



Interaction between time-varying tone inharmonicity, fundamental frequency and spectral shape affects felt tension and timbral semantics

Asterios Zacharakis* and Konstantinos Pasiadis

School of Music Studies, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Received 24 October 2019, Accepted 8 February 2021

Abstract – Musically induced tension has been the subject of thorough study in the music cognition literature but its relationship with timbre is still poorly investigated. This study examines how the dynamic variation of a tone's inharmonicity may affect a number of auditory qualities, namely brightness, roughness and mass along with felt tension under different acoustical conditions (i.e., F_0 , spectral shape and type of inharmonicity). Fifty-six musically trained participants gave real-time continuous ratings on eight time-varying stimuli upon the aforementioned qualities. Static ratings over the initial purely harmonic parts of the stimuli were also obtained by a subgroup of the listening panel. The fundamental frequency exhibits the strongest influence on the responses of the four qualities, followed by the type of inharmonicity and the spectral shape to a lesser degree. The profile patterns of mass and brightness proved to be strong predictors for tension profile patterns while the roughness profile magnitudes show a significant main effect on the magnitude of tension profiles. Overall, these results demonstrate that time-varying inharmonicity affects continuous responses on both timbral semantics and tension, while indicating that felt tension may be influenced by underlying timbral qualities in a dynamic context.

1 Introduction

The alternation between tension and relaxation induced by music listening has been argued to constitute an important mechanism through which music conveys emotional content and becomes pleasurable (e.g., [1, 2]). Therefore, investigation and modelling of the musical parameters that are capable of inducing tension to a listener has been a popular research topic in the last few decades (e.g. [1, 3–7]). However, none of the above studies has dealt with this topic from the perspective of timbre, as most focused on the role of harmony, melody and dynamics. This is justified by the fact that both timbre and musically induced tension are complex, multidimensional entities with not even a clear-cut definition. Still, the few studies that have examined timbre and tension in combination have indicated that aspects of the former influence the latter. In one of the earliest attempts by Paraskeva and McAdams [8], it has been suggested that a mere change of orchestration (solo piano vs. orchestra) can affect the ratings of completeness – defined

as the opposite of tension – for the same musical phrase. A later work by Pressnitzer et al. [9] focused on the timbral quality of roughness and showed that it was capable of carrying tension-relaxation information even in isolated non-tonal orchestral chords. In addition, a few more recent studies suggested that timbre can affect the emotion conveyed by melodies [10] and even isolated chords [11].

Although these studies support the idea that musical timbre is indeed affecting the experience of tension and other emotions, they mostly use the term timbre as an equivalent of instrumental identity and as a result they report no acoustic correlates of tension. A recent study by Farbood and Price [12] adopted a different approach by using isolated synthesised sound stimuli in a pairwise presentation set-up and confirmed that differences in tension ratings could be systematically attributed to mere timbre differences decoupled from instrumental identity. Furthermore, it was able to identify certain acoustic parameters (i.e., tone inharmonicity, roughness and spectral flatness) as the most influential for tension elicitation.

The findings of Farbood and Price [12] demonstrated the significance of inharmonicity in inducing tension. In addition, the perceptual salience of inharmonicity was highlighted in two occasions featured in our own previous research. The first finding concerned its relationship with two semantic dimensions of timbre. The notion of timbral semantics refers to the verbal terms used to describe timbral qualities and can be used to break the multidimensional

*Corresponding author: aszachar@mus.auth.gr

The display of Audio files embedded in this article depends on the software used (PDF reader, video player, installed codec, direct display in the browser, etc.). Please see the [Adobe Acrobat page](#) for more explanation. It is best to listen to the Audio files with headphones as otherwise the changes in the stimuli might not be perceptible as intended. You can also find the audio files at <https://doi.org/10.5281/zenodo.4626964>

concept of timbre down to more clearly defined qualities. Our previous work examined an extended descriptive vocabulary and identified the three most salient general concepts for timbre description [13]. The first was named *luminance* and encapsulated terms such as bright and deep, the second was called *texture* and contained terms such as rough, soft, rounded and harsh and the third was called *mass* containing terms such as dense, rich, full, thick or light. It was shown that inharmonicity was related to both the luminance and the mass dimensions of the luminance-texture-mass framework for timbral semantics [13]. The second finding concerned its sensitivity under background noise. In [14] we showed that although inharmonicity was indeed a prominent acoustic correlate for the dimensions of a timbre space of synthetic sounds, the feature became less perceptually salient in the condition where sounds were presented together with white noise in the background (i.e., inharmonicity could not predict any dimension of the timbre space obtained for the noisy condition). The above evidence suggests that inharmonicity may constitute a factor at play in both timbrally induced tension and the semantics of timbre.

In the work by Farbood and Price [12] the examined tones were static with fixed amounts of inharmonicity and the responses were given based on a pairwise comparison with relation to felt tension. However, since musically induced tension is a dynamic entity, it is very often measured through continuous real-time ratings in response to musical stimuli (e.g., [5]–[7], [15–20]). The presented study adopted this rating paradigm to obtain continuous responses regarding the brightness, roughness, mass and tension of synthetic tones with time-varying inharmonicity. This way, it becomes possible to examine whether the findings by Farbood and Price [12] regarding tension and inharmonicity apply to dynamically changing sounds as well. In addition, with this experiment we seek to explore a potential relationship between felt tension and the semantics of timbre. It is possible that this relation may also be mediated by acoustic parameters such as the fundamental frequency, the spectral content and the type of inharmonicity alteration itself. Therefore, this work examined two different conditions for each of these acoustic attributes to test for their potential influence on the ratings of the qualities.

The main questions addressed with this study are the following: (1) Does a dynamic alteration of inharmonicity influence felt tension profiles similarly to what was suggested by Farbood and Price for static sounds [12]? In addition, are semantic qualities of timbre similarly affected? (2) Can F_0 , spectral shape or the type of introduced inharmonicity account for potential changes in timbral semantics and felt tension? (3) Finally, how do possible changes of timbral semantics may relate to a possible change of felt tension?

2 Method

2.1 Main listening test

2.1.1 Listening panel

Fifty-six musically trained listeners (mean age: 22.5 years old, standard deviation: 5.54 years, average years of

musical practice: 12.9, 34 female, 22 male) took part in the main listening test. The participants were students from the School of Music Studies of the Aristotle University of Thessaloniki and they were given course credit as compensation for their participation. None of them reported any hearing loss.

2.1.2 Stimuli

The stimuli were produced using a custom built Max/MSP additive synthesiser. Two different pitches (220 and 440 Hz) and two different types of spectral shapes (one sawtooth wave and one square wave) were used. The sawtooth waves consisted of 30 partials and the square waves of 15 partials. The inharmonicity of these four initial stimuli was varied in two different manners. The first introduced a positive displacement of frequencies according to the stiff piano string equation $f_n = n f_0 (1 + B n^2)^{1/2}$ described in [21]. This resulted in the higher partials being more heavily displaced compared to the lower ones, with B reaching a maximum value of 3×10^{-4} . The second introduced a random (positive or negative) displacement for each partial as a percentage of its harmonic frequency (with a maximum of 4%). Both B (in the first case) and the displacement of each partial (in the second case) linearly increased towards the specified maximum level and then decreased back to zero within a total duration of 30 s. This inverted U temporal profile (shown in Fig. 4) was favoured for inharmonicity alteration as it allowed to get responses on both rising and falling inharmonicity. The spectrograms of the the stimuli which were overall 8 (2 pitches \times 2 spectral shapes \times 2 types of inharmonicity variation) are shown in Figure 1.

At this point we would like to introduce the abbreviations of the stimuli which will take the following form throughout the manuscript: F_0 -spectral shape type-inharmonicity type. The possible conditions accordingly can vary between 220 and 440 Hz, sawtooth (saw) or square wave (sqr) and stiff string (str) or random (rnd).

2.1.3 Procedure and apparatus

The task of the listeners was to provide real-time continuous assessment of the stimuli on four different qualities, namely: brightness, roughness, mass and tension. The choice of the timbral semantics was made based on the luminance-texture-mass framework. Brightness and luminance are both translated as $\phi\omega\tau\epsilon\iota\nu\acute{o}\tau\eta\tau\alpha$ in Greek. Roughness ($\tau\rho\alpha\chi\acute{\upsilon}\tau\eta\tau\alpha$) was picked as a prominent representative of the texture semantic dimension that has a meaningful application for sound description as a noun in Greek. For example, adjectives such as harsh, soft or round may be frequently used to characterise sounds but their conversion to noun would result in a rather atypical term for the same purpose in the Greek language. Similarly, mass ($\mu\acute{\alpha}\zeta\alpha$) was maintained as a term since prominent representative nouns falling into this semantic category such as fullness, richness or thickness would also be uncommon for sound description in Greek. Tension was defined similarly to Farbood and Price [12] as the opposite of the feeling of relaxation.

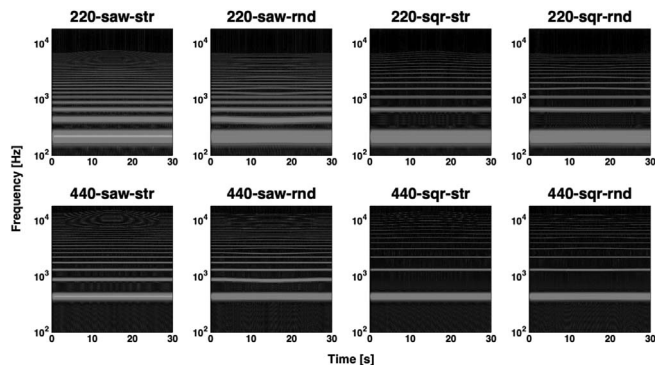


Figure 1. Spectrograms of the 8 stimuli. Stimuli abbreviations: F_0 (220 or 440 Hz) – spectral shape type (sawtooth or square wave) – inharmonicity type (stiff string or random).





In addition, extra care was taken to uncouple the concept from sound intensity since in Greek, the word *ένταση* is commonly used for both. The brightness, roughness and mass qualities were not particularly defined and a brief elaboration of these concepts was only provided to participants requesting further clarifications (e.g., “roughness refers to how rough as opposed to how smooth a sound is”). Each quality was rated on a separate block of trials (random order of quality-specific blocks). Stimuli within each of the four blocks were also presented in random order. The input device used for acquiring the real-time assessment of the qualities was a Kensington Expert wireless trackball mouse. Participants were directed to use movement to the right to indicate increase of the quality and movement to the left to indicate decrease of the quality. Movement upwards or downwards did not provide any information and was not recorded. The presentation of the stimuli was made through the use of a MacBook Pro computer and a pair of PreSonus HD7 circumaural headphones. The initial static conditions (first 3-s) were loudness equalised in an informal listening test within the research team using as a reference the static part of stimulus 1 (220-saw). The RMS playback level of the reference was 70 dB SPL (A-weighted) measured through a dB meter calibration with the use of a dummy head BK 4128 HATS. This process resulted in RMS playback levels for the 220-sqr, 440-saw and 440-sqr stimuli at 71.3, 68.4 and 69 dBA SPL respectively. The inharmonicity alteration did not induce measurable changes in playback level for any of the stimuli. The continuous ratings were acquired by a custom designed LabVIEW GUI that sampled the horizontal axis coordinate of the trackball mouse every 5 ms and offered participants a real-time visualisation of the quality-profiles they were creating. The overall process lasted around 30 min for most of the participants. Ethical approval had been granted for this experiment by the research committee of the Aristotle University.

2.2 Complementary listening test

The set up of the main listening test was adequate for providing information regarding the temporal evolution of the quality-profiles. However, it could not offer any perspective

with respect to the real magnitudes of the assessed qualities as it was assumed that they were all beginning from zero independently of the stimulus. Since the comparative magnitudes between profiles may be proven an informative factor concerning our research questions this complementary experiment intended to provide us with the stimulus dependent offsets for the qualities under study. The listening panel was a subset of the main experiment’s panel consisting of 26 listeners (mean age: 21 years old, standard deviation: 4.14 years, average years of musical practice: 11.5, 16 female, 10 male).

2.2.1 Stimuli

The stimuli were the four initial stimuli of the main listening test: one sawtooth wave (30 harmonics) in two different fundamental frequencies: 220 Hz  and at 440 Hz  and one square wave (15 harmonics) in two different fundamental frequencies: 220 Hz  and 440 Hz . The stimuli were 3-s long as this duration essentially constituted the initial stable part of the stimuli in the main listening test.

2.2.2 Procedure and apparatus

The participants were first presented with the four stimuli in random order for familiarisation purposes and then rated each one of them in four scales: brightness, roughness, mass and tension. The rating was made using a slider on the screen anchored by the full extent of each attribute and its negation (e.g., not bright – very bright) and with a hidden scale ranging from -1 to 1 . Thus, the rating for each stimulus was a single value on each scale. The GUI of this experiment was built in Matlab and the equipment for stimulus playback was the same as in the main listening test. The duration of the test was less than 5 min.

3 Analysis and results

3.1 Static ratings

The agreement among participants for the static ratings of the complementary listening test was very high for timbre description (Cronbach’s alpha for brightness = .96, roughness = .89 and mass = .90) but very poor for tension (Cronbach’s alpha = $-.48$). This indicates that the 3-s duration of a static synthesised timbre was adequate for inducing coherent ratings for timbral qualities (brightness, roughness and mass) but not for felt tension. This finding supports the assumption that tension is a cognitive, higher-level concept that requires some context to become meaningful in contrast to timbral semantics that seem to be related to lower-level perceptual processes.

Figure 2 (top) shows the boxplots of the ratings on the four qualities. A Shapiro-Wilk test for normality was passed by all ratings except for brightness of the 440 Hz square wave (at significance level, $p = .05$) and a Levene’s test within each block of qualities showed that the hypothesis of equal variance was not violated for any of the blocks

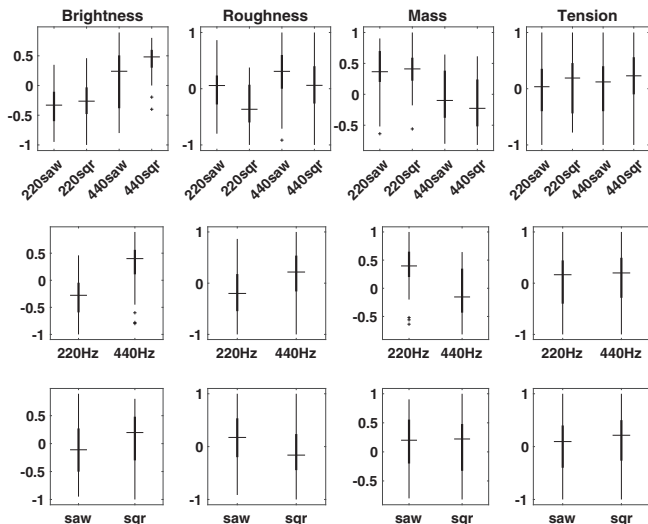


Figure 2. The top panels show the boxplots of the ratings on the four qualities for the four 3-s static stimuli. The medians were used as initial values for the average quality-profiles of the main experiment shown in Figure 4. The middle and the bottom panels show the boxplots as a function of F_0 and spectral shape to better illustrate the effects identified from the two-way repeated measures ANOVA.

($p = .05$). Subsequently, two-way, repeated-measures ANOVAs were conducted to examine the effect of F_0 and spectral shape on each of the qualities. This revealed significant main effects of F_0 on brightness ($F(1, 25) = 44.0$, $p < .001$, $\eta^2 = .64$), roughness ($F(1, 25) = 9.7$, $p = .005$, $\eta^2 = .28$) and mass ($F(1, 25) = 23.5$, $p < .001$, $\eta^2 = .49$) but not on tension ($F(1, 25) = 1.14$, $p = .3$). As shown in Figure 2 (middle) the relation between brightness and roughness with F_0 is proportional but the relation is inversely proportional between mass and F_0 . Significant main effects of the spectral shape were also evident on brightness ($F(1, 25) = 6.17$, $p = .02$, $\eta^2 = .20$) and roughness ($F(1, 25) = 11.18$, $p = .003$, $\eta^2 = .31$). Figure 2 (bottom) shows that the spectrally richer sawtooth waves induced lower brightness and higher roughness ratings in comparison to the spectrally poorer square waves. No significant interaction effects were identified between the fundamental frequency and spectral shape in any of the qualities. Tension was notably the only quality whose mean ratings did not feature any significant differentiation between the four initial conditions in accordance to the previously reported poor inter-rater agreement.

3.2 Continuous ratings

The raw continuous responses of the main listening test were sub-sampled from the original 20 Hz down to 2 Hz by calculating the mean values of adjacent non-overlapping rectangular time windows (.5-s = 10 original samples). These mean values corresponded to the new samples resulting in 60-sample time series. Next, each of the participants' profiles were normalised within each of the four qualities by the maximum absolute value of his/her ratings on this particular quality. This way the initially unbounded rating

scale was transformed to profiles ranging between -1 and 1 by assuming that the maximum (or minimum) value of the scale was reached at least once by each participant for each quality.

3.2.1 Assessment of coherence

In order to assess the coherence of the collection of our responses we generated a set of one thousand collections for each of the eight stimuli, each one containing independent random permutations of the 56 original time series. The Cronbach's alpha corresponding to each of the four qualities was calculated by taking the average over the eight stimuli. This resulted in one thousand Cronbach's alpha values for each of the four qualities. Table 1 shows the Cronbach's alpha of the original data averaged over each quality in comparison to the 50th and 95th (in parenthesis) percentiles of the average Cronbach's alpha values as calculated from the randomly permuted distributions. All Cronbach's alpha values of the original data well exceeded the 95th percentile of these distributions and were, therefore, deemed statistically significant. The mean Cronbach's alpha indicated that roughness and tension were more consistently rated compared to brightness and mass.

3.2.2 Creating categorical variables from time evolving profiles

The continuous responses acquired from this experiment were too short (30-s long) to be analysed with the activity analysis method proposed by [22]. At the same time, they exhibited non-stationarities due to the inherent nature of assessing a stimulus featuring a certain pattern of time evolution. One possibility for analysing such data would be to apply typical time series analysis techniques (e.g., ARIMA or ARIMAX models) that have been previously used for continuous responses of musical stimuli [23]. However, we opted for a more intuitive approach by treating each individual time-varying response as a profile characterised by a certain pattern with a certain strength.

Two categorical variables were constructed through cluster analysis of the profiles. The first variable concerned the type of profile pattern and the second the magnitude of the profile. The categories of patterns resulted from the three following analysis steps:

1. The maximum pairwise Pearson's r coefficients were calculated for all acquired profiles (8 stimuli \times 4 qualities \times 56 participants) at the optimal time lag as a metric of pattern similarity. Although the optimal time lags were eventually less than two seconds in most cases, this step was taken in order to account for possible time lags in the responses between participants.
2. A distance matrix with distances represented by $1 - r$ was subjected to k -medoids cluster analysis. The requested number of clusters was four based on silhouette values and interpretability of the resulting average profiles that are presented in Figure 3.

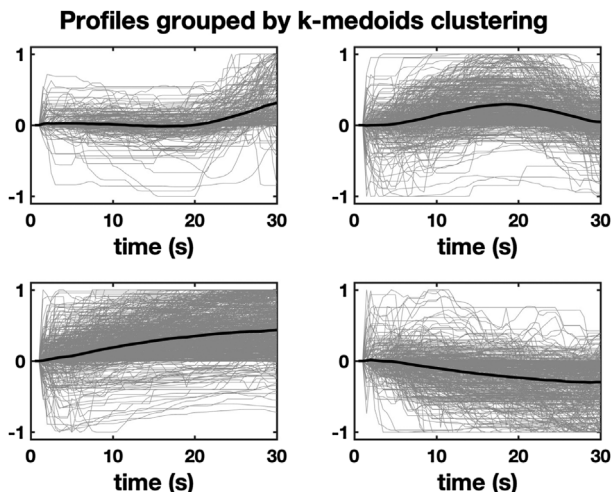


Figure 3. The thick black line represents the mean of the profiles within each cluster. The 1st cluster grouped profiles that increased in the end, the 2nd cluster featured inverted U profiles, the 3rd and 4th clusters grouped increasing and decreasing profiles respectively.

Table 1. Comparison between the mean Cronbach’s alpha values for each quality of the original data and the 50th and 95th percentile (in parenthesis) Cronbach’s alpha values based on 1000 collections of randomly permuted data for each of the 56 time series. All Cronbach’s alpha values were deemed statistically significant since they exceeded the 95th percentile of the 1000 unrelated collections.

	Cronbach’s alpha			
	Brightness	Roughness	Mass	Tension
Original data	.65	.85	.69	.89
Random permutation	-.03 (.08)	-.03 (.08)	-.03 (.08)	-.03 (.07)

3. The types of profile pattern were coded into a nominal variable that represented the index of the cluster where it belonged.

The k -medoids clustering of the magnitudes was based on a dissimilarity matrix with Euclidean distances between the maximum absolute value of each profile. Three clusters were requested from the k -medoids clustering based on silhouette values and the distribution of magnitude values that resulted in the following ranges $0 < x \leq .33$, $.33 < x \leq .75$ and $.75 < x \leq 1$. Each magnitude was coded into an ordinal variable category corresponding to its relevant cluster.

3.2.3 Influences of F_0 , spectral shape and type of inharmonicity

The main listening experiment followed a repeated measures design whereby the same participants rated the same four variables multiple times under different acoustical conditions. Thus, eight generalised estimating equations

(GEEs) models for repeated measures with multinomial (for patterns) and ordinal (for magnitudes) distributions were employed using each of the quality-related categorical variables that were described above as dependent variables and the acoustical conditions as predictors. Table 2 presents the model effects for these models. F_0 was the most influential predictor overall. It featured a main effect on the profile patterns of brightness, mass and tension and on the profile magnitude of mass. It was also involved in interaction effects on the patterns of roughness, mass and tension and on the magnitude of roughness. The type of inharmonicity exhibited also a main effect on mass magnitude and its interaction with F_0 contributed to prediction of the mass and tension patterns. Finally, spectral shape influenced the magnitudes of roughness and tension directly and interacted with F_0 and inharmonicity for predicting the patterns of roughness and brightness, respectively.

Generalised estimating equation models for multinomial distributions in essence apply multiple binary logistic models for each pair between the categories of the dependent variable. This results in a different beta coefficient for each comparison. In our case, where dependent variables consist of either three or four categories with no clear reference (i.e., a category to which compare all the rest) a presentation of all significant beta coefficients for each model would not be easy to decipher. Instead we employ Figure 4 to facilitate a possible interpretation concerning the identified influences of acoustical conditions on quality profiles (presented in Table 2) rather than as a strict quantification. Figure 4 shows the average profiles for each quality (i.e., patterns and magnitudes) and each of the 8 stimuli (i.e., the different experimental conditions) along with the inharmonicity variation pattern. The median values of the static ratings presented in Figure 2 were used as the initial values of the corresponding profiles.

The most characteristic feature of Figure 4 was the prominence of the average inverted U pattern in the 440-Hz condition for the majority of cases. At the same time, decreasing average patterns appeared only in 220-Hz condition (for brightness) and also increasing profiles were more prominent in 220-Hz condition overall. These observations justify the omnipresence of F_0 as a profile pattern predictor for all qualities that was identified through the GEE models. The average patterns of tension seemed to be also influenced by the type of inharmonicity as suggested by Table 2. Indeed, the average patterns of 440-str resembled an inverted U more closely in contrast to the increasing patterns of 440-rnd.

In accordance to the GEE modelling, the average magnitudes of roughness seemed to be influenced by an interaction between spectral shape and F_0 . More specifically, the 220-saw combinations seemed to feature a higher average magnitude compared to the 220-sqr conditions for the corresponding inharmonicity types. However, the opposite was true for the 440-Hz condition. Finally, the magnitudes for mass seemed to be greater for the 220-Hz condition. At this point, it has to be noted that the mere presentation of the average profiles is not possible to clearly reflect all the identified effects of the acoustical conditions on the quan-

Table 2. Model effects for the generalised estimating equations models for repeated measures and multinomial distribution with the four qualities as dependent categorical variables and the acoustical conditions as predictors. Two different models were created for each quality incorporating categories based on the profile pattern and the maximum magnitude respectively.

Criterion	Depend. variable	Predictor	Wald Chi-Square	df	Sig.
Pattern	Brightness	F_0	23.6	1	.000
		Spect. shape \times Inh. type	6.0	1	.014
	Roughness	$F_0 \times$ Spect. shape	4.8	1	.028
	Mass	F_0	4.4	1	.036
		$F_0 \times$ Inharm. type	4.0	1	.045
	Tension	F_0	24.9	1	.000
$F_0 \times$ Inharm. type		8.1	1	.004	
Inharm. type		5.3	1	.022	
Magnitude	Brightness	–			
	Roughness	Spect. shape	4.2	1	.040
		$F_0 \times$ Spect. shape	4.1	1	.042
	Mass	F_0	6.7	1	.010
		Inharm. type	4.7	1	.030
	Tension	Spectral shape	7.3	1	.007

tised quality profiles, especially when these effects are weak. The purpose of this presentation was to highlight the nature of the most notable identified effects.

3.2.4 Influence of timbral semantics on felt tension

Two stepwise multinomial logistic regression models were employed with the tension profile pattern and magnitude categories as dependent variables and the corresponding categories of the three timbral semantics as predictors. Table 3 presents the model fitting criteria and a goodness-of-fit metric for these models. The pattern of the tension profile was predicted by the patterns of all three timbral qualities

with mass being the most informative based on the Akaike information criterion and the $-2 \log$ likelihood, followed by brightness and with a weak contribution from roughness. The magnitude of tension, on the other hand, was predicted primarily by the magnitude of roughness and secondarily by the magnitude of mass.

Figure 5 is used to shed some light on the specific nature of the above general relationships by presenting the frequencies of occurrence for tension pattern and magnitude categories with respect to the categories of the three timbral semantics (with prerequisite of being significant predictors in the multinomial regression models). It can be noticed that it was more likely to observe an increasing pattern for tension when there was also an increasing pattern in either mass, brightness or roughness. This was most prominent for mass and roughness since it was also quite likely for a rising tension pattern to coincide with a decreasing brightness pattern. The probabilities of observing an inverted U pattern for tension were also higher when having an inverted U primarily in roughness and secondarily in mass. A decreasing tension pattern coincided heavily with a decreasing brightness profile although a decreasing brightness also coincided strongly with increasing and inverted U tension patterns.

The tension magnitudes were more likely to fall in the low category [0 – .33] when magnitudes primarily of roughness and secondarily of mass fell in the same category. For the rest categories, differences in frequency of occurrence were marginal.

3.2.5 Overall perspective

The subsections above have separately presented the influence of acoustic parameters on brightness, roughness, mass and tension for the static and continuous ratings, as

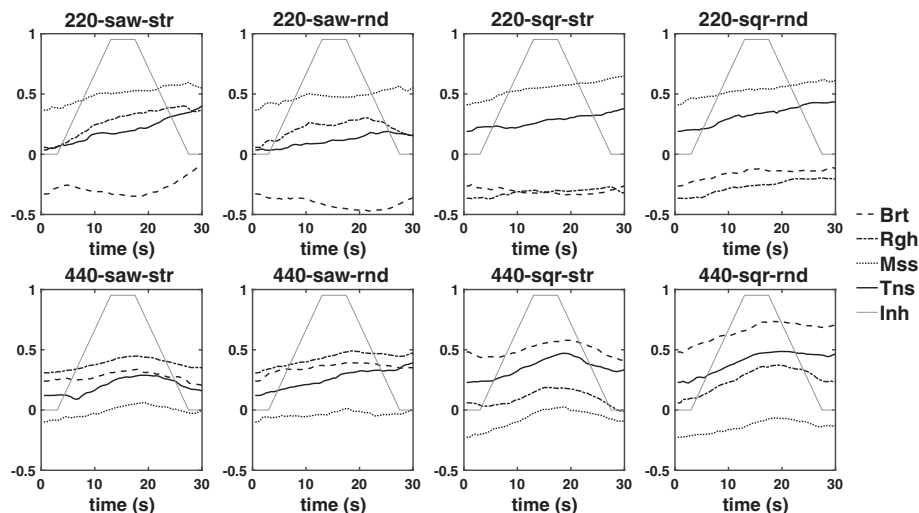


Figure 4. Mean values of the brightness, roughness, mass and tension profiles for each of the stimuli. 220-saw-str: 🎧, 220-saw-rnd: 🎧, 220-sqr-str: 🎧, 220-sqr-rnd: 🎧, 440-saw-str: 🎧, 440-saw-rnd: 🎧, 440-sqr-str: 🎧, 440-sqr-rnd: 🎧 (please use headphones for listening to the accompanying audio files). The grey line shows the function of inharmonicity variation over time: zero 0–3 s, linear increase 3–13 s, one 13–17 s, linear decrease 17–27 s. The initial offset for each quality and each stimulus is the corresponding median value from the static ratings (see Fig. 2).

Table 3. Model fitting criteria, effect selection tests and one goodness-of-fit metric for regression models with tension as dependent variable and the three timbral semantics as predictors. Two different models were created based on the profile pattern (multinomial logistic regression) and the maximum magnitude (ordinal regression). The Akaike information criterion (AIC) measures the model quality by balancing between goodness of fit and parsimony. The lower the values the better the quality. $-2 \log$ likelihood is a measure of the unexplained variance by the model therefore lower values are also desirable. A Chi-square test assesses the significance of each additional predictor by measuring whether the difference in $-2 \log$ likelihood resulting from its inclusion is significant. Finally, McFadden Pseudo R^2 is a metric of the model’s goodness-of-fit. As there is not a straightforward interpretation of this metric, its values should not be judged by the standards for a good fit in a regression analysis [24]. Although in general a larger value means that more variance is explained by the model, such metrics for logistic regression are most useful for comparing competing models for the same data.

Criterion	Depend. variable	Predictor	AIC	$-2 \log$ likelihood	Chi-Square	df	Sig.	McFadden Pseudo R^2
Pattern	Tension	Intercept	528.8	22.8				
		Mass	474.6	450.6	72.2	9	.000	.121
		Brightness	448.3	406.3	44.3	9	.000	
		Roughness	442.1	382.1	24.1	9	.004	
Intercept	241.2	237.2						
Magnitude	Tension	Roughness	229.6	217.6	19.5	4	.001	.036
		Mass	223.2	203.2	14.5	4	.006	

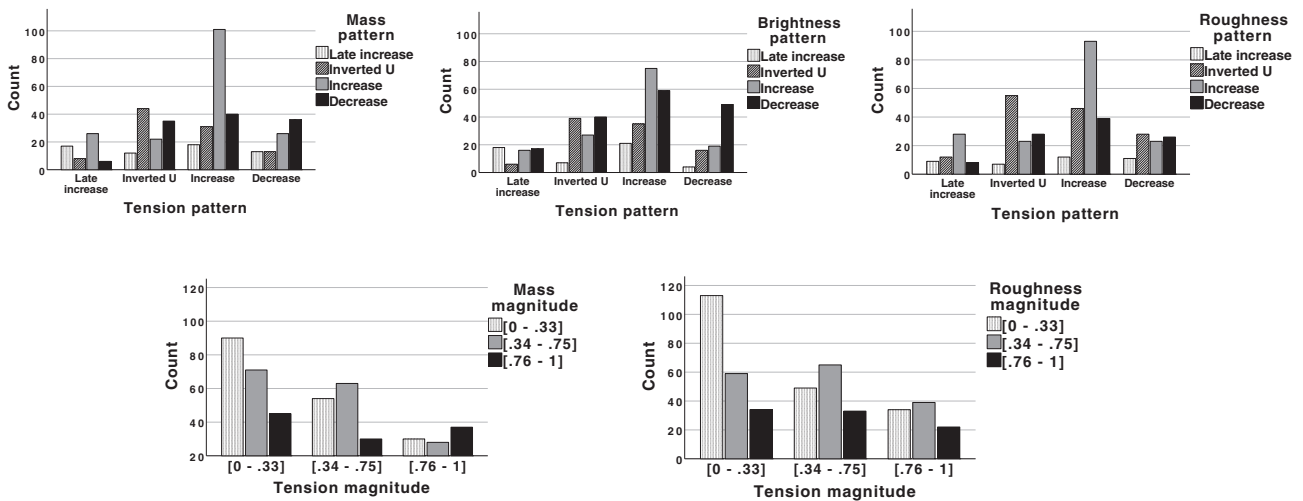


Figure 5. Frequencies of occurrences for tension patterns (top) and tension magnitudes (bottom) within the corresponding categories of the significant predictors.

well as the prediction of dynamically varying tension from timbral semantics. Figure 6 summarises the findings for the dynamically varying stimuli presented in Tables 2 and 3 highlighting main effects with black and interactions with dashed grey arrows. F_0 was the most influential parameter overall, with a direct influence on brightness, mass and tension patterns as well as mass magnitude, and four significant interactions with the other two acoustical parameters. The type of inharmonicity had significant main effects on tension pattern and mass magnitude and showed three significant interactions. Spectral shape also featured main effects on roughness and tension magnitudes and three significant interactions. Figure 6 also shows that mass was identified as the strongest predictor for tension (thicker line), albeit brightness and roughness also played a role.

Overall, this analysis has examined three distinct characteristics of the quality profiles, namely their baseline (i.e., initial values), shape (i.e., pattern) and strength of

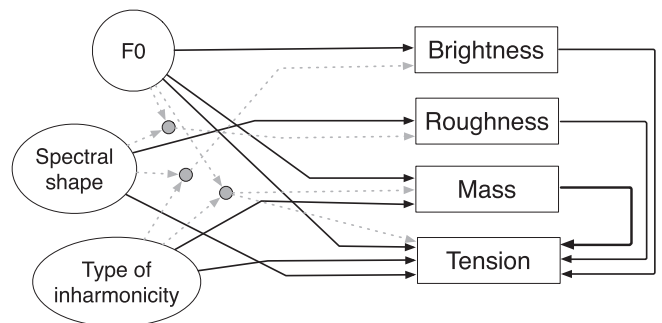


Figure 6. The web of relationships between the four qualities and the acoustical conditions of this experiment in the general context of continuous inharmonicity alteration. Black lines indicate a main effect and dashed grey lines indicate an interaction effect. F_0 was the most influential acoustic parameter for predicting the variation of the qualities under study followed by inharmonicity type and spectral shape. Finally, as shown on the right, mass was the best predictor of tension but brightness and roughness also had a significant effect.

Table 4. Summary of all significant main effects and interactions from acoustic parameters on predicting the ratings of the four examined qualities. * $p < .05$, ** $p < .001$, *ns*, non significant.

Quality	Profile character	F_0	Spectral shape	Inharm. type	$F_0 \times$ Spect. Shape	$F_0 \times$ Inharm. type	Spect. shape \times Inharm. type
Brightness	Pattern	**	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	*
	Magnitude	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>
	Baseline	**	*	–	<i>ns</i>	–	–
	Overall	**	*	<i>ns</i>	<i>ns</i>	<i>ns</i>	*
Roughness	Pattern	<i>ns</i>	<i>ns</i>	<i>ns</i>	*	<i>ns</i>	<i>ns</i>
	Magnitude	<i>ns</i>	*	<i>ns</i>	*	<i>ns</i>	<i>ns</i>
	Baseline	**	**	–	<i>ns</i>	–	–
	Overall	**	**	<i>ns</i>	*	<i>ns</i>	<i>ns</i>
Mass	Pattern	*	<i>ns</i>	<i>ns</i>	<i>ns</i>	*	<i>ns</i>
	Magnitude	*	<i>ns</i>	*	<i>ns</i>	<i>ns</i>	<i>ns</i>
	Baseline	**	<i>ns</i>	–	<i>ns</i>	–	–
	Overall	**	<i>ns</i>	*	<i>ns</i>	*	<i>ns</i>
Tension	Pattern	**	<i>ns</i>	*	<i>ns</i>	*	<i>ns</i>
	Magnitude	<i>ns</i>	*	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>
	Baseline	<i>ns</i>	<i>ns</i>	–	<i>ns</i>	–	–
	Overall	**	*	*	<i>ns</i>	*	<i>ns</i>

shape (i.e., magnitude). Table 4 summarises the effect of the significance of each acoustic parameter for each of these three profile characteristics. The data support that knowledge of F_0 is necessary to adequately predict the value of all the qualities under study at a given moment in time either through a main effect or through an interaction. The type of inharmonicity was here not useful for predicting roughness and spectral shape was not to significant for predicting mass in any form.

The average profiles shown in Figure 4 can be also used to highlight the importance of all three characteristics, namely baseline, pattern and magnitude. Note that the profiles do not initiate from the same point. Apart from the overall shape and magnitude of the profiles, these initial values – directly related to the initial acoustic properties of each stimulus – are useful to obtain a general perspective comparing the profiles of a given quality across different stimuli or comparing profiles of different qualities on the same stimulus.

4 Discussion

This paper investigated the complex relationships between timbral semantics and tension on the one hand, and time varying inharmonicity, spectral shape and F_0 with the aforementioned qualities on the other hand. The major findings are summarised below.

1. Through static ratings of brightness, roughness, mass and tension on 3-s long stimuli it was shown that the four initial conditions of our experiment (i.e., different F_0 s and spectral shapes) were capable of inducing consistently different judgments of timbral semantics but not of felt tension. This is despite the sawtooth wave stimuli being harmonically richer, thus possessing more energy in the higher frequencies compared to

the square wave stimuli, a fact that according to ([1] in Chapter 15, Creating Tension) is associated with anger or fear. This finding corroborates the notion that judgements on timbral attributes are sensory and do not require much context to form, in comparison to judgements on tension which are cognitive and seem to need some context before becoming meaningful.

2. A longer (30-s) continuous inharmonicity alteration of synthetic complex tones was capable of inducing consistent changes in the perception of both timbral qualities and felt tension thus confirming the findings of [12] for dynamically changing sounds. In this case, the ratings on tension and roughness gave the most coherent collections, but ratings on brightness and mass were also consistent well above chance.
3. The most consistent acoustic influence on the ratings (both static and continuous) of all four qualities was the fundamental frequency. F_0 had a statistically significant main effect on the profile patterns of brightness, mass and tension in addition to the magnitude of tension and also on the baseline values of brightness, roughness and mass. The type of inharmonicity used for the time-alteration of the stimuli had a significant main effect on mass magnitude and tension pattern and contributed to the prediction of brightness, mass and tension patterns through interactions (Table 2). This observation implies that mass and brightness might be more sensitive to the type of inharmonicity alteration than roughness. Finally, the spectral shape featured a weak main effect on the baselines of brightness ($\eta^2 = .20$) and roughness ($\eta^2 = .31$) and showed direct influence in predicting the magnitudes of roughness and tension and the patterns of brightness and roughness through interactions.
4. Mass was the most informative predictor of tension profile pattern followed by brightness. Roughness, on the other hand, was the most informative predictor

of tension profile magnitude. The mass-tension relation observed here supports Huron’s ([1] in Ch.15 Creating Tension) argument that “massive” sounds evoke more fear and therefore are usually employed in climactic moments in music. Besides, the roughness-tension association also confirms previous authors [1, 9].

This study also gives rise to a number of interesting questions that warrant further investigation in future work:

Influence of F_0 on tension

The influence of F_0 on tension that was identified in this work was not reported in [12] although three different F_0 s had been tested (110, 220 and 440 Hz). Since the experimental set up of [12] consisted of a pairwise comparison of induced tension it might be the case that the gradual variation of an acoustic parameter and the subsequent continuous response is responsible for this discrepancy. A comparative pairwise assessment of tension between the initial purely harmonic part of our stimuli and the point of maximum inharmonicity could shed more light in a potential interplay between F_0 and the nature of the requested task.

Varying patterns of response

It is also worth noting that all our stimuli featured the same inverted U inharmonicity variation pattern (Fig. 4). However, this was matched by only one out of the four pattern categories of Figure 3. The other categories were better described by an increasing trend, a decreasing trend and even a pattern of strong late increase. What could be the cause of this mismatch? One could attribute the increasing or decreasing trends to a measurement error in the form of a drift during rating. However, participants were provided with real-time visual feedback of the profile they were creating, therefore they should have been able to adjust for such drifts in case their rating was violating their intentions. Besides, this rationale fails to explain why there is not a similar drift present in every profile.

An alternative hypothesis is to attribute this phenomenon to a possible F_0 mediated response bias or some form of differentiated sensitivity between rising and falling inharmonicity (inverted U profiles were more frequent for the 440-Hz condition as implied by Fig. 2). A similar phenomenon has been shown to exist as a bias of distance estimation for an approaching versus a receding tone [25]. The distance of the approaching tones (but not of noise) was underestimated (i.e., intensity raise was overestimated) in contrast to the distance of the receding tones which was estimated more accurately by blindfolded participants. In our case, while participants were generally quite likely to report a rising pattern in response to increasing inharmonicity for all four qualities and for both F_0 conditions, the probability of reporting a falling pattern as a response to decreasing inharmonicity (that would eventually result in an inverted U profile) was higher for the 440-Hz condition. In short, participants tended to report more closely to the

pattern of the inharmonicity variation in the 440-Hz condition. While increased sensitivity of inharmonicity perception for higher frequencies has been reported for static sounds [26] this fact alone cannot fully account for the observed effect.

The most notable discrepancy from the inharmonicity pattern was the late-increase response pattern. It could be argued that this pattern, which despite being the less frequent was relatively equally distributed among qualities, might be reflecting a sense of musical resolution caused by the elimination of inharmonicity and the return to a purely harmonic signal. Note that the increase seems to be echoing even during the last 3-s where the stimuli have returned back to their purely harmonic state resembling a similar lag in continuous responses reported in [27]. Such a response pattern should not be entirely unexpected with respect to tension given its association with the formation of expectations. When it comes to timbral qualities, however, it could imply some sort of directional effect (i.e., different perception of a movement towards inharmonicity compared to movement towards pure harmonicity or even a respective response bias) that has not been previously reported to the best of our knowledge. Further experimentation to investigate a possible F_0 -mediated difference in perception of rising versus falling inharmonicity is mandated based on the evidence presented in this paper. A better understanding of how dynamic variation of inharmonicity is perceived could be useful in domains such as sound synthesis, sound design or data sonification.

Causality between timbre and felt tension

Finally, while it is reasonable to assume that there exists a causal relation between acoustical conditions and the ratings of the qualities under study, causality between the qualities themselves may not be deemed as straightforward. Based on previous evidence [13, 28], this paper has made the assumption that auditory brightness, roughness and mass are discrete concepts, thus there was no effort to predict one from the other. On the contrary, tension was assumed to be a cognitive concept that could potentially be affected by the sensory timbral qualities. Although, the evidence presented here supports this initial hypothesis it is not possible to make a definitive claim on the causality of the identified relationships. It has to be acknowledged that the strict laboratory conditions of this experiment may challenge our assumptions of independence between timbral semantics or even the definition of tension as a cognitive concept. In other words, there is a chance that our observations result from the limited sonic diversity of our stimuli where the alteration of a single acoustical dimension (i.e., inharmonicity) was not capable of inducing unrelated profiles in a number of qualities. However, despite the controlled nature of this experiment, these results pave the way for further investigation on the relation between timbre and tension in more naturalistic scenarios to reach the ultimate goal of incorporating timbral aspects in dynamic models of musically induced tension.

Conflict of interest

Author declared no conflict of interests.

Data availability statement

This article includes audio files embedded in the article. The audio files are also available in Zenodo, under the reference <https://doi.org/10.5281/zenodo.4626964> [29].

Acknowledgments

The authors would like to thank all the participants of this study. They also wish to thank three anonymous reviewers for their very helpful comments on previous versions of this manuscript. 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.

References

1. D.B. Huron: Sweet anticipation: Music and the psychology of expectation. MIT press, London, UK, 2006.
2. S. Koelsch: Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience* 15 (2014) 170–180.
3. E. Bigand, R. Parncutt, F. Lerdahl: Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics* 58 (1996) 125–141.
4. E. Bigand, R. Parncutt: Perceiving musical tension in long chord sequences. *Psychological Research* 62 (1999) 237–254.
5. F. Lerdahl, C.L. Krumhansl: Modeling tonal tension. *Music Perception* 24 (2007) 329–366.
6. M.M. Farbood: A parametric, temporal model of musical tension. *Music Perception* 29 (2012) 387–428.
7. M.M. Farbood, F. Upham: Interpreting expressive performance through listener judgments of musical tension. *Frontiers in Psychology* 4 (2013) 998.
8. S. Paraskeva, S. McAdams: Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. Aristotle University of Thessaloniki, Greece, 1997, pp. 438–441.
9. D. Pressnitzer, S. McAdams, S. Winsberg, J. Fineberg: Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & Psychophysics* 62 (2000) 66–80.
10. J.C. Hailstone, R. Omar, S.M. Henley, C. Frost, M.G. Kenward, J.D. Warren: It’s not what you play, it’s how you play it: Timbre affects perception of emotion in music. *The Quarterly Journal of Experimental Psychology* 62 (2009) 2141–2155.
11. I. Lahdelma, T. Eerola: Single chords convey distinct emotional qualities to both naive and expert listeners. *Psychology of Music* 44 (2016) 37–54.
12. M.M. Farbood, K.C. Price: The contribution of timbre attributes to musical tension. *The Journal of the Acoustical Society of America* 141 (2017) 419–427.
13. A. Zacharakis, K. Pasiadis, J.D. Reiss: An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception* 31 (2014) 339–358.
14. A. Zacharakis, M.J. Terrell, A.J. Simpson, K. Pasiadis, J.D. Reiss: Rearrangement of timbre space due to background noise: Behavioural evidence and acoustic correlates. *Acta Acustica united with Acustica* 103 (2017) 288–298.
15. C.K. Madsen, W.E. Fredrickson: The experience of musical tension: A replication of nielsen’s research using the continuous response digital interface. *Journal of Music Therapy* 30 (1993) 46–63.
16. C.L. Krumhansl: A perceptual analysis of Mozart’s piano sonata k. 282: Segmentation, tension, and musical ideas. *Music Perception* 13 (1996) 401–432.
17. M. Lehne, M. Rohrmeier, D. Gollmann, S. Koelsch: The influence of different structural features on felt musical tension in two piano pieces by Mozart and Mendelssohn. *Music Perception* 31 (2013) 171–185.
18. M. Lehne, M. Rohrmeier, S. Koelsch: Tension-related activity in the orbitofrontal cortex and amygdala: an fMRI study with music. *Social cognitive and affective neuroscience* 9 (2013) 1515–1523.
19. M. Goodchild, B. Gingras, S. McAdams: Analysis, performance, and tension perception of an unmeasured prelude for harpsichord. *Music Perception* 34 (2016) 1–20.
20. B. Gingras, M.T. Pearce, M. Goodchild, R.T. Dean, G. Wiggins, S. McAdams: Linking melodic expectation to expressive performance timing and perceived musical tension. *Journal of Experimental Psychology: Human Perception and Performance* 42 (2016) 594.
21. H. Fletcher: Normal vibration frequencies of a stiff piano string. *The Journal of the Acoustical Society of America* 36 (1964) 203–209.
22. F. Upham, S. McAdams: Activity analysis and coordination in continuous responses to music. *Music Perception* 35 (2018) 253–294.
23. F. Bailes, R.T. Dean: Comparative time series analysis of perceptual responses to electroacoustic music. *Music Perception* 29 (2012) 359–375.
24. D. McFadden: Quantitative methods for analyzing travel behaviour of individuals: some recent developments. University of California Berkeley, CA, Institute of Transportation Studies, 1977.
25. J.G. Neuhoff: An adaptive bias in the perception of looming auditory motion. *Ecological Psychology* 13 (2001) 87–110.
26. B.C.J. Moore, R.W. Peters, B.R. Glasberg: Thresholds for the detection of inharmonicity in complex tones. *The Journal of the Acoustical Society of America* 77 (1985) 1861–1867.
27. E. Schubert: Modeling perceived emotion with continuous musical features. *Music Perception* 21 (2004) 561–585.
28. A. Zacharakis, K. Pasiadis: Revisiting the luminance-texture-mass model for musical timbre semantics: A confirmatory approach and perspectives of extension. *Journal of the Audio Engineering Society* 64 (2016) 636–645.
29. Interaction between time-varying tone inharmonicity, fundamental frequency and spectral shape affects felt tension and timbral semantics. [Online] Available at: <https://doi.org/10.5281/zenodo.4626964> [Accessed: Mar 22 2021].

Cite this article as: Zacharakis A. & Pasiadis K. 2021. Interaction between time-varying tone inharmonicity, fundamental frequency and spectral shape affects felt tension and timbral semantics. *Acta Acustica*, 5, 21.