

Asymmetric beat/tactus: Investigating the performance of beat-tracking systems on traditional asymmetric rhythms

Thanos Fouloulis,^{*1} Aggelos Pikrakis,^{#2} Emiliios Cambouropoulos^{*3}

^{*}*Dept. of Music Studies, Aristotle University of Thessaloniki, Greece*

[#]*Dept of Computer Science, University of Piraeus, Greece*

¹afoululis@mus.auth.gr, ²pikrakis@unipi.gr, ³emiliios@mus.auth.gr

ABSTRACT

Theories of western metrical structure commonly hypothesize an isochronous beat level (tactus) upon which the concept of metre is built. This assumption is challenged by this study. It is proposed that time at the tactus level may be measured by isochronous or asymmetric temporal ‘scales’ depending on the musical data (just like asymmetric pitch scales are adequate for organizing tonal pitch space). This study examines the performance of beat tracking systems on music that features asymmetric rhythms (e.g. 5/8, 7/8) and proposes potential improvement of theoretical and practical aspects relating to beat perception that can allow the construction of more general idiom-independent beat trackers. The tactus of asymmetric/complex musical rhythms 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). Two state-of-the-art beat-tracking systems (Dixon 2007; Davies & Plumbley 2007) and a beat/tempo induction system (Pikrakis et al. 2004) are tested on a number of traditional Greek (dance) songs that feature asymmetric rhythms. The beat output of the algorithms is measured against the corresponding beat structures indicated by expert musicians, and the algorithms are compared to each other. As expected, the beat-trackers cannot cope well with asymmetric rhythms. The metre/tempo induction system performs better in processing asymmetric rhythms; it does not always find the correct beat level but this level exists implicitly in the model (in between sub- and super-beat levels).

I. INTRODUCTION

In recent years a number of 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 etc) (Mckinney et al. 2007; Dixon 2007; Davies et al. 2009).

The aim of this study is to examine the performance of beat tracking systems on music that features asymmetric rhythms (e.g. 5/8, 7/8) and to propose potential improvement of theoretical and practical issues relating to beat perception that can allow the construction of more general idiom-independent beat trackers. The tactus of asymmetric/complex musical rhythms 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). Should such asymmetric non-isochronous beat levels be considered as exceptions or even ‘anomalies’ to standard isochronous beat definitions and be treated thus accordingly in beat tracking systems? Or a broader definition of beat is possible that can naturally accommodate asymmetric rhythms?

In section II, some properties of asymmetric metres, and more specifically asymmetric beats, are discussed. Then, in section III, two state-of-the-art beat-tracking systems (Dixon 2007; Davies & Plumbley 2007) and a metre/tempo induction system (Pikrakis et al. 2004) are described. Finally, in section IV the three metre/beat models are tested on a number of traditional Greek songs that feature asymmetric rhythms. The beat output of the algorithms on these songs is measured against the corresponding beat structures indicated by expert musicians (we also use knowledge regarding corresponding dance movements as most of the songs can be danced), and the algorithms are compared to each other.

II. ASYMMETRIC BEAT AND METRE

Musical time is commonly organized around a (hierarchical) 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 consist 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. Performing a metrical analysis of these patterns

reveals a structure where long (three's) and short (two's) units are 'sandwiched' in between a lower isochronous sub-beat level and a higher isochronous metric level (Cambouropoulos, 1997) – see examples in Figure 1.

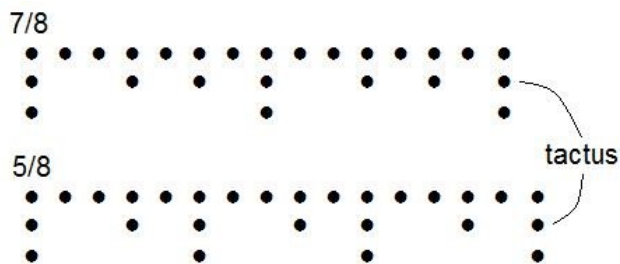


Figure 1. Asymmetric metrical structures

III. BEAT/METRE MODELS

A. Simon Dixon's BeatRoot

BeatRoot, a beat tracker application presented by Simon Dixon is designed to track beats in expressively performed music. The system consists of three main components which perform onset detection, tempo induction and beat tracking. Input data can be digital audio or symbolic representation such as MIDI. (Dixon 2007)

The first main stage of processing, onset detection process, detects the salient rhythmic events using a spectral flux difference function. The list of onset (event) times is then used by the tempo induction module to extract periodicities through an all-order inter-onset interval (IOI) analysis. Likely tempo hypotheses at various metrical levels are generated by using a clustering algorithm which groups similar IOIs that represent the various musical units (e.g. half notes, quarter notes, etc).

These hypotheses, together with the event times become the input to the beat tracking subsystem. This subsystem relies on a multiple agent architecture to evaluate several different hypotheses concerning the rate and timing of musical beats. Initially, each agent begins with a tempo (rate) hypothesis from the tempo induction subsystem and an onset time that is picked from the first few onsets, corresponding to agent's phase. Then the agent predicts further beats according to its rate and phase hypothesis, using tolerance windows so that it can cope with deviations from perfectly metrical time. Agents that coincide in predicting the time and rate of the beat with other agents as well as those that find no corresponding event to their predictions are terminated. (Dixon 2006; Dixon 2007)

BeatRoot's limitations are focused in the lack of higher level musical knowledge, such as notions of off-beats or expected rhythmic patterns and its tendency to prefer faster rates as they are easier to track, giving corresponding agents higher scores.

B. Davies and Plumbley beat tracking model

In Davies and Plumbley model the first stage of processing, the onset detection function, transforms the audio signal into a more suitable representation using spectral difference, phase deviation and complex domain onset detection functions. The

three different driving functions are used independently, which results in three separate sequences of beat times to be examined later in order to infer the sole output of the system (Davies & Plumbley 2007).

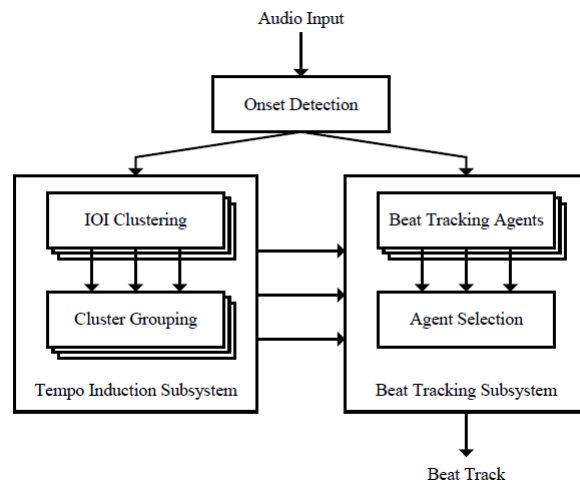


Figure 2. Simon Dixon's BeatRoot system architecture

Outputs of each driving function are passed through an autocorrelation function in order to proceed with the estimate of beat period. Further analysis is required which is based in the use of perceptually weighted (Rayleigh) shift-invariant comb filterbank giving emphasis to periodicities close to 500ms while covering a range of an upper limit of 1.5s. The pair of peaks which are strongest in the filterbank output function and whose periodicities are most closely related by a factor of two are selected as tempo candidates. Beat locations are estimated independently. The known beat period from a previous step serves as a basis for the creation of a suitable impulse train. Then the beat locations are found by cross-correlating this impulse train with each driving function.

Davies and Plumbley beat tracking system works in a two state mode. The first, the General State, extracts the beat period and beat alignment through a process of repeated induction without any prior knowledge of the input. Because no special effort is made to enforce continuity, by using solely the General State can lead to common beat tracking errors such as the switching between different metrical levels i.e. changing to half or double rate without any tempo change and switching between in phase (on-beat) and out-of-phase (off-beat). The role of General State is to infer an initial beat period and detect tempo changes. It uses a Rayleigh weighting within the shift-invariant comb filterbank which gives emphasis to periodicities close to 500ms.

In order to maintain continuity within a given beat period hypothesis a Context-Dependent State is incorporated in the model (Figure 3). The Context-Dependent State operates in a similar way to the General State based on prior knowledge regarding the tempo, time-signature and the locations of past beats. The previous Rayleigh weighting is in this case replaced with a Gaussian weighting, which puts a limitation in the range of observable periodicities and forces the prediction of beats at regular beat period intervals. The system switches between the

two states in order to accommodate beat consistency and tempo changes.

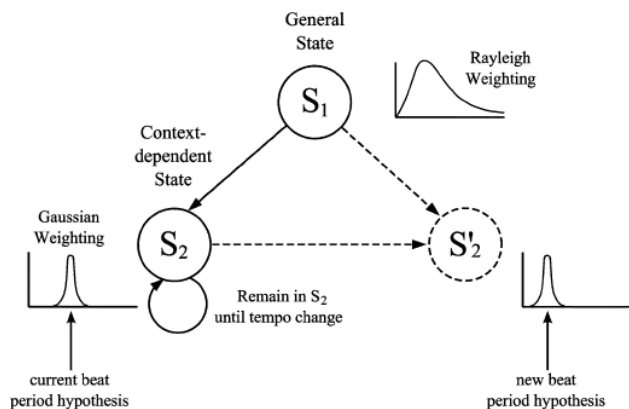


Figure 3. Davies and Plumbley two-state model

C. Pikrakis Meter and Tempo induction model

Pikrakis et al. developed a model that focuses on meter and tempo extraction on polyphonic audio recordings. Audio is processed on a segment by segment basis in non overlapping long-term windows of 10s (Pikrakis et al. 2004). An inner moving short-term window generates a sequence of feature vectors considering energy and chroma based mel frequency cepstral coefficients (MFCCs) (Figure 4).

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, music meter and beat can be jointly estimated by indentifying certain local minima (valleys). Although the beat or the music meter does not always coincide with the global minimum of the plot they can be spotted as approximate multiple or submultiple of that or other local minima.

Further analysis is required in order to associate certain periodicities with the actual beat and meter. Two different ranges of SSM diagonal indices are considered suggesting that beat and meter candidates are lying within respectively. Candidates are examined in pairs using two separate criterions. 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. Finally, the meter of the whole audio is selected regarding its frequency of appearance through histograms that are formed using the calculated meter values per segment.

Tempo estimation process, based on previous results about beat lag, can either extract a value per long-term segment or an average value for the whole audio.

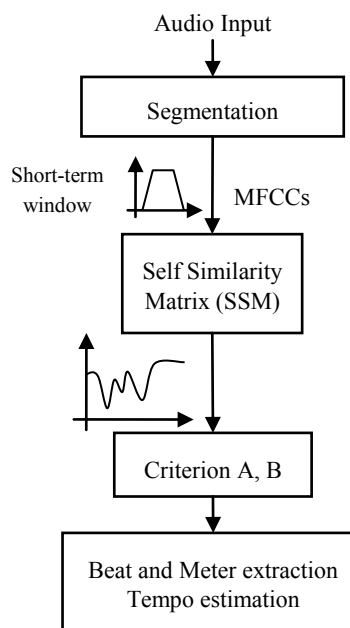


Figure 4. Overview of the architecture of meter and tempo induction model presented by Pikrakis

IV. COMPARISON - EVALUATION

In this study we used a set of 30 Greek traditional songs in order to examine the performance of two state-of-the-art beat-tracking systems and a meter/tempo induction system in processing 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 first two models focus on finding and tracking the beat whereas the latter performs meter extraction; therefore, no direct comparison between them is implied. 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.

We asked three professional musicians to evaluate the performance of these models (using knowledge regarding corresponding dance movements as most of the songs can be danced) and to provide appropriate annotations. The results are summarized in tables 1, 2. For every song relevant information is mentioned such as time signature, tempo and meter (meter is expressed as a rhythmic pattern comprised of groupings of basic temporal units - in this case groups of two or three eighth notes). Regarding the two beat tracking models, the professional musicians assigned the output of the models to the corresponding metrical level and, then, evaluated them. The professional musicians labeled as correct/right (R) the models' output in the cases it coincided with the spontaneous tactus when hearing the song, as wrong (W) when the model tracked the beat erroneously in an isochronous level, and as accepted (A) when the model could correctly track an isochronous periodicity, but at a higher or lower metrical level or shifted to out-of-phase (off-beat).

Pikrakis' et al. model provides a joint estimation of meter and tempo so the corresponding values were evaluated as pairs by the professional musicians. It is more convenient to outline the output of the model as (meter) $x:1$ with tempo t to depict a

pair of dominant periodicities which relate to each other with a proportion of $x:1$ in a rate of t . This notation is more suitable in cases when the real meter can be inferred from the output of the algorithm, for instance, in cases where the estimated meter-tempo is doubled, e.g., estimated 4/8 with 260bpm instead of 2/4 with 130bpm. The professional musicians labeled as correct/right (R) the models' output in the cases it coincided with the real meter and tempo or could be easily transformed to represent it. As accepted (A) were characterized the cases when the model picked up the dominant periodicities but in a different scale. In other words the extracted meter value for a specific tempo corresponded to a number of measures instead of one, e.g. estimated meter 7/8 with tempo 130bpm which corresponded to two measures (i.e., 14/8) instead of the original meter 7/8 with tempo 260bpm. Wrong (W) cases were the ones when the algorithm indicated wrong periodicities as dominant e.g. asymmetric meters (5/8, 7/8) as binary or when the original meter/tempo couldn't be derived from the output.

As can be inferred by the results (table 1), the two beat tracking models encounter difficulties in tracking an asymmetric tactus. Their outputs typically compose an isochronous beat, regardless of the input, resulting in switching between on-beat and off-beat in the majority of songs. Simon Dixon's BeatRoot has a tendency to produce faster beats which is probably due to the agents' functionality and the fact that faster rates are easier to track. Examining table 1 for this model it is obvious that there is no output assigned at the metrical level. Davies and Plumbley's model exhibits a slightly better performance shifting in many cases in a higher metrical level which is isochronous and can be marked as acceptable. Their model seems more prone to out-of-phase errors considering for example a wrong grouping of 3-2-2 instead of the correct 2-2-3 in the case of a song with time signature 7/8.

On the other hand the meter and tempo induction model introduces a better performance in processing asymmetric rhythms (table 2). It seems to cope well with non-binary meters in most of the cases as it can recognize the original time signature. In some instances it seems to designate as more dominant periodicities the ones that refer to a span of two measures (songs no. 14, 17, 18, 19, 23, 26). Many of these results could probably be corrected by performing post processing using knowledge from musicology. For example it is more probable for a song marked as 9:1 with tempo 130 to be 9/8 with double tempo. It is worth pointing out that even in cases that the model outputs a wrong meter calculation the actual meter value resides in the histograms but with a lower peak. This is the case for songs no. 10, 13, 15, 16, 30. The majority of the songs in which the algorithm falls into a wrong output are cases with too fast or too slow tempi.

V. CONCLUSION

In this work we examined the performance of two state-of-the-art beat-tracking systems and a meter/tempo induction system in processing asymmetric rhythms. The beat-trackers encounter difficulties in identifying and tracking a non isochronous beat. Their output is typically isochronous which is acceptable in some cases such as when it coincides with sub-beat or metrical level. This performance could

probably be improved by incorporating knowledge about hierarchic metrical structure so that processing and combining data from multiple metrical levels could be more effective.

The metre/tempo induction system is more successful in dealing with non-binary meters. The original time signature is recognized correctly in most cases even if it may correspond sometimes to a two measure span. As periodicities (e.g. phrases) are not restricted within the time span of one measure those cases can be qualified as acceptable or even correct. Generally, the correct beat level seems to exist implicitly in the model which provides the base to improve sufficiently its performance by using an appropriate post-processing method.

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 CADM-TR-09-06*, Queen Mary University of London, Centre for Digital Music.
- Davies, M. E. P. & Plumbley, M. D. (2007). Context-dependent beat tracking of musical audio. In *IEEE Transactions on Audio, Speech and Language Processing*. 15(3):1009-1020.
- Dixon, S. (2006). MIREX 2006 Audio beat tracking evaluation: BeatRoot. Retrieved from http://www.musicir.org/evaluation/MIREX/2006/abstracts/BT_dixon.pdf.
- Dixon S. (2007). Evaluation of the audio beat tracking system BeatRoot. *Journal of New Music Research*, 36(1):39-50.
- Hockman J. A. (2008). An Overview of Beat Tracking Techniques. Retrieved from http://www.music.mcgill.ca/~hockman/coursework/MUMT_611/final/jhockman_611final.pdf
- 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.
- 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. Performance of the two beat tracking models in tracking an asymmetric beat using a set of 30 Greek traditional songs. For both models it is indicated whether they track the beat correctly (Beat level - B) or they shift to metrical level (M) or sub-beat level (SB), or they produce out-of-phase output (*). Annotations (AN) were provided by three professional musicians: Correct/Right (R), Wrong (W), Accepted (A).

	Song's Name	Time Signature	Tempo	Metre	Dixon's Beatroot				Davies & Plumbly model			
					SB	B	M	AN	SB	B	M	AN
1	Sousta Rodou	2/4	144 (1/4)			1/4		R		1/4		R
2	Mpalos	2/4	82 (1/4)		1/8			A		1/4*		A
3	Ehe geia panagia (Hasapiko)	2/4	130 (1/4)			1/4*		A		1/4*		A
4	Tsamikos	3/4	98 (1/4)		1/8			A		1/4		R
5	Apopse mavromata mou	3/4	104(1/4)			1/4		R		1/4		R
6	Valtetsi	3/4	108(1/4)			1/4		R		1/4*		A
7	Armenaki	4/4	180(1/4)		1/4			A		2/4		R
8	Louloudi ti marathikes	4/4	127(1/4)		1/4			A	1/4			A
9	Zagorisios - Kapesovo	5/8	94 (1/8)	2-3	1/8			A	1/8			A
10	Mpaintouska Thrakis	5/8	420 (1/8)	2-3		2/8		W			5/8*	A
11	Itia	6/8	201(1/8)	4-2		2/8		A		2/8		A
12	Enas aitos kathotane	6/8	209(1/8)	4-2		2/8		A		2/8		A
13	Zonaradiko Thrakis	6/8	429 (1/8)	3-3		3/8		R		3/8		R
14	Perasa ap' tin porta sou	7/8	264(1/8)	3-2-2		2/8		W		2/8		W
15	Tik Tromakton Pontos	7/8	488 (1/8)	2-2-3		3/8		W			7/8*	A
16	Mantilatos Thrakis	7/8	483 (1/8)	2-2-3		3/8		W			7/8*	A
17	Mantili Kalamatiano	7/8	273 (1/8)	3-2-2		2/8		W		3/8		W
18	Milo mou kokkino	7/8	268 (1/8)	3-2-2		2/8		W		3/8		W
19	Na diokso ta synnefa	7/8	266 (1/8)	3-2-2		2/8		W		2/8		W
20	Oles oi melahroines	8/8	381 (1/8)	3-3-2		2/8		A			4/8*	W
21	Dyo mavra matia agapo	8/8	396(1/8)	3-3-2		2/8		A			4/8	A
22	Feto to kalokairaki	9/8	136(1/8)	2-2-2-3	1/8			A	1/8			A
23	Karsilamas	9/8	256 (1/8)	2-2-2-3	3/16			W		2/8		W
24	Zeimpekiko neo	9/8	60 (1/8)	2-2-2-3	1/16			A	1/16			A
25	Amptaliko neo	9/8	104 (1/8)	3-2-2-2	1/8			A	1/8			A
26	Tsiourapia Makedonias	9/8	276 (1/8)	2-2-2-3		2/8		W		2/8		W
27	Karsilamas - Ti ithela	9/8	288 (1/8)	2-2-2-3		2/8		W		3/8		W
28	Ela apopse stou Thoma	9/8	185 (1/8)	2-2-2-3	1/8			A		2/8		W
29	Zagorisios Ipeiros	10/8	115 (1/8)	2-2-2-2-2		3/8		W		2/8		R
30	Patrounino makedonias	11/8	240 (1/8)	3-2-2-2-2		2/8		W		2/8		W

Table 2. Meter/Tempo induction evaluation for Pikrakis et al. model using a set of 30 Greek traditional songs. Annotations (AN) were provided by three professional musicians who examined the output values as pairs.

					Pikrakis et al. model		
	Song's Name	Time Signature	Tempo	Metre	Calc. Tempo	Calc. Meter	AN
1	Sousta Rodou	2/4	144 (1/4)		285	4:1	R
2	Mpalos	2/4	82 (1/4)		171	4:1	R
3	Ehe geia panagia (Hasapiko)	2/4	130 (1/4)		260	4:1	R
4	Tsamikos	3/4	98 (1/4)		98	3:1	R
5	Apopse mavromata mou	3/4	104(1/4)		206	6:1	R
6	Valtetsi	3/4	108(1/4)		214	6:1	R
7	Armenaki	4/4	180(1/4)		181	4:1	R
8	Louloudi ti marathikes	4/4	127(1/4)		260	8:1	R
9	Zagorisios -Kapesovo	5/8	94 (1/8)	2-3	97	2:1 or 5:1	A
10	Mpaintouska Thrakis	5/8	420 (1/8)	2-3	83	4:1	W
11	Itia	6/8	201(1/8)	4-2	208	6:1	R
12	Enas aitos kathotane	6/8	209(1/8)	4-2	206	6:1	R
13	Zonaradiko Thrakis	6/8	429 (1/8)	3-3	77	4:1	W
14	Perasa ap´tin porta sou	7/8	264(1/8)	3-2-2	130	7:1	A
15	Tik Tromakton Pontos	7/8	488 (1/8)	2-2-3	73	2:1	W
16	Mantilatos Thrakis	7/8	483 (1/8)	2-2-3	69	2:1 or 3:1	W
17	Mantili Kalamatiano	7/8	273 (1/8)	3-2-2	132	7:1	A
18	Milo mou kokkino	7/8	268 (1/8)	3-2-2	133	7:1	A
19	Na diokso ta synnefa	7/8	266 (1/8)	3-2-2	130	7:1	A
20	Oles oi melahroines	8/8	381 (1/8)	3-3-2	193	4:1	A
21	Dyo mavra matia agapo	8/8	396 (1/8)	3-3-2	200	4:1	A
22	Feto to kalokairaki	9/8	136 (1/8)	2-2-2-3	139	9:1	R
23	Karsilamas	9/8	256 (1/8)	2-2-2-3	130	9:1	A
24	Zeimpekiko neo	9/8	60 (1/8)	2-2-2-3	61	9:1	R
25	Amptaliko neo	9/8	104 (1/8)	3-2-2-2	109	9:1	R
26	Tsiourapia Makedonias	9/8	276 (1/8)	2-2-2-3	146	9:1	A
27	Karsilamas - Ti ithela	9/8	288 (1/8)	2-2-2-3	96	9:1	A
28	Ela apopse stou Thoma	9/8	185 (1/8)	2-2-2-3	181	9:1	R
29	Zagorisios Ipeiros	10/8	115 (1/8)	2-2-2-2-2	114 or 228	10:1	R
30	Patrounino makedonias	11/8	240 (1/8)	3-2-2-2-2	65 or 222	2:1	W