Tension perception in Greek traditional folk music: examining the role of timbral semantics.

Asterios Zacharakis
Aristotle University of Thessaloniki
aszachar@mus.auth.gr

Konstantinos Pastiadis
Aristotle University of Thessaloniki
pastiadi@mus.auth.gr

Athena Katsanevaki
Aristotle University of Thessaloniki
atekatsa@mus.auth.gr

ABSTRACT

This paper presents an empirical experiment aiming to investigate the potential influence of timbral semantics on tension induction in Greek traditional folk music. To this end, a group of seventeen listeners rated the evolution of auditory *luminance*, *texture* and *mass* together with the felt *tension* over sixteen musical excerpts in real-time. Correlation and regression analyses between these four quality profiles for each particular stimulus showed that all three examined timbral qualities had instances of very strong association with tension. Although auditory mass featured the greatest number of such instances, no safe conclusion can be reached based on current findings regarding the most influential timbral semantic dimension for tension induction. Instead, it seems that a combination of conditions (i.e., musical parameters) can either maximise or minimise the influence of each timbral dimension.

1. INTRODUCTION

Music is believed to draw a significant amount of its appeal from its ability to stimulate emotional responses that alternate between tension and relaxation (e.g., Huron, 2006; Lehne & Koelsch, 2015). Krumhansl (2015) sums up the cognitive view of musical tension by stating that it is created when an expected event is delayed or when the context is ambiguous. In general, the tension-relaxation phenomena in Western music have mostly been studied with a focus on harmony, melody, dynamics and rhythm, while timbre –being a domain-general psychological feature as Farbood (2012) puts it- has been largely underrepresented with the notable exception of Pressnitzer et al. (2000). However, Farbood & Price (2017) have recently investigated some timbral attributes with respect to tension and reported that higher degrees of roughness, inharmonicity and spectral flatness of musical tones were associated with higher tension ratings. The authors of this work suggest that the effect of timbre on tension should be further investigated in more ecologically valid settings.

In Greek traditional folk music extreme manipulations of expectations are not the norm. With the exception of improvisational parts, melodic and rhythmic structures tend to be repetitive, the orchestrations are generally fixed throughout a given song and climactic moments - that according to Huron (2006) constitute the epitome of a tension-relaxation schema in music making- are rare. Does this mean that tension build up is not intended by the folk music creators and in turn not experienced by the listeners? Or could it be that in more predictable musical creations, the timbral characteristics of a piece may have an important role to play with respect to conveying tension? If it so, which of the timbral qualities are most influential?

This study will aim to address the above questions based on the previously developed luminance-texture-mass (LTM) framework for musical timbre semantics (Zacharakis, Pastiadis & Reiss, 2014; 2015; Zacharakis & Pastiadis, 2016). The LTM framework suggests that the most salient semantic dimensions of musical timbre are luminance (i.e., bright vs. dull), texture (i.e., rough vs. smooth) and mass (i.e., full vs. empty). Therefore, this work will seek to identify for possible associations between the three dimensions of the LTM framework and felt tension in Greek folk music.

2. METHOD

A selection of 16 instrumental excerpts from various types of Greek traditional folk music (e.g., dances, dirges, Akritika, from Epirus, Asia Minor, Aegean islands, etc.) also including a variety of lead instruments was presented to seventeen listeners. Most excerpts constitute introductory or improvisational parts and range from 24 seconds to 68 seconds long. Vocal parts were avoided partly due the increased timbral heterogeneity (Sandell, 1995) that is introduced by a human voice and partly due to the semantic charge of the lyrics whose impact could not be controlled. The stimuli can be accessed online at: http://ccm.web.auth.gr/timbreandtension.html. Stimuli were equalised in loudness through informal listening within the research team and their RMS playback level was measured to be between 65 and 75 dB SPL (A-weighted, slow response). The presentation of the stimuli was made using a pair of PreSonus HD7 headphones. All listeners reported an equal loudness for all stimuli.

The participants (17) were students at the School of Music Studies of the Aristotle University of Thessaloniki (6 male, mean age: 21, average years of musical practice: 12.4) and they received course credit as compensation.

Participants rated continuously and in real-time the change in the timbral qualities *luminance*, *texture* and *mass* according to the LTM model, plus the felt *musical tension*[1] of each excerpt. The timbral qualities were orally

[1] Felt musical tension refers to the amount of tension that is actually felt by the participant as opposed to the amount of tension that he/she thinks that the stimulus is supposed to express.

elaborated very briefly by defining their two extremes to ensure that listeners have a common understanding of the concept. Positive luminance was defined as auditory brightness and negative luminance as auditory dullness, positive texture as auditory roughness and negative texture as auditory smoothness, finally positive mass as auditory fullness and negative mass as auditory emptiness. At this point, it has to be noted that the LTM framework for timbral semantics has been developed by empirical experiments on monophonic timbres. Therefore, the assessment of polyphonic music based on the LTM components is a novel element of this study. The concept of tension was elaborated as *inner tension (εσωτερική ένταση)* in order to avoid confusion that may have arisen due to the fact that the word for tension in Greek coincides with the term for sound volume. Inner tension was defined similarly to Farbood & Price (2017) as: *less tension corresponds to a feeling of relaxation or resolution, while more tension corresponds to the opposite direction*.

The rating device used was a small (16.9 x 21 cm) Wacom Intuos draw pen tablet set up like a mouse. Movement of the pen on the tablet on the right indicated increase of the quality under judgement while movement on the left indicated decrease of the quality. The values obtained were not limited by the physical dimensions of the tablet since the participant could simply reposition the pen on the tablet to get more available space just like he/she could do with a normal mouse. The data acquisition interface was custom designed in LabVIEW. It sampled the pen's horizontal axis coordinate every 5 milliseconds and offered participants a real-time visualisation of the profiles they were creating.

Rating on each of the four components was made in blocks of random order for each participant. In addition, the presentation order of the stimuli within each block was also randomised. As a result, all listeners eventually listened to each stimulus four times. The duration of the experiment was a little over an hour for most of the participants not including breaks (which they were advised to take whenever necessary in order to keep their concentration levels high).

3. ANALYSIS & RESULTS

Raw responses were subsampled by calculating the mean value over adjacent non-overlapping rectangular time windows (.5 secs = 10 samples). The resulting time series were subjected to first-order differentiation and replacement of positive/negative values with 1 and -1, respectively. The time series were next integrated and each participant's data were normalised within each quality by his/her maximum rating on this particular quality. Finally, the profiles were smoothed using a cubic spline interpolant and linear trends were removed from each individual participant's time series to ensure 'stationarity' (Dean & Bailes, 2010).

The inter-participant reliability analysis showed a good agreement for all qualities under study with luminance, texture, mass and tension featuring a Cronbach's Alpha of .81 and .86, .88 and .86 respectively.

Therefore, the processed time series were averaged over every stimulus and each of the four qualities under study. Figure 1 presents the mean profiles along with their 95% confidence intervals. Correlation analysis between the averaged time series revealed the relationships between timbral semantics and tension. The Pearson's correlation coefficient between tension and the timbral qualities luminance, texture and mass are presented in table 1. Overall, tension variation seems to be associated with variation in all three timbral semantic components to different extends depending on the particular stimulus.

Stimulus	Lead instrument	Luminance	Texture	Mass	Tension MIRToolbox
In a foreign land since a little boy	Qanun	.50**	.71**	.85**	.30*
The Rasti	Bagpipes	.94**	.89**	.89**	.41**
Zonaradikos Dance	Bagpipes	-	.50**	-.34*	.30*
The King throws a party	Violin	.80**	.91**	.92**	-
Dirge and Stroto Pogonisio	Clarinet	.23*	.75**	.40**	.35**
Lament	Lute	.69**	.95**	.87**	-
Karsilamas	Politiki Lyra	.79**	.24*	.69**	.41**
Tik dance	Pontiaki Lyra	.89**	.76**	.96**	.27*
Dirge from Epirus	Ney	.87**	.25*	.92**	.63**
Servikos dance	Ney	.71**	.96**	.31*	.52**
Yannis and the dragon	Violin	-	.78**	.80**	.77**
If you are going to foreign lands	Ney	.90**	.97**	.74**	-
Trygona	Santouri	.65**	-.48**	.72**	.39**
Sousta dance of Patmos	Bagpipes	.68**	.56**	.48**	.32*
The little Vlach boy	Clarinet	.95**	.90**	.97**	-.31*
Sebastian dance	Tabouras	.92**	.34*	.27*	-.60**

Table 1. Pearson's correlation coefficients between tension profiles and the profiles of the timbral qualities luminance, texture, mass together with tension calculated by the MIR Toolbox's miremotion. (**: p<.001, *:p<.05)

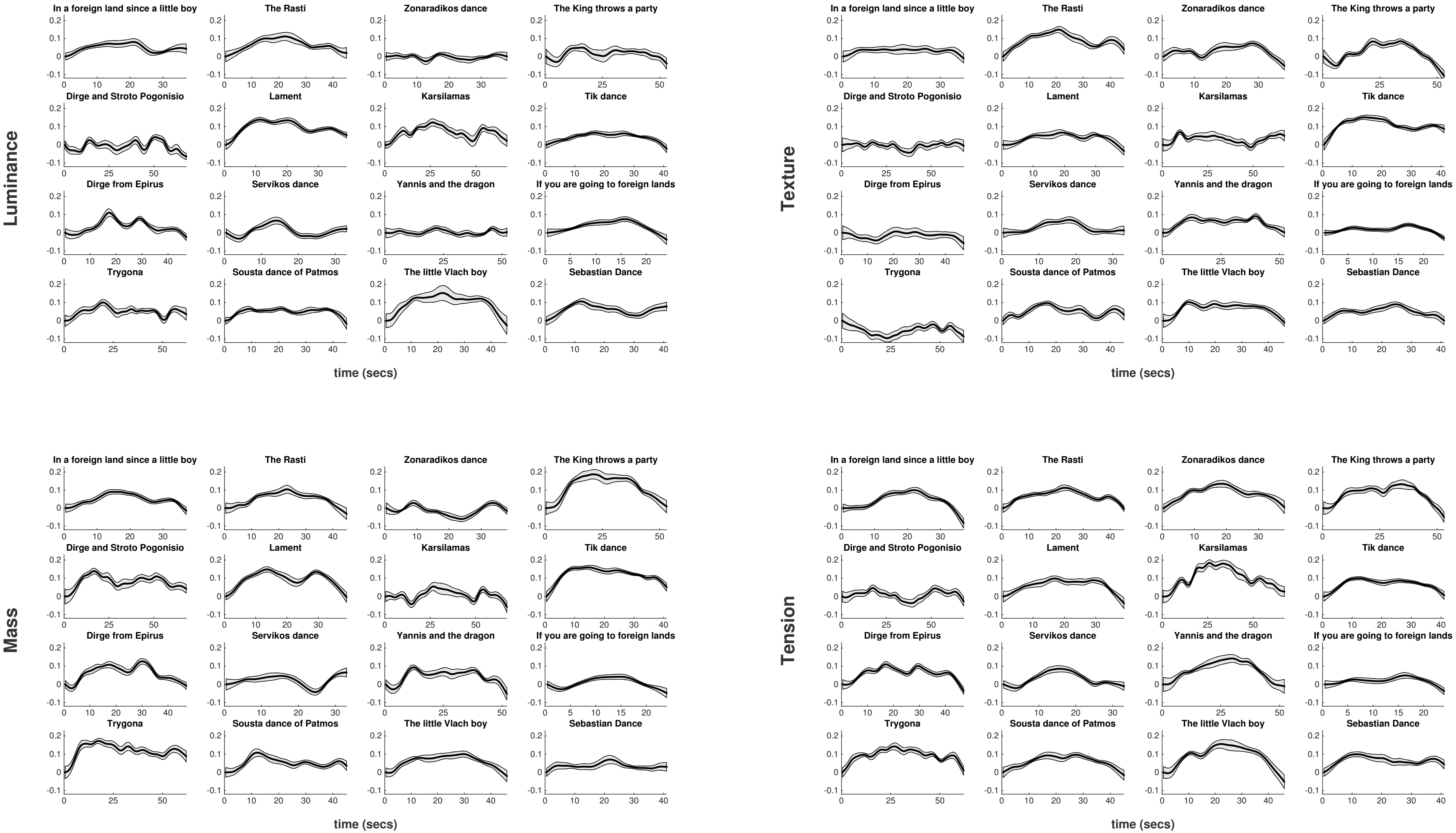

Figure 1. Mean temporal profiles corresponding to luminance, texture, mass and tension (in black) for each stimulus along with the 95% confidence intervals area in light grey. The values on the y-axis result from the normalisation process described in section 3.

In addition, a computational tension prediction calculated from the audio signal according to the *miremotion* function (Eerola et al., 2009) included in the MIR Toolbox (Lartillot & Toiviainen, 2007) was also compared to the empirically acquired tension profiles. The MIR tension estimation is based, among others, on calculation of dynamic, tonal and rhythmic variations. The window used for calculating tension through miremotion was 4 seconds long with 50% overlap. Subsequently the time series of the tension calculation were linearly extrapolated to exactly match the number of samples corresponding to the empirical tension profile for each stimulus. Finally, the time series of the miremotion tension were also smoothed using a cubic spline interpolant. Figure 2 shows the sixteen tension profiles that resulted from the above procedure and the last column of table 1 presents the Pearson's correlation coefficients between the empirical and the calculated tension profiles for each stimulus. With the exception of 'Yannis and the Dragon' where the relatonships between the MIR-tension and the

Stimulus	**Standardised beta**				**R-squared**
	Luminance	Texture	Mass	Tension MIRToolbox	
In a foreign land since a little boy			.85		.72
The Rasti	.90			.15	.91
Zonaradikos Dance		.51		.34	.36
The King throws a party			.92		.84
Dirge and Stroto Pogonisio		.71		.21	.62
Lament			.87		.76
Karsilamas	.75			.29	.72
Tik dance			.96		.93
Dirge from Epirus			.80	.29	.92
Servikos dance		.96			.93
Yannis and the dragon			.52	.41	.73
If you are going to foreign lands		.98			.95
Trygona			.66	.17	.55
Sousta dance of Patmos	.66			.24	.52
The little Vlach boy			.97		.94
Sebastian dance	.92				.84

Table 2. Multiple regression models for each stimulus using tension as dependent variable and the three timbral qualities plus tension calculated by the MIR Toolbox as predictors.

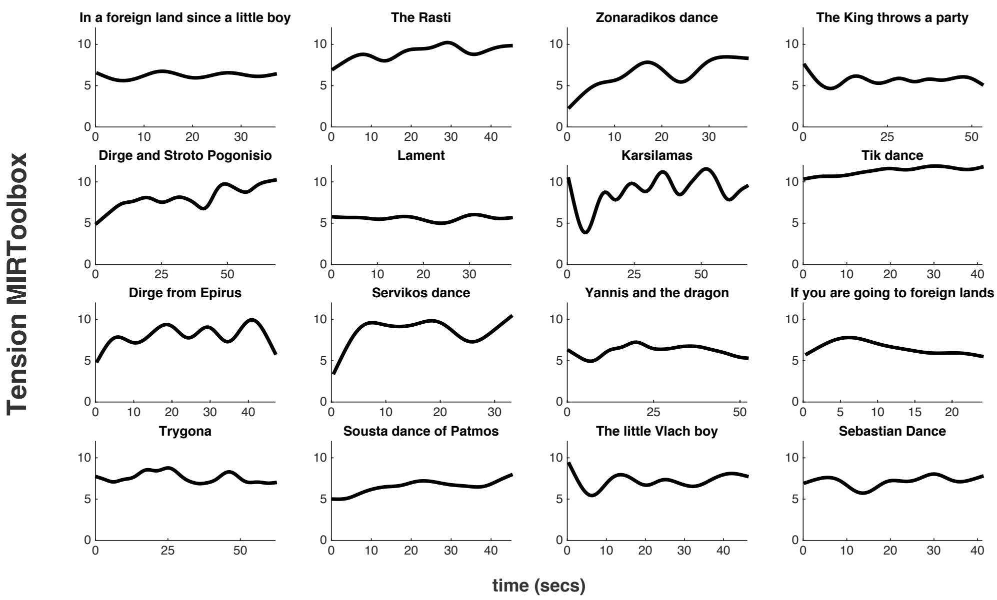

Figure 2. Temporal tension profiles calculated from the miremotion function of the MIR Toolbox.

LTM components with tension is comparable, in all other cases the MIR tension is more weakly correlated to tension compared to the LTM components, albeit in most cases the correlations are statistically significant.

Wanting to examine the influence of the timbral semantics along with other musical structures represented by tension-MIR we performed a two-step linear multiple regression analysis with one of the LTM components and MIR-tension as predictors and tension as the dependent variable. In the first step we picked the best predictor out of the LTM components and in the second step we examined whether additional inclusion of the MIR-tension contributed towards a better prediction of tension. The models were evaluated based on a combined maximisation of the explained variance (i.e., R-squared) and minimisation of the Akaike Information Criterion (AICc; Hurvich & Tsai, 1989). Table 2 succinctly presents the models that were favoured through this process.

Mass features eight appearances as the best predictor while *luminance* and *texture* appear only four times each, thus implying that *mass* may be more influential than the other two semantic components for tension perception. Also, in half of the stimuli, pairing MIR-tension with one of the LTM predictors contributes to a better model. Overall, the amount of tension variance explained (reflected by the R-squared values) is in many cases very high, as expected by the generally high correlations between tension and the LTM components.

4. DISCUSSION

This study constitutes a preliminary attempt to investigate a potential relationship between timbral semantics and tension perception using traditional folk Greek music as a vehicle. It should be viewed as a stimulation for further discussion on this thorny topic rather than a study that provides definitive answers having in mind that this approach is novel not only in the field of folk music but also in music perception literature in general. Using real-world polyphonic music as stimuli is particularly challenging for various reasons. First of all, in such a scenario all musical parameters (e.g., melodic contour, rhythmical patterns, dynamics, expressivity, timbre, etc.) vary concurrently and cannot be easily isolated. Secondly, the timbral heterogeneity (Sandell, 1995) inherent in some of our polyphonic stimuli makes the rating of timbral semantics of a whole excerpt of music a non-trivial task. Considering this fact, the agreement exhibited for the timbral ratings in particular is impressive.

All tension profiles except for 'Dirge and Stroto Pogonisio' feature one or two tension peaks at some point during the stimulus that are statistically significant (i.e., the lowest limit of their 95% confidence intervals is higher than the highest limit of the 95% interval of the 1st second of the stimulus). Such tension profiles have been typically identified in music perception literature (Krumhansl, 2015). The 'Dirge and Stroto Pogonisio' features a statistically significant minimum at the second half of the stimulus length. Even this sole appearance of decreasing tension profile shows that our participants were not biased towards reporting a bell-type tension profile at all instances. The LTM profiles feature a higher variability in shape, including increasing, decreasing, relatively stable and fluctuating profiles.

At this point, it has to be noted that the design of this experiment does not allow making any judgement with respect to the absolute value of the qualities under study for each stimulus. This is because participants judged only the variation of the qualities in time but did not have a way to inform us of the initial value of each quality. That is to say, a quality profile that does not vary much (e.g., the luminance profile for the 'Zonaradikos dance') does not necessarily imply a low overall judgement of this quality. The information on the initial absolute values of the four qualities for all our stimuli is going to be obtained and exploited in future work.

Such information will help to better examine the influence of timbre on perceived tension and to properly validate the perspective suggested by Huron (2006) according to which tension can be viewed as the dynamic subcategory of a generalised feeling of uneasiness that he calls *dissonance*. The other branch of this general dissonance is the static *sensory dissonance* in which timbral information belongs to. In this sense, the stress induced by static sensory means, such as timbre, should not be strictly viewed as tension but rather as a feeling of uneasiness. However, participants' agreement on tension ratings –reported in the present as well as in other studies-suggests that the terms tension and uneasiness can probably be used interchangeably.

Despite the above described caveats, our results demonstrate that timbral semantics are in many cases very strongly correlated with tension perception and that tension as calculated by the miremotion from the MIR Toolbox can -in half of the stimuli- be used in combination with LTM components to better account for felt tension variation. That is, of course, not to say that timbre is the most important attribute for tension perception. Since correlations do not imply causal relationships, it could well be the case that an underlying musical parameter (such as melodic contour, rhythmical density or dynamics) is a latent variable, affecting both timbre and tension perception. This hypothesis will also be investigated in future work.

In general, probably the most important finding of this study is that all three components of the LTM timbral semantics framework can potentially influence tension perception. This, however, does not always happen in a consistent manner as demonstrated by table 1. As an example, the two stimuli featuring the highest tension peaks ('Karsilamas' and 'The little Vlach boy') also feature the highest (‘The little Vlach boy’) and third highest luminance peaks (‘Karsilamas’). At the same time, however, the lowest ranked stimuli in terms of maximum luminance (‘Yannis and the Dragon’ and ‘Zonaradikos dance’) are ranked third (‘Zonaradikos dance’) and fifth (‘Yannis and the Dragon’) in terms of maximum tension. The same also stands for texture and mass. These results may pose a partial challenge to past findings supporting that sonorities of higher roughness and mass are tension-provoking means in music (Pressnitzer et al., 2000; Huron, 2006; Farbood & Price, 2017) by suggesting that this may only be conditionally true.

A similar type of inconsistency emerges from a preliminary musicological analysis of our excerpts. For ‘Karsilamas’ (featuring one of the highest maxima in both luminance and tension profiles) the peak in the profiles seems to coincide with a rise in melodic pitch. The same stands for ‘Servikos dance’, ‘Rasti’, ‘The King throws a party’ and ‘Yannis and the dragon’. A notable exception, however, is ‘The little Vlach boy’ where while pitch decreases, tension rises to reach the maximum peak out of all tension profiles. In this case, the violation of the norm could be attributed to the existence of strong melodic attractions and existence of chromatisism. The above examples, demonstrate that tension perception is a multifaceted phenomenon that is not influenced by one attribute alone.

Some other types of idiosyncratic elements that affect the profiles of our qualities can also be reported. The small fluctuations evident in the profiles of ‘Trygona’ could probably be due to the impulsive character of Santouri, while local rises in tension, luminance or mass could be attributed to glissandi or melodic embellishments (e.g., in ‘Rasti’, ‘Dirge and Stroto Pogonisio’, ‘Dirge from Epirus’ and ‘Servikos dance’). A detailed musicological analysis in respect with the acquired LTM and tension profiles will be another scope of future work.

Finally, another issue that warrants further investigation is a fine-tuning of the window size for the MIR-tension calculation. The selection of a four-second window was made in this work as a reasonable choice that would successfully simulate human reaction time and working memory. However, other possible lengths and/or temporal adjustments may yield better associations with the empirical data.

5. ACKNOWLEDGEMENTS

This study was supported by a post-doctoral scholarship issued by the Greek State Scholarships Foundation (grant title: "Subsidy for post-doctoral researchers", contract number: 2016-050-050-3-8116), which was co-funded by the European Social Fund and the Greek State.

The authors would like to thank the participants of this study.

6. REFERENCES

Dean, R. T., & Bailes, F. (2010). Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empirical Musicology Review, 5(4),* 152-175.

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models. In *Proceedings of the International Society of Music Information Retrieval Conference (ISMIR 10)* (pp. 621-626).

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29(4)*, 387-428.

Farbood, M. M., & Price, K. C. (2017). The contribution of timbre attributes to musical tension. *The Journal of the Acoustical Society of America*, *141*(1), 419-427.

Huron, D. B. (2006). *Creating tension* in Sweet anticipation: Music and the psychology of expectation. MIT press: London, UK.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297-307.

Krumhansl, C. L. (2015). Statistics, structure, and style in music. *Music Perception, 33(1)*, 20-31.

Lartillot, O., & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)* (pp. 237-244).

Lehne, M., & Koelsch, S. (2015). Towards a general psychological model of tension and suspense. *Frontiers in Psychology*, 6, 79.

Pressnitzer, D., McAdams, S., Winsberg, S., & Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & psychophysics*, *62*(1), 66-80.

Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining" blended" instrument pairings in orchestration. *Music Perception, 13(2)*, 209-246.

Zacharakis, A., & Pastiadis, K. (2016). Revisiting the luminance-texture-mass model for musical timbre semantics: A confirmatory approach and perspectives of extension. *Journal of the Audio Engineering Society*, *64(9)*, 636-645.

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2014). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception, 31(4)*, 339-358.

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2015). An interlanguage unification of musical timbre: bridging semantic, perceptual, and acoustic dimensions. *Music Perception*, *32(4)*, 394-412.