Then only those peaks (λ_u, π_u) are selected that correspond to a theoretic scale note in ν^θ. For each theoretical scale degree pitch ν_l^θ, the closest pitch λ_u is chosen. If this way, one pitch λ_u is associated with several theoretical scale degree pitches ν_l^θ, pitch λ_u is assigned only to the closest theoretical scale degree pitch ν_l^θ. Then the pitches λ_u are selected that lie within a $\Delta_A = 150$ cents distance from their assigned theoretical scale degree pitch μ_l^θ. If two or more pitches λ_u fulfill this condition, the peak with highest associated amplitude π_u is selected: Formally, the estimated scale degree pitches $\hat{\nu} = (\lambda_{t_1}, \ldots, \lambda_{t_{L'}})$ are defined by $t_l = \arg\max_{|\lambda_t - \nu_l^\theta| \leq \Delta_A}(\pi_t)$.

[5] Exceptions with $L = 8$ theoretical scale degree pitches exist in some variations of echoi not included here.

If no pitch λ_u lies within the range Δ_A around the theoretical pitch ν_l^θ, t_l is not defined and the estimated scale $\hat{\nu} = (\hat{\nu}_1, \dots, \hat{\nu}_{L'})$ has less degrees than the theoretical scale ν^θ, i.e. $1 \leq L' < L$.

From the estimated empirical scale degrees $\hat{\nu}$ and for all recordings of the same echos, consider N_l pitches

$$\mathbf{p_l} = (p_{l,1}, \dots, p_{l,N_l}) \tag{8}$$

that lie within a 2-moria distance of the estimated scale degree pitch $\hat{\nu}_l$ of scale degree l, selected from the instantaneous pitch trajectory $p = (p_1, \dots, p_{N_i})$, consisting of N_i pitches of all recordings of a particular echos i. The 2-moria distance defines half the size of the smallest theoretical scale interval (cf. Table 1). This is done to apply statistical testing to check whether pitches $\mathbf{p_l}$ are drawn from a distribution with mean equal to the theoretical scale degree pitches ν_l^θ (see details below).

Now let us explain how to calculate the *scale steps* $\delta = (\delta_1, \dots, \delta_L)$ between consecutive scale degrees of one echos. Instead of calculating the histograms for all the recordings of one echos all together as done in the previous paragraphs, we now calculate the histograms individually for each recording $1 \leq m \leq M_i$ of echos i: $c^{m,i} = c^{p^{m,i},b}$ (Equation 3) where $p^{m,i}$ is the instantaneous pitch trajectory of recording m of echos i and b are the 216 bins per octave (Equation 7). From histogram $c^{m,i}$, peaks (λ, π) with pitches λ and amplitudes π are selected via Algorithm 1 (peak picking). From the peaks (λ, π) the estimated scale degree pitches $\hat{\nu} = (\nu_1, \dots, \nu_{L'})$ are calculated. From the estimated scale steps $\hat{\nu}$, the *scale steps* $\delta = (\delta_1, \dots, \delta_L)$ can be calculated as the difference between the lth and the $l-1$th estimated scale degree pitches $\delta_l = \hat{\nu}_l - \hat{\nu}_{(l-2)\text{mod}L+1}$. The scale steps are the intervals between consecutive scale degrees. From the peak amplitudes $\pi_l^{m,i}$ associated with the empirical scale degree pitches $\nu_l^{m,i}$, we define the *average scale degree amplitude* by

$$\bar{\pi}_l = \frac{1}{\sum_{i=1}^{I} M_i} \sum_{i=1}^{I} \sum_{m=1}^{M_i} \pi_l^m \tag{9}$$

where M_i are the number of recordings of echos $1 \leq i \leq I$. This concept of average scale degree amplitudes indicates prominences of scale degrees and is similar to Krumhansl's probe tone ratings, constant Q profiles or harmonic pitch class profiles but it generalizes to other than 12-equal temperament tunings (Purwins et al., 2000).

4.5.2 Statistical Testing

To assess the deviation between theoretic and empirical scale degree pitches and steps, we apply a chain of tests as an analytical instrument. As a first step, the Shapiro-Wilk test is applied, to determine whether the pitches $\mathbf{p_l}$ are normally distributed around the estimated scale degree pitches $\hat{\nu}_l$ across all M_i instances of the same echos i. If the p value is above significance level $\alpha = 0.05$, we assume normal distribution and apply the t-test to $\mathbf{p_l}$. In case the Shapiro-Wilk rejects the normality hypothesis, the Wilcoxon signed-rank test is applied instead. The hypothesis to be tested by the Wilcoxon signed-rank test is the following. For echos i, the l-th estimated scale degree pitches $\mathbf{p_l}$ are derived from a distribution with median equal to the l-th theoretical scale interval ν_l^θ. We perform $n = 48$ (6 scale degree pitches[6] times 8 echos types) tests. In multiple testing, there is an increased chance to falsely reject a hypothesis. In order to counterbalance this effect, we apply the Bonferroni correction (Dunn, 1961). To reject a hypothesis with a significance level of $\alpha = 0.05$, we determine the Bonferroni corrected threshold for the p values of the individual $n = 48$ tests as: $\alpha_B = \frac{\alpha}{n} = \frac{0.05}{48} \approx 0.001$. If $p < \alpha_B$, we reject the null hypothesis and conclude that the theoretical scale degree pitch $\nu_l^{\theta,i}$ deviates significantly from the empirical scale degree pitch $\hat{\nu}_l^i$ for echos i.

In order to determine if some empirical scale degree pitches deviate from theory more than others, we calculate the differences between theoretical scale degree pitch and the N_m instantaneous scale degree pitch $p_{l,n}^{m,i}$,

$$\Delta_{l,n}^{m,i} = \nu_l^{\theta,i} - p_{l,n}^{m,i} \tag{10}$$

for recordings $m = 1, \ldots, M_i$ of echos i, and scale degree $l = 1, \ldots, L$. For each scale degree l of echos i, we calculate the *mean scale degree difference* defined by

$$\bar{\Delta}_l^i = \frac{1}{\sum_{m=1}^{M_i} N_m} (\sum_{m=1}^{M_i} \sum_{n=1}^{N_m} \Delta_{l,n}^{m,i}) \tag{11}$$

. To assess practice-theory deviation of scale degree pitches across all echoi, the *mean absolute scale degree difference*

$$\bar{\Delta}_l = \frac{1}{I} \sum_{i=1}^{I} |\bar{\Delta}_l^i| \tag{12}$$

is computed from the mean deviations for each scale degree l,($1 \leq l \leq L'$), and across all echoi i, ($1 \leq i \leq I$).

Another objective of this study is to assess the deviation between theoretic scale steps and scale steps in practice. For each of the *theoretical scale steps* $\delta^\theta = (\delta_1^\theta, \ldots, \delta_Q^\theta) = (67, 100, 133, 167, 200, 233, 333)$ (in cents) corresponding to the $Q = 7$ steps of $4, 6, 8, 10, 12, 14$, and 20 moria we aim at estimating the *empirical scale steps* $\delta^\pi = (\delta_1^\pi, \ldots, \delta_Q^\pi)$. We define the set of all scale steps

$$\mathcal{N} = \{\delta : \exists l \in \{1, \ldots, L\}, m \in \{1, \ldots, M\} \text{ so that } \delta = \delta_l^m\}, \tag{13}$$

for scale degree $l = 1, \ldots, L$ and recordings $m = 1, \ldots, M$. Then we collect the scale steps δ_l^m that are within a ± 1 moria (16.67 cents) neighborhood $\mathcal{N}_q$ around the q-th theoretical scale step δ_q^θ, i.e.,

$$\mathcal{N}_q = \{\delta : \exists l \in \{1, \ldots, L\}, m \in \{1, \ldots, M\} \text{ so that } \delta = \delta_l^m \text{ and } |\delta - \delta_q^\theta| \leq 16.67\}. \tag{14}$$

Note that for scale degree $q = Q$ of 333 cents, we define a neighborhood of ± 4 moria around the theoretical value to account for the relatively large distance between this and the previous theoretical interval at 233 cents.

[6]The scale degree pitches are normalized with respect to scale degree I, therefore only 6 scale degree pitches remain.

The empirical scale steps δ^{π} can be approximated by the *estimated empirical steps* $\hat{\delta} = (\bar{\delta}_1, \ldots, \bar{\delta}_Q)$, defined by the empirical means

$$\bar{\delta}_q = \frac{1}{|\mathcal{N}_q|} \sum_{\delta \in \hat{\mathcal{N}}_q} \delta \tag{15}$$

for $q = 1, \ldots, Q$. To assess the significance of theory-practice deviations in the sizes of the scale steps, first the Shapiro-Wilk test on normality is performed. If the hypothesis of normality is rejected, the Wilcoxon signed-rank test is applied. Otherwise, the t-test is employed. For the Wilcoxon signed-rank test, the null hypotheses to be tested are that the δ in the set $\mathcal{N}_q$ around the theoretical scale steps δ_q^{θ} is drawn from a distribution with median δ_q^{θ}. Since we perform 7 tests for $Q = 7$ scale steps, a significance level of of $\alpha = 0.05$ corresponds to a corrected Bonferroni threshold of $\alpha_B = \frac{\alpha}{7} \approx 0.007$ (cf. Table 3).

We also construct the *scale step histogram* using the notation for the set $\mathcal{N}$ all scale degree steps (Equation 13):

$$d = c^{\mathcal{N},b} \tag{16}$$

with an octave division of 216 bins according to Equation 7. This histogram is then smoothed with a moving average filter of 3 bins length (cf. Figure 7).

The prominence of empirical scale degree pitches is analysed via two approaches. First, to investigate whether some scale degrees are particularly emphasized in Byzantine Chant, which ones are these and to what degree this happens, the average scale degree amplitude $\bar{\pi} = (\bar{\pi}_1, \ldots, \bar{\pi}_L)$ averaged across all recordings of all echoi is considered. Also, the average scale degree amplitude of Byzantine echoi are compared to a related music tradition, the Turkish makams.

Another aim is to assess the deviation between the prominent scale degrees in theory and practice. As defined in theory, the set of the most prominent scale tones is one of the elementary characteristics of the echos. This set usually consists of 2 or 3 scale degrees and it may be differently defined in the sub-categories of chants (Heirmoi etc.) of the same echos (cf. Table 4). For comparison, the *echos average scale degree amplitude*

$$\bar{\pi}_l^i = \frac{1}{M_i} \sum_{m=1}^{M_i} \pi_l^m \tag{17}$$

is calculated for each echos i (cf. Figure 8). The three scale degrees $(l_{t_1}, l_{t_2}, l_{t_3})$ with the highest amplitudes $\bar{\pi}_l^i$ are taken as the set of the most prominent scale degrees of the empirical data (cf. Table 4).

5 Results

In this section, we present our findings referring to three particular aspects of pitch histograms in Byzantine echoi in performance practice, namely the tuning of scale degrees, the interval sizes between consecutive scale degrees (steps), and the prominence of scale degrees. The smoothed histograms for all echoi can be found in Figure 9.

5.1 Tuning of Scale

The tuning of the Byzantine scales is investigated by comparing the pitches of empirical and theoretical scale notes. A series of statistical tests is employed to determine for which scale degrees practice and theory of tuning deviate. For all recordings m of a given echos i, the histogram c across all instantaneous pitches p is calculated. From that, the echo-specific estimated scale degree pitches $\hat{v}$ are calculated. For each estimated scale degree pitch $\hat{v}_l$ of scale degree l, all instantaneous pitches $\mathbf{p_l}$ are collected that lie within a ± 2 moria range around $\hat{v}_l$. For each scale degree pitch and each echos the Shapiro-Wilk test on normality is performed on $\mathbf{p_l}$. The results show that for all scale degrees and all echoi the normality hypothesis is rejected, given significance level $\alpha = 5\%$. The Wilcoxon Signed-Rank Test (W-test) with significance level $\alpha = 0.05\%$, corresponding to a corrected Bonferroni treshold of $\alpha_B = 0.001\%$, is therefore applied to test the null hypothesis that the median of the empirical pitches is the same as the theoretical pitch of a particular scale degree in a particular echos.

Table 2 reveals that the majority of empirical scale degrees of all echoi differ significantly from theory, except for the scale degrees V of First, VII of First Plagal, VI of Second Plagal, and II of Third and Grave. The Second, Fourth and Fourth Plagal echoi have all their empirical scale degrees significantly deviating from theory. In the sequel, we will discuss in detail significant theory-practice deviations of more than 2 moria (cf. Table 2). The VI. scale degree of the First and First Plagal has a relatively large mean scale degree difference (Equation 11) value with negative sign, i.e., the empirical scale degree pitch is smaller than the theoretical one. This could be due to the fact that for these echoi the VI. scale degree is diminished when the melody is descending, according to theory. Other large negative mean scale degree differences appear for the VII. scale degree of Second Authentic as well as the III. and VII. scale degrees of Second Plagal. These scale degrees are reached by relatively large theoretical scale degree steps; the VII scale degree of Second Authentic is reached by the VI-VII scale step of 14 moria (233 cents) whereas the III. and VII. scale degrees of Second Plagal are both reached by a scale step of 20 moria (333 cents, cf. Table 1). The empirical scale degrees are smaller than the theoretical ones, i.e. in practice, the theoretically largest scale degree steps (14 and 20 moria) tend to be diminished. In practice, in the Fourth echos, the V is significantly diminished compared to its theoretical scale degree pitch. This could be related to the fact that the Fourth is the only echos, in which the first tetrachord is extended by two moria up to 32 moria in comparison to the other echoi which are based on a first tetrachord of 30 moria. An interpretation could be that the singer tends to diminish the abnormally high tetrachord pitch of this echos by diminishing the adjacent scale degree, i.e., the V. In Fourth, also scale degree step VII-I (theoretically 10 moria) tends to be diminished towards the more common step VII-I of 8 moria.

For scale degree pitches, the question is addressed, how much practice deviates from theory across all echoi. For all chants of all echoi, the mean absolute scale degree differences (Equation 12) are calculated between empirical and theoretical scale degree pitches. It is investigated, whether the absolute scale degree difference is less for some empirical scale degrees in comparison to other scale degrees. For scale degrees II-VII the mean absolute scale degree differences

Echos	**Feature**	**Scale Degrees**					
		II	III	IV	V	VI	VII
First (13)	p (W-test)	0	0	0	0.0032	0	0
	Mean SD Difference	−16.65	11.25	10.45	0.84	−43.89	5.79
First Plagal (15)	p (W-test)	0	0	0.0001	0	0	0.8767
	Mean SD Difference	28.06	−5.65	−0.91	4.98	−50.06	−0.06
Second (7)	p (W-test)	0	0	0	0	0	0
	Mean SD Difference	5.79	−5.56	−6.55	10.36	20.95	−49.64
Second Plagal (7)	p (W-test)	0	0	0	0	0.1592	0
	Mean SD Difference	11.20	−44.32	5.06	10.79	0.47	−138.47
Third (18)	p (W-test)	0.0048	0	0	0	0	–
	Mean SD Difference	0.45	−21.70	−11.19	−5.48	−11.16	–
Grave (8)	p (W-test)	0.0233	0	0	0	0	–
	Mean SD Difference	−0.44	−27.81	−11.05	−17.42	−10.53	–
Fourth (11)	p (W-test)	0	0	0	0	0	0
	Mean SD Difference	5.27	5.58	−5.21	−39.14	11.13	38.85
Fourth Plagal (16)	p (W-test)	0	0	0	0	0	0
	Mean SD Difference	22.06	11.17	16.63	17.86	17.00	22.25

Table 2: Significance of the mean scale degree difference between practice and theory for all echoi (number of instances in brakets). Since all echoi are aligned to pitch 0 for scale degree I, only II-VII are shown. For scale degrees II-VII the p-value of the test statistic of the Wilcoxon Signed-Rank test (W-test) and the mean scale degree difference (in cents, Equation 11) between practice and theory are indicated. Zero p-values correspond to values smaller than 10^{-5}. Practice-theory differences (cf. Table 1) greater/smaller than ±2 moria (33 cents) are colored and discussed in the text. For the VII. scale degree of the Third and the Grave echoi, no peaks were detected in the corresponding echos histogram (cf. Figure 9), hence no further analysis was considered (indicated with '-' symbol).

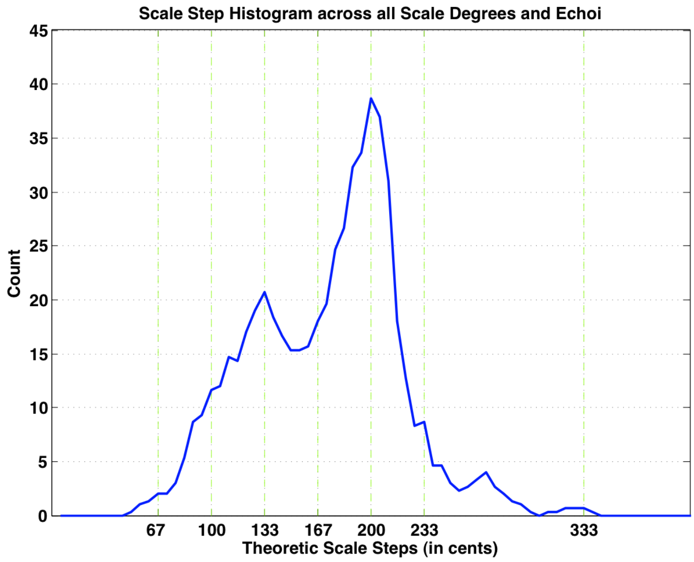

Figure 7: Scale step size histogram computed with 216 bin resolution and smoothed using a moving average with a filter length of 3 bins .

are $\bar{\Delta} = (11.2, 16.6, 8.4, 13.4, 20.7, 42.5)$. The practice-theory deviation for scale degree IV is significantly less than for the VI. and the VII. scale degree. The stability of the fourth can be related to the role of the fourth as the frame of the first tetrachord building block of the scale. From Figure 9 one can infer that the pitch content of the echoi is mostly concentrated in the first 4 scale degrees that lie within the first tetrachord range.

5.2 Scale Steps

The theoretical scale steps are compared to the estimated empirical steps. The questions addressed are: 1) Which intervals are most frequently used in Byzantine music? 2) Which ones deviate significantly between theory and practice?

In Figure 7, the smoothed scale step histogram $d = c^{\mathcal{N},b}$ is shown, based on the scale degree steps between all L consecutive scale degrees in all I echoi and the bins $b = (0, \ldots, \frac{71}{216})$. A moving average filter of 3 bins length is used for smoothing. The histogram displays the frequencies of occurrence of these scale steps in practice. The most frequently used scale steps in Byzantine music are the whole tone (with theoretical size of 200 cents) and the minimal tone (133 cents), followed by the minor tone (167 cents) and the semitone (100 cents). The 14-moria (233 cents) theoretical scale step occurs also with significant frequency whereas the smallest and largest theoretical scale steps (67 and 333 cents respectively) are not used so often.

To assess the significance of theory-practice deviations in the sizes of the scale steps, we calculate the set of empirical scale steps $\mathcal{N}_q$ (Equation 14) around the $Q = 7$ theoretical scale step $\delta_q^\theta = (4, 6, 8, 10, 12, 14, 20)$. First the Shapiro-Wilk test is performed in order to determine whether $\delta \in \mathcal{N}_q$ is normally

Theoretical scale Steps δ_q^θ (in moria)	4	6	8	10	12	14	20
δ_q^θ (in cents)	66.7	100	133.3	166.7	200	233.3	333.3
Estimated Empirical Steps $\hat{\delta}_q$ (in cents)	81.4	107.5	136.8	174.6	201.0	231.7	277.1
Deviation from Theory $\hat{\delta}_q - \delta_q^\theta$ (in cents)	14.7	7.5	3.5	7.9	1.0	-1.6	-56.2
p (Wilcoxon test)	**0.0007**	**0**	**0.0028**	**0**	0.1024	0.1957	**0**

Table 3: Scale degree step size analysis. The estimated empirical steps (Equation 15) are compared to the theoretical scale steps. Scale degree step sizes significantly different from theory are highlighted in boldface. Zero p-values correspond to values smaller than 10^{-5}. It appears that the first four scale degree steps differ significantly from theory. The largest absolute theory-practice differences can be found for scale degree step sizes of 4 and 20 moria.

distributed for $1 \leq q \leq Q$. With a significance level of $\alpha = 5\%$ the normality hypothesis is rejected for 6 out of the 7 scale steps and the Wilcoxon signed-rank test is applied to test the null hypothesis that $\mathcal{N}_q$ has a mean equal to the theoretical scale step δ_q^θ at the $\alpha = 0.05$ significance level corresponding to a corrected Bonferroni threshold of $\alpha_B = 0.007$. In Table 3, results show that theoretic steps of 67 cents, the semitone (100 cents), the minor (133.3 cents), and minimal tone (166.7 cents), and the interval of 333 cents differ significantly from practice. On the other hand, for the whole tone and the interval of 233 cents there is no significant difference between empirical and theoretical steps. The smallest and largest scale steps (67 and 333 cents respectively) display the largest difference between theory and practice. For the majority of the steps, the estimated empirical steps $\hat{\delta}_q$ are larger than the theoretical steps δ_q^θ. The two largest theoretical intervals (233 and 333 cents respectively), are diminished in practice.

5.3 Prominence of Scale Degrees

The prominence of scale degrees for all Byzantine recordings is studied via the average scale degree amplitudes $\bar{\pi}_l$ and compared to the average scale degree amplitudes of Turkish makams in Figure 8. Results reveal that the most prominent scale degrees in Byzantine Chant are the I, III, II, IV in descending order with I (the tonic) having the highest amplitude. From Table 4, we can see that in the shown echoi / chant genres according to theory, the most prominent scale degrees are I, III, IV, II in descending order starting from I. Only the order of II and IV is exchanged when comparing theory to practice. Melodic concentration on mainly the first four scale degrees is further supported by the short pitch range of Byzantine melodies and the importance of the tetrachordal entity as a pitch frame in scale and melody construction. For comparison, makams from religious Turkish music of the artists Kani Karaca and Bekir Sıdkı Sezgin have been chosen. The Turkish makams have been processed the same way as the Byzantine Chants as indicated in Figure 1. In Figure 8 (bottom) the average scale degree amplitudes $\bar{\pi}_l$ for 69 Turkish makams are shown. In contrast to the Byzantine Chant, for the Turkish makams, a strong emphasis of the

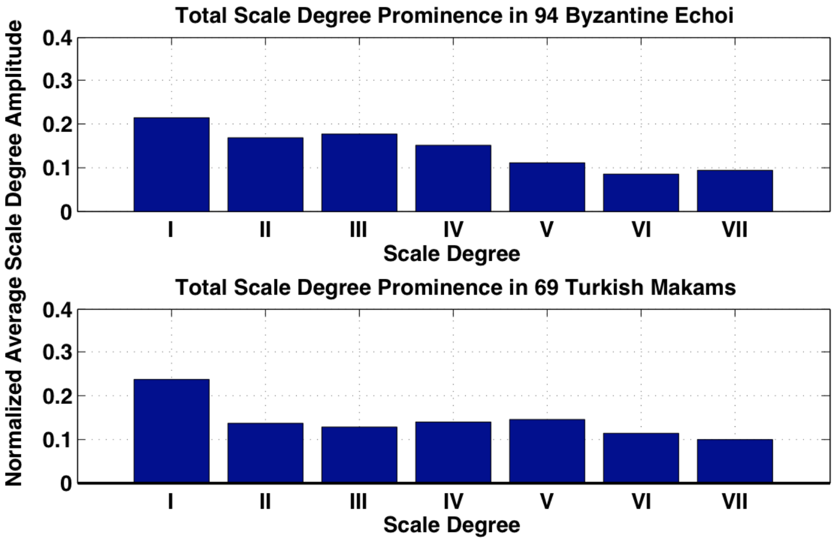

Figure 8: Average scale degree amplitude $\bar{\pi}_l$ (in frames) in Byzantine echoi (top) and scale degree histograms of religious Turkish makams (bottom). The decreasing emphasis with a relatively strong III in Byzantine scale degrees is contrasted with a strong V in Turkish makams.

I and to a lesser extent of the V and a deemphasis of the III, II, and IV can be observed. The high prominence of I and V in Turkish makams appears as the most important distinguishing feature between Byzantine and Turkish scale degree prominence.

The prominence of scale degrees for each echos individually is studied via a comparison of the set of theoretically most prominent scale degrees with the three highest average scale degree amplitudes in practice as shown in Table 8. In 14 of 21 echoi/chant genres, the empirical scale degree prominences are consistent with the theoretical ones. Theoretical and empirical scale degree prominence coincide for First Heirmoi, Third, and Fourth Heirmoi. Whereas the II is a theoretically prominent scale degree only for the Grave, the II appears among the three highest average scale degree amplitudes for six echoi (for First Authentic only for the genre Stichera). The three highest average scale degree amplitudes appear in the three adjacent scale degrees I-III except of the First Authentic of Heirmoi type, the Second Plagal, and the Fourth Plagal echos. In the Second Plagal echos, the three empirically most prominent scale degrees do not contain the I. scale degree, in contrast to Chrysanthine theory. The V. scale degree, never listed among the most prominent scale degrees in Chrysanthine theory, appears as the second most prominent scale degree for the Second Plagal and the Fourth Plagal. For the majority of echoi, the empirical set of most prominent scale degrees consists of the scale degrees I, II, III, with I usually being the most prominent scale degree.

Echos	**Chant Genre**	**Theory/ Practice**	**Theoretically Most Prominent Scale Degrees/ Highest Average Scale Degree Amplitudes**				
First	Heirmoi	Theory	I			IV	
		Practice	I		III	IV	
	Stichera	Theory	I		III		
		Practice	I	II	III		
First Plagal	Heirmoi	Theory	I		III		
		Practice	I	II	III		
	Stichera	Theory	I			IV	
		Practice	I	II	III		
Second	Stichera/Papadika	Theory	I		III		
		Practice	I	II	III		
Second Plagal	Stichera/Papadika	Theory	I			IV	
		Practice			III	IV	V
Third	All chants	Theory	I		III		
		Practice	I	II	III		
Grave	Heirmoi/Stichera	Theory	I	II		IV	
		Practice	I	II	III		
Fourth	Heirmoi	Theory	I		III		
		Practice	I	II	III		
Fourth Plagal	Heirmoi/Stichera	Theory	I		III		
		Practice	I		III		V
Count	Prominence	Theory	10	1	6	4	0
		Practice	9	7	10	2	2

Table 4: Scale tone prominence in theory and practice. Chrysanthine theory provides two or three most prominent scale degrees respectively. Highlighted cells correspond to the three most prominent scale degrees in practice as suggested by the echos average scale degree amplitudes π_l^i (Equation 17). The darkest colour represents the most prominent amplitude and the lightest colour the third largest amplitude. In the two bottom rows, for theory and practice, prominences are counted across all echoi/ chant genres. In practice, scale degrees II and V are more prominent than in Chrysanthine theory.

6 Discussion

We would like to review the processing steps and discuss how the statistical results could have been biased.

The extraction of the pitch (f_0) trajectory has been introduced in Section 4.1. Its evaluation in Section 4.1.1 indicated that for one recording of the Grave echos the estimated pitch trajectory is relatively distorted. It appears that this distortion is due to the presence of drone tones in the pauses of the singing voice, when the singing voice does not dominate the relatively low voice drone tone. As a result, the I. could be overemphasized in the echos average scale degree amplitude (Equation 17). However, this artifact is not likely to affect the empirical scale degree pitches $\hat{\nu}$ a lot. Previously filtering out the drone pitch could reduce noise in the f_0 detection of the melody.

In Section 4.4.2, Algorithm 2 for tonic detection is evaluated. As an evaluation result the tonic was detected incorrectly in two out of 20 of the excerpts. These are exactly two of the few exceptions in which the final note of the chant does not end on the tonic. This occurs for instance when the ending of one chant harmonically prepares the start of the following chant in the course of the Byzantine liturgy. Given the small sample size in the tonic detection experiment, we cannnot state the exact percentage of echoi with this non-tonical ending notes with high confidence. However a percentage of 10% would be a reasonable estimate. This introduces noise in the plots of the echoi in Figure 9 and a bias in the statistical results in Table 2 and 3. This bias is limited by the fact that the Wilcoxon signed-rank test tests for the median. Therefore, distorted pitches would bias the outcome much less than in the case of a test (e.g. t-test) that considers the mean. In future work, the tonic detection algorithm could be improved to handle these cases. The cross-correlation method as applied in (Gedik, Ali C & Bozkurt, 2008) could be used for this objective, at the cost of introducing another bias in estimating the theory-practice deviations by using music theory information more excessively when processing the data from performance practice. This bias could be reduced by only correlating the down-transpositions of the theoretical scale degree pitches in Table 1 by a minor third, fourth, and fifth. When employing the cross-correlation method using the theoretical scale degree pitches (Table 1), one would have to decide how to account for the different amplitudes across the scale degrees. The scale degree amplitudes could be estimated empirically. But when using the empirical data to determine the theoretical scale degree amplitudes, this, again, would introduce a bias to the subsequent tests.

In Section 4.2, a slight shifting of the reference pitch during the performance could happen. As a consequence, the peaks of the pitch histogram c would be smeared out. This effect is likely to be small in the case of professional singers. Also the vibrato range contributes to the width of the peaks. The relatively irregular patterns of vibrato and other micro-intonation in the singing voice makes it hard to derive sharp pitch histogram peaks in the case of the singing voice. The bandwidth h for kernel smoothing is selected manually. A change of the smoothing bandwidth h affects the histogram peak locations $\lambda = (\lambda_1, \ldots, \lambda_U)$ in Section 4.3. However, due to the large sample size of the instantaneous pitch vector p this bias may be limited. When considering the instantaneous pitches $\mathbf{p_l}$ (Equation 8) around scale degree l in Section 4.5.1, the tails of the distribution for each scale degree are cut off. Also there is a dependency between adjacent scale degrees and between the empirical peak positions and the nearest theoretical scale degree pitch to which they have been mapped. However, an informal inspection reveals that the empirical peaks are relatively close to the theoretical peaks, which indicates that this bias is limited. Since the empirical scale degree pitches $\hat{\nu} = (\hat{\nu}_1, \ldots, \hat{\nu}_L)$ are normalized with respect to scale degree I, the estimation error for scale degree pitches in the middle (IV and V) in between two octaves of scale degree I are larger than for the scale degree steps close to I, such as II and VII. This effect appears to be limited since in Table 2 there are much larger deviations for scale degrees VI, VII and III than for scale degrees IV and V.

However, the impact of the bias of the preprocessing on the statistical results is limited. This paper presents a novel statistical framework for the quantitative analysis of scale degrees that leaves some refinement of the statistical analysis for future work.

7 Conclusion

In this paper, a new method has been introduced to empirically study particular aspects of pitch usage in performance practice, namely tuning, steps, and prominence of scale degrees. The method has been applied to the analysis of Byzantine Chants. In particular, a methodology has been introduced to test to what degree performance practice of Byzantine Chant follows the widely known Chrysanthine theory. The theoretic hypotheses have been tested on a corpus of recordings of chants of the octoechos. An approach combining pitch estimation with appropriate post filtering, kernel smoothing, and statistical tests has been applied to the recordings. Among the novel methods proposed here are the histogram computation and smoothing and the tonic detection algorithm. The former comprises the use of Gaussian kernels in histogram computation and suitable tuning of the smoothing factor. The latter includes a robust method of tonic detection from the last phrase note. In addition, a statistical framework has been introduced to study scale degrees.

In general, the analysis shows that performance practice follows Chrysanthine theory. However, the analysis results also indicate subtleties in performance practice of Byzantine Chant: The singer shows the tendency to diminish the largest step sizes ($14, 20$ moria: II-III, VI-VII in Second Authentic/Plagal) among all scale steps within the octoechos. The theoretical scale degree IV in the Fourth Authentic is exceptionally higher than in all other echoi. In this echos, the singer diminishes the theoretical V in practice. In Fourth, also scale degree step VII-I (theoretically 10 moria) tends to be diminished towards the more common step VII-I of 8 moria. This gives support to the conjecture that the singer levels the extreme step and scale degrees particularities within the octoechos. The IV., the frame of the first tetrachord is the scale degree with least theory-practice deviation. The scale degree tuning within the first tetrachord deviates less from theory than the scale degrees outside the first tetrachord.

The analysis of the scale steps reveals that the most frequently used scale steps are the whole tone and the minimal tone, followed by the minor tone and the semitone. The smallest and largest theoretical steps do not occur as often. Whereas the smallest scale step (67 cents), the semitone, the minor and the minimal tone are significantly larger in practice compared to theory, the opposite is true for the largest scale step of 333 cents.

Regarding the prominence of the scale degrees, the scale degrees within the first tetrachord occur more often than the other scale degrees with a decreasing prominence of scale degrees I (maximum), III, II, and IV.

The present study, as a first step towards a computational analysis of pitch patterns of Byzantine Chant, can be extended in many ways. In particular, the analysis of micro-intonation in the range of a few cents, embellishments and cadential clauses, the consideration of rhythmical/metrical context and of melodic contour could aid the understanding of pitch usage. Furthermore, beyond the eight echoi, the echoi genres (Heirmoi, Stichera and Papadika) should be studied. In this study we intended to minimize the employed musical prior kowledge. However for an effective echos recognition application, the tonic detection algorithm could be improved in future work by integrating feedback from tonic detection via the cross-correlation method as applied in (Gedik, Ali C & Bozkurt, 2008).

It has been shown in this article that the presented system serves for empirically testing a theoretic model of Byzantine Chant. The empirical findings can then help to refine the theoretical model and to let it reflect the musical performance practice. Furthermore, the methodology introduced here has applications to a wide range of oral music traditions, in particular to the ones that share similarities with the modal system of Byzantine music, e.g. Greek, Cypriot and Turkish folk music that have been historically influenced by Byzantine music.

8 Acknowledgments

H. P. was supported in part by the German "Bundesministerium für Bildung und Forschung" (BMBF), Grant BFNT, No. 01GQ0850. We are grateful to Daniel Bartz from Berlin Institute of Technology for helpful discussion and proof reading the manuscript.

References

Akkoç, C. (2002). Non-deterministic scales used in traditional Turkish music. *Journal of New Music Research*, *31*(4), 285–293(9).

Angelopoulos, L. (1986). *The importance of simon karas research and teaching regarding the taxonomy and transcription of the effect of the signs of cheironomy: Oral interpretation of the written interpretation.* Retrieved 30/03/2011, from http://www.psaltiki.net/journal/2.1/angelopoulos/angelopoulos_karas.html

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* New York: Springer.

Bozkurt, B. (2008, March). An Automatic Pitch Analysis Method for Turkish Maqam Music. *Journal of New Music Research*, *37*(1), 1–13.

Braun, J. (1980). Musical Instruments in Byzantine Illuminated Manuscripts. *Early Music*, *8*(3), 312–327.

Chordia, P., & Rae, A. (2007). Raag recognition using pitch-class and pitch-class dyad distributions. In S. Dixon, D. Bainbridge, & R. Typke (Eds.), *Proceedings of the international conference on music information retrieval* (pp. 431–436).

Damarlaki, I. (1999). *Pws tha mathw na psallw praktika [learn how to perform the byzantine chants] [audio cd]*. Crete: Polychronakis.

de Cheveigne, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111*(4), 1917. doi: 10.1121/1.1458024

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, *56*(293), 52–64.

Gedik, Ali C, & Bozkurt, B. (2008). Automatic Classification of Turkish Traditional Art Music Recordings by Arel Theory. In *Proceedings of the fourth conference on interdisciplinary musicology (cim08)* (Vol. 56). Thessaloniki, Greece.

Gómez, E. (2006). *Tonal Description of Music Audio Signals* (Vol. 192). Unpublished doctoral dissertation, Universitat Pompeu Fabra. doi: 10.1086/508205

Ioannidis, D. (2005). *H theoria tis byzantinis mousikis stin praxi [theory of byzantine music in practice]*. Athens: Ioannidis Dimitrios.

Kakoulidis, G. I. (1999). *Pws na vriskoume praktika tous okto ihous tis byzantinis mousikis [how to estimate empirically the eight modes of byzantine music]*. Athens.

Karas, S. (1970). *Ta geni kai diastimata eis tin byzantini mousiki [genres and intervals in byzantine music]*. Athens.

Kartsonas, P. (n.d.). *Praktiki methodos ekmathisis tis psaltikis tehnis [practice methods for learning byzantine chant] [audio cd]*. The Hunt of St. George Mount Athos.

Levy, K., & Troelsgård, C. (2011). *Byzantine chant.* Oxford Music Online. Retrieved 24/08/2011, from `http://www.oxfordmusiconline.com/subscriber/article/grove/music/04494`

Mavroeidis, M. (1999). *Oi mousikoi tropoi stin anatoliki mesogeio [the musical modes in east mediterranean]*. Athens: Fagotto.

Moelants, D., Cornelis, O., & Leman, M. (2009). Exploring African Tone Scales. *10th International Society for Music Information Retrieval Conference (ISMIR)*, 489–494.

Purwins, H., Blankertz, B., Dornhege, G., & Obermayer, K. (2004). Scale degree profiles from audio investigated with machine learning. In *Audio engineering society convention 116.*

Purwins, H., Blankertz, B., & Obermayer, K. (2000). A new method for tracking modulations in tonal music in audio data format. In *Neural networks, 2000. ijcnn 2000, proceedings of the ieee-inns-enns international joint conference on* (Vol. 6, pp. 270–275).

Rabiner, L. (1977). On the use of autocorrelation analysis for pitch detection. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *25*(1), 24–33.

Serrà, J., Koduri, G. K., Miron, M., & Serra, X. (2011). Assessing the tuning of sung indian classical music. In *Ismir* (pp. 157–162). Retrieved from `http://ismir2011.ismir.net/papers/OS2-2.pdf`

Sheather, S. J., & Jones, C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical*

Society Series B Methodological, *53*(3), 683–690.
Thoukididi, A. (2003). *Kripida byzantinis mousikis [foundations of byzantine music]*. Cyprus: Kykkos Church.
Toiviainen, P., & Eerola, T. (2006). Visualization in comparative music research. In *Computational statistics compstat 2006* (pp. 209–219). Springer.
Tzanetakis, G., Ermolinskyi, A., & Cook, P. (2002). Pitch Histograms in Audio and Symbolic Music Information Retrieval. *Journal of New Music Research*, *32*(2), 143–152.
Tzanetakis, G., Kapur, A., Schloss Andrew, W., & Wright, M. (2007). Computational Ethnomusicology. *Journal of Interdisciplinary Music Studies*, *1*(2), 1–24.
Winnington-Ingram, R. P. (1936). *Mode in Ancient Greek Music.* Cambridge [Eng.]: The University press.
Zannos, I. (1990). Intonation in Theory and Practice of Greek and Turkish Music. *Yearbook For Traditional Music*, *22*(1990), 42.

9 Appendix: Pitch Histograms and Theoretical Scale Degree Pitches for all Echoi

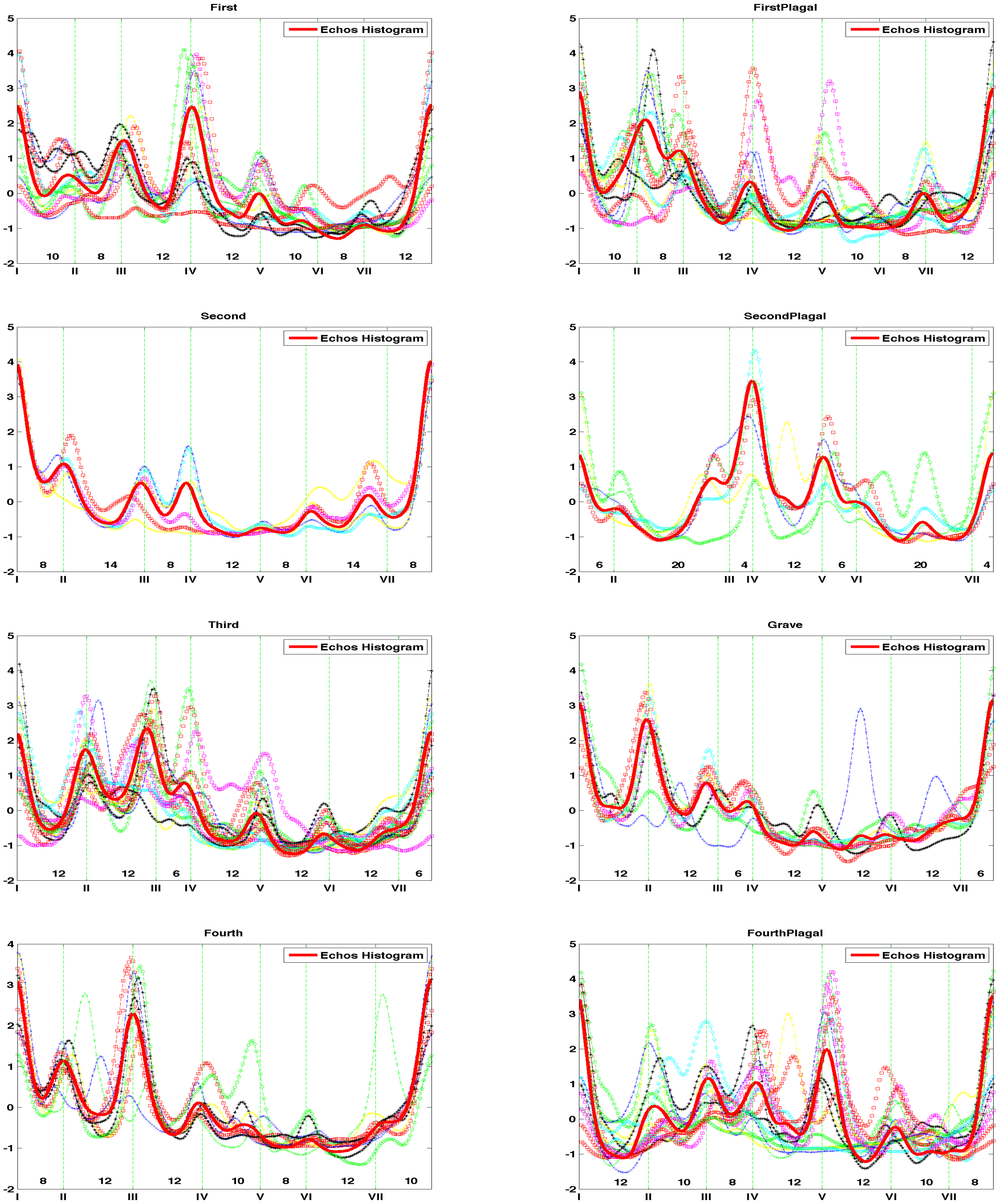

Figure 9: For all recordings of all echoi, the pitch histograms are displayed. The vertical lines indicate pitches of scale degrees according to Chrysanthine theory (cf. Table 1). The y axis represents the normalized histogram count. The bold red lines represent the echos histogram computed from pitch trajectories across all recordings of the same echos.