TRADITIONAL ASYMMETRIC RHYTHMS: A REFINED MODEL OF METER INDUCTION BASED ON ASYMMETRIC METER TEMPLATES

Thanos Fouloulis
Dept. of Music Studies,
Aristotle Univ. of Thessaloniki
thanos.fouloulis@gmail.com

Aggelos Pikrakis
Dept. of Computer Science,
University of Piraeus
pikrakis@unipi.gr

Emilios Cambouropoulos
Dept. of Music Studies,
Aristotle Univ of Thessaloniki
emilios@mus.auth.gr

ABSTRACT

The aim of this study is to examine the performance of an existing meter and tempo induction model (Pikrakis et al, 2004) on music that features asymmetric rhythms (e.g. 5/8, 7/8) and to propose potential improvement by incorporating knowledge about asymmetric rhythm patterns. We assume knowledge of asymmetric rhythms in the form of metric templates (consisting of both isochronous and asymmetric pulse levels). Such knowledge seems to aid the induction process for symmetric/asymmetric rhythms and thus improve the performance of the aforementioned model.

1. INTRODUCTION

In recent years a number of meter induction and beat tracking models have been implemented that attempt to identify perceptually pertinent isochronous beats in musical data. Such models assume an isochronous tactus within a certain tempo range (usually centered around the spontaneous tempo). The performance of such systems is usually measured against musical datasets drawn from Western music (e.g. classical, rock, pop, jazz) that features almost exclusively symmetric rhythmic structures (e.g. 3/4, 4/4, 6/8) (Mckinney et al. 2007; Dixon 2007; Davies et al. 2009). The tactus of asymmetric/complex musical rhythms, however, is non-isochronous; for instance, a 7/8 song is often counted/taped/danced at a level 3+2+2 (not at a lower or higher level). Such models fail to identify asymmetric beat levels (Fouloulis et al, 2012).

Musical time is commonly organized around a (hierarchic) metrical structure of which the most prominent level is the beat level (tactus) (Lerdahl and Jackendoff, 1983). Such a metric structure facilitates the measurement of time and the categorical perception of musical temporal units (durations, IOIs). In western music, an isochronous beat level is almost always assumed; any divergences from isochronous beat are treated as 'special cases' or even 'anomalies'.

A central assumption of this paper is that the beat level (tactus) of metrical structure need not be isochronous. It is asserted that metrical structure is learned implicitly (through exposure in a specific idiom), that it may be asymmetric and that the tactus level itself may consists of non-isochronous units. It is maintained that an acculturated listener may use spontaneously an asymmetric tactus to measure time, as this is the most plausible and parsimonious way to explain and organize rhythmic stimuli within specific musical idioms.

Rhythm and pitch share common cognitive underlying mechanisms (Parncutt, 1994; Krumhansl, 2000). Asymmetric structures are common in the pitch domain. Major and minor scales, for instance, are asymmetric. Listeners learn pitch scales through exposure to a specific musical idiom, and then automatically organize pitch and tonal relations around the implied asymmetric scales. Asymmetric scales are actually better (cognitively) than symmetric scales (e.g. 12-tone chromatic scale or whole-tone scale) as they facilitate perceptual navigation in pitch/tonal spaces. It is, herein, assumed that asymmetric beat structures may arise in a similar fashion to asymmetric pitch scales, and may organize certain rhythmic structures in an accurate and more parsimonious manner.

In more formal terms, the kinds of asymmetric beat structures mentioned in this study may be described as series of repeating asymmetric patterns consisting of long (three's) and short (two's) units. Such asymmetric patterns are 'sandwiched' in between a lower isochronous sub-beat level (commonly at the 1/8 duration) and a higher isochronous metric level (e.g. 5=3+2 or 7=3+2+2) (Fouloulis et al, 2012). Such hierarchic metric structures are considered in this paper as a whole rather than a number of independent isochronous and asymmetric pulse levels.

2. METER AND TEMPO INDUCTION MODEL

2.1 Original model architecture

In this study we examine a potential improvement in the performance of an existing model (Pikrakis et al, 2004) that focuses on meter and tempo extraction on polyphonic audio recordings. The existing version processes audio in non overlapping long-term windows while using an inner moving short-term window to generate sequences of feature vectors considering energy and mel frequency cepstral coefficients (MFCCs) (Figure 1). For every long-term window a Self Similarity Matrix (SSM) is formulated based on the assumption that its diagonals can reveal periodicities corresponding to music meter and beat. Calculating the mean value of each diagonal and plotting it against the diagonal index each audio segment reveals a "rhythmic signature" that can be further analyzed in order to infer the actual beat and meter. Two different ranges of SSM diagonal indices in this "rhythmic signature" are considered suggesting that beat and meter candidates are lying within respectively.

The original model relies on two criteria to associate certain periodicities to music meter and tempo. In the first criterion beat candidates are selected as the two neighbouring local minima that possess larger values. Meter candidates are validated in relation to beat candidates according to the accepted set of music meters under investigation. Calculating the sum of corresponding mean values for every pair, the music meter of a segment can be determined as the one that exhibits the lowest value. The second criterion differentiates in that it takes into account the slope (sharpness) of the valleys of each pair and not just their absolute values.

The meter of the whole audio is selected taking into account its frequency of occurrence through histograms that are formed using the calculated meter values per segment. Tempo estimation process, based on previous results about beat lag, is jointly extracted per long-term segment or as average for the whole audio.

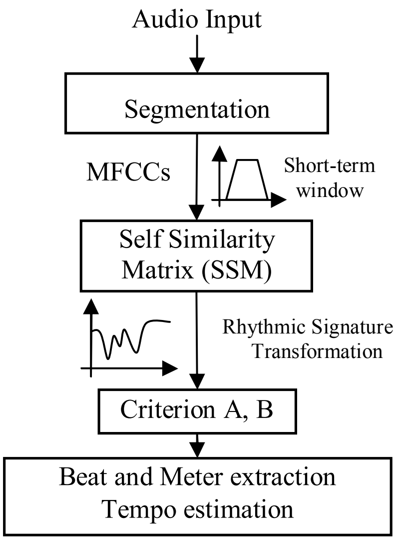

Figure 1. Overview of the architecture of the original meter and tempo induction model.

2.2 Refined model architecture

The main motivation behind the refined model relies on the assumption that meter induction can be assisted by querying an audio recording against known metric templates. Knowledge about metric structure is incorporated into the model by including a set of both symmetric and asymmetric templates in a form of a template library (Figure 2). During the induction process and for a given tempo hypothesis each "rhythmic signature" of an audio recording can be evaluated in turn with the contents of the template library so that we can conclude to the most prominent one.

2.3 Template Generation

Templates were generated for the following time signatures 2/4, 3/4, 4/4, 5/8 (3+2), 6/8, 7/8 (3+2+2), 8/8 (3+3+2) and 9/8(3+2+2+2) using MIDI and audio drum sounds for a reference tempo of 260bpm (1/8). The corresponding audio files were then transformed into their respective template "rhythmic signatures" by using the same procedure as before. In Figure 3 "rhythmic signatures" of 7/8 and 5/8 templates on a tempo of 260bpm (1/8) are presented. The lowest local minima (valleys) on these templates match strong periodicities and can be considered as meter candidates. The distance in the x-axis D_t between two successive meter candidates is generally altered accordingly to tempo changes and is utilized during the induction process in order to scale the template according to the calculated tempo.

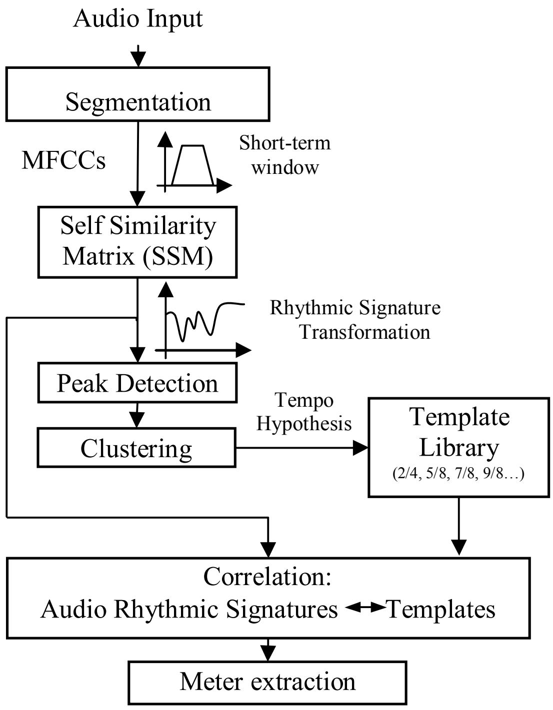

Figure 2. Overview of the architecture of the refined meter and tempo induction model.

3. IMPLEMENTATION DETAILS

The refined system keeps the initial audio processing steps of the original model but we refer to them anyway for the sake of comprehension. In the first step audio recordings are processed on a segment by segment basis in non overlapping long-term windows of 10s. Sequences of feature vectors are extracted using a "chroma based" variation of standard MFCCs, which yields significantly better results by emphasizing beat and meter candidates. This approach instead of assuming equally spaced critical band filters in the mel scale makes use of a critical band filter bank consisting of overlapping triangular filters, that is aligned with the chromatic scale of semitones (starting from 110 Hz and reaching up to approximately 5KHz) (Pikrakis et al, 2004).

Feature vectors in each long-term window are extracted by means of a short-term processing technique. The values for the length, w_s and hop size h_s of the short-term window were chosen as 100ms and 10ms respectively. Then, the sequences of feature vectors are utilized to form self-similarity matrices (SSM), using the Euclidean function as a distance metric, in order to reveal the dominant periodicities inside each segment. This can be achieved by computing the mean value B_k for each diagonal k and plot the value against the diagonal index. Local minima in this curve correspond to strong periodicities that are prominent in the specific time frame. We can consider the function B(k) as the "rhythmic signature" of the long-term segment from which it is extracted.

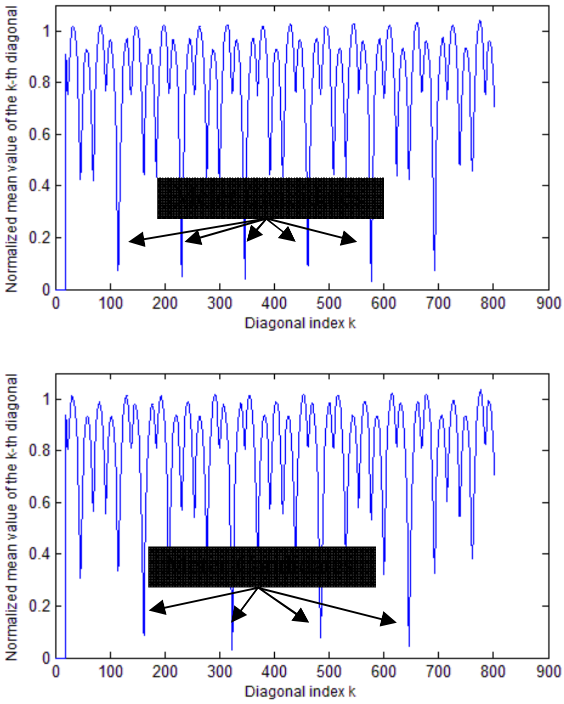

Figure 3. "Rhythmic signatures" for a 5/8 (top) and a 7/8 (bottom) template.

3.1 Peak detection - smoothing

Each "rhythmic signature" is then processed using a peak detection algorithm to extract the diagonal indices k that correspond to the most salient local minima (valleys). The peak detection algorithm uses the first derivative of the signal and relies on the fact that the first derivative of a peak has a downward-going zero-crossing at the peak maximum. To avoid picking false zero-crossing due to the noise we use a technique that initially smooths the first derivative of the signal using a rectangular window, and then it takes only those zero crossings whose slope exceeds a certain pre-determined minimum (slope threshold). The smoothing algorithm simply replaces each point in the signal with the average of m adjacent points defined by smooth width. For example, for a 3-point smooth (m = 3) (O'Haver, 2013):

$$S_j = \frac{Y_{j-1} + Y_j + Y_{j+1}}{3} \quad (1)$$

where S_j the j-th point in the smoothed signal, Y_j the j-th point in the original signal, and n is the total number of points in the signal.

3.2 Clustering valleys

In order to account for light tempo changes and also slight deviations from strict metronomical performance we cluster the detected valleys by using the notion of valley bins. Each bin is defined by a diagonal index mean value m_b and a tolerance window e. Each time a new valley is assigned to a relative bin the bin mean value m_b is updated. The time equivalent T_k for a local minimum k is $T_k = k * step$ (Pikrakis et al, 2004) where step is the short-term step of the moving window (10 ms for our study). In this work the width of the tolerance window was defined to be 8*step =80ms.

Valleys are weighted by taking into account their frequency of occurrence in the sequence of "rhythmic signatures", their slope and their amplitude. This relies on the assumption that meter periodicities are prominent in the majority of the "rhythmic signatures" and exhibit steeper slopes and narrower valleys. Therefore, bins which are more populated and contain sharper valleys are discriminated. The next step is to pick the two most important valleys bins that have successive mean values m_b. If the previous assumption is right and those two successive valley bins correspond to meter candidates then the distance D_s in x-axis between them can be compared to the corresponding distance D_t of each template.

This comparison determines the stretching/expanding factor f_t for each template that is needed to compensate for the tempo difference between the tempo of the real audio file and the reference tempo (260bpm - 1/8) that was specified during template generation. The product of this step is to conclude in tempo hypothesis using factor f_t and then perform a "time scaling" for each template of the template bank.

3.3 Meter extraction

The final step of the algorithm performs a correlation analysis between each "rhythmic signature" of the audio and every time-scaled template. In particular, each template is slided through every rhythmic signature and a correlation coefficient is calculated. Finally, the template for which the correlation coefficient has a maximum value is considered as the winner. The results for a 7/8 and a 9/8 song are presented in figure 4.

4. RESULTS AND DISCUSSION

In a previous study (Fouloulis et el. 2012) we tested the original version of the model (Pikrakis et al. 2004) against a set of Greek traditional songs that featured mostly asymmetric rhythms with time signatures of 2/4, 3/4, 5/8, 6/8, 7/8, 8/8, 9/8, 10/8 and 11/8. The majority of the songs were derived from educational material and most of them start with an introductory rhythmic pattern in order to indicate the correct way of tapping/counting.

In this study we used a similar set of 30 Greek traditional songs and examined the model's performance after incorporating templates with time signatures of 2/4, 3/4, 4/4, 5/8, 6/8, 7/8, 8/8, and 9/8 (Table 1). The preliminary results are encouraging, indicating that this architecture may prove to be quite effective and may assist the induction process. In general the model seems to retain its significant behavior in processing non-symmetric meters but some more tweaking is needed in order to further improve performance.

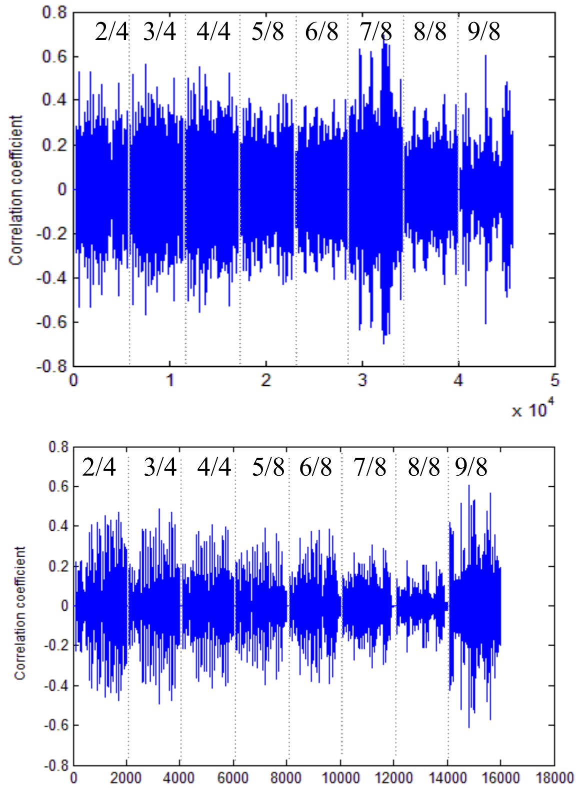

Figure 4. Correlation analysis for a 7/8 (top) and a 9/8 song (bottom).

In cases (tracks 19, 20, 21, 26, 28 and 29) when the original model tended to designate as more dominant periodicities the ones that referred to a span of two measures the refined model's output corresponds to the correct tempo and meter.

For tracks no. 4-6, the refined model assumes a 6/8 meter instead of 3/4. This seems to be supported but the nature of the performance (Figure 5). For track 7 and 8, it indicates an asymmetric 8/8 while the notation of the song indicates an isochronous pulse; again this is due to the performance elements that introduce asymmetric features.

It is worth pointing out that the instances in which the algorithm falls into a wrong estimation are songs with too fast tempi (songs 10, 16, 18, 22 and 23). In all these cases the actual meter value resides in the correlation plot but with a lower peak.

5. FURTHER RESEARCH

The architecture presented above still has many open issues that need to be explored. First of all it is necessary to evaluate its performance using a larger data set. Secondly, the results could probably be improved if further musicological/cogntive knowledge is incorporated. For example, constraints about tempo hypotheses that exceed some limits (e.g. too slow or too fast rates) could be integrated. Additionally, a wider range of more refined templates can be generated (by assigning a variety of sounds to the various metric midi templates), allowing a more effective discrimination between different metric structures for a given tempo.

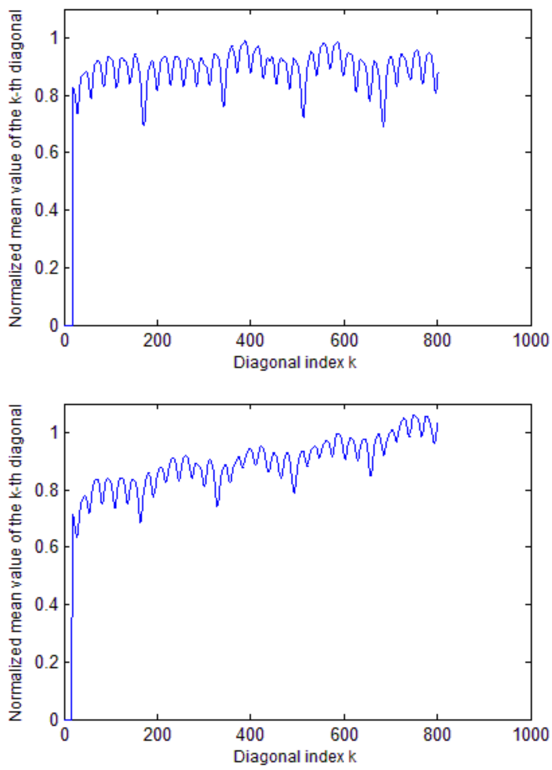

Figure 5. "Rhythmic signatures" from one segment of song no. 14 (top) and song no. 6 (bottom). Patterns seem to support the fact that even if song no. 6 is notated as 3/4 it can be considered as 6/8 due to performers' musical idiom.

6. CONCLUSION

In this study we investigate a potential improvement in the performance of an existing model (Pikrakis et al, 2004) in inducting meter from audio recordings of folk music by embedding knowledge about asymmetric/complex rhythmic structures. Templates of common asymmetric rhythm patterns were generated and then imported into the system. The preliminary results in this ongoing research are very encouraging, indicating that this architecture may prove to be quite effective and can assist the induction process.

7. REFERENCES

Cambouropoulos, E. (1997) Musical Rhythm: Inferring Accentuation and Metrical Structure from Grouping Structure. In *Music, Gestalt and Computing - Studies in Systematic and Cognitive Musicology*, M. Leman (ed.), pp. 277-293. Springer, Berlin.

Davies, M. E. P., Degara N. & Plumbley M. D. (2009). Evaluation Methods for Musical Audio Beat Tracking Algorithms. *Technical Report C4DM-TR-09-06*, Queen Mary University of London, Centre for Digital Music.

Dixon S. (2007). Evaluation of the audio beat tracking system BeatRoot. *Journal of New* Music *Research*, 36(1):39-50.

Fouloulis, A., Pikrakis, A., & Cambouropoulos, E (2012). Asymmetric beat/tactus: Investigating the performance of beat-tracking systems on traditional asymmetric rhythms. In *Proceedings of the Joint Conference ICMPC-ESCOM 2012*, Thessaloniki, Greece. 23-28 July 2012.

Krumhansl, C.L.(2000) Rhythm and pitch in music cognition. *Psychological Bulletin* 126, pp. 159–179 .

Lerdahl, F. and Jackendoff, R. (1983) *A generative Theory of Tonal Music,* The MIT Press, Cambridge (Ma).

Mckinney, M. F., Moelants, D., Davies, M. E. P., and Klapuri, A. (2007). Evaluation of audio beat tracking and music tempo extraction algorithms. *Journal of New Music Research*, 36(1):1–16.

O'Haver, T C., (2013). A Pragmatic Introduction to Signal Processing. Retrieved from: http://terpconnect.umd.edu/~toh/spectrum/IntroToSignalProcessing.pdf

Parncutt, R. (1994) Template-Matching Models of Musical Pitch and Rhythm Perception. *Journal of New Music Research,* 23:145-167.

Pikrakis, A., Antonopoulos, I., & Theodoridis, S. (2004). Music meter and tempo tracking from raw polyphonic audio. *Proceedings of ISMIR 2004*, Spain

Table 1. Meter and tempo induction results of the original and refined model.

					Refined Model with embedded templates		**Original model**	
	Song's Name	**Time Signature**	**Tempo**	**Meter**	**Calc. Tempo**	**Calc. Meter**	**Calc. Tempo**	**Calc. Meter**
1	Sousta Rodou	2/4	144 (1/4)		**141 (1/8)**	**2/4**	285	4:1
2	Mpalos	2/4	82 (1/4)		**84 (1/8)**	**2/4**	171	4:1
3	Ehe geia panagia (Hasapiko)	2/4	130 (1/4)		**258 (1/8)**	**4/8**	260	4:1
4	Tsamikos	3/4	98 (1/4)		**196 (1/8)**	**6/8**	98	3:1
5	Apopse mavromata mou	3/4	104(1/4)		**209 (1/8)**	**6/8**	206	6:1
6	Valtetsi	3/4	108(1/4)		**218 (1/8)**	**6/8**	214	6:1
7	Armenaki	4/4	180(1/4)		**357 (1/8)**	**8/8**	181	4:1
8	Louloudi ti marathikes	4/4	127(1/4)		**256 (1/8)**	**8/8**	260	8:1
9	Zagorisios -Kapesovo	5/8	94 (1/8)	**2-3**	**96 (1/8)**	**5/8**	97	2:1 or 5:1
10	Mpaintouska Thrakis	5/8	420 (1/8)	**2-3**	**250 (1/8)**	**3/8**	83	4:1
11	Dio palikaria apo to Aivali	5/8	239(1/8)	**3-2**	**241 (1/8)**	**5/8**	-	-
12	Esvise to keri kira Maria	5/8	249(1/8)	**3-2**	**243 (1/8)**	**5/8**	-	-
13	I Kiriaki	5/8	300(1/8)	**3-2**	**293 (1/8)**	**5/8**	-	-
14	Itia	6/8	201(1/8)		**202 (1/8)**	**6/8**	208	6:1
15	Enas aitos kathotane	6/8	209(1/8)		**209 (1/8)**	**6/8**	206	6:1
16	Perasa ap΄tin porta sou	7/8	264(1/8)	**3-2-2**	**74 (1/8)**	**2/8**	130	7:1
17	Tik Tromakton Pontos	7/8	488 (1/8)	**2-2-3**	**499 (1/8)**	**7/8**	73	2:1
18	Mantilatos Thrakis	7/8	483 (1/8)	**2-2-3**	**199 (1/8)**	**3/8**	69	2:1 or 3:1
19	Mantili Kalamatiano	7/8	273 (1/8)	**3-2-2**	**273 (1/8)**	**7/8**	132	7:1
20	Milo mou kokkino	7/8	268 (1/8)	**3-2-2**	**265 (1/8)**	**7/8**	133	7:1
21	Na diokso ta synefa	7/8	266 (1/8)	**3-2-2**	**266 (1/8)**	**7/8**	130	7:1
22	Oles oi melahroines	8/8	381 (1/8)	**3-3-2**	**83 (1/8)**	**2/8**	193	4:1
23	Dyo mavra matia agapo	8/8	396(1/8)	**3-3-2**	**97 (1/8)**	**2/8**	200	4:1
24	Marmaromenios vasilias	8/8	198(1/8)	**3-3-2**	**195 (1/8)**	**8/8**	-	-
25	Feto to kalokairaki	9/8	136(1/8)	**2-2-2-3**	**139 (1/8)**	**9/8**	139	9:1
26	Karsilamas	9/8	256 (1/8)	**2-2-2-3**	**255 (1/8)**	**9/8**	130	9:1
27	Amptaliko neo	9/8	104 (1/8)	**3-2-2-2**	**246 (1/8)**	**9/8**	61	9:1
28	Tsiourapia Makedonias	9/8	276 (1/8)	**2-2-2-3**	**296 (1/8)**	**9/8**	109	9:1
29	karsilamas - Ti ithela	9/8	288 (1/8)	**2-2-2-3**	**290 (1/8)**	**9/8**	146	9:1
30	Ela apopse stou Thoma	9/8	185 (1/8)	**2-2-2-3**	**184 (1/8)**	**9/8**	96	9:1