An Interlanguage Unification of Musical Timbre: Bridging Semantic, Perceptual, and Acoustic Dimensions
Author(s): Asterios Zacharakis, Konstantinos Pastiadis and Joshua D. Reiss
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 32, No. 4 (April 2015), pp. 394-412
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/10.1525/mp.2015.32.4.394
Accessed: 13/05/2015 05:09

AN INTERLANGUAGE UNIFICATION OF MUSICAL TIMBRE: BRIDGING SEMANTIC, PERCEPTUAL, AND ACOUSTIC DIMENSIONS

ASTERIOS ZACHARAKIS
Queen Mary University of London, London, United Kingdom

KONSTANTINOS PASTIADIS
Aristotle University of Thessaloniki, Thessaloniki, Greece

JOSHUA D. REISS
Queen Mary University of London, London, United Kingdom

THE CURRENT STUDY EXPANDS OUR PREVIOUS work on interlanguage musical timbre semantics by examining the relationship between semantics and perception of timbre. Following Zacharakis, Pastiadis, and Reiss (2014), a pairwise dissimilarity listening test involving participants from two separate linguistic groups (Greek and English) was conducted. Subsequent multidimensional scaling analysis produced a 3D perceptual timbre space for each language. The comparison between perceptual spaces suggested that timbre perception is unaffected by native language. Additionally, comparisons between semantic and perceptual spaces revealed substantial similarities which suggest that verbal descriptions can convey a considerable amount of perceptual information. The previously determined semantic labels "auditory texture" and "luminance" featured the highest associations with perceptual dimensions for both languages. "Auditory mass" failed to show any strong correlations. Acoustic analysis identified energy distribution of harmonic partials, spectral detail, temporal/spectrotemporal characteristics and the fundamental frequency as the most salient acoustic correlates of perceptual dimensions.

Received: July 5, 2013, accepted June 21, 2014.

Key words: musical timbre perception, timbre spaces, multidimensional scaling, semantic description, acoustic correlates

RELATIONAL MEASURES OF TIMBRE RESEARCH comprise two popular paradigms: dissimilarity rating and verbal description. The most popular approach for the study of timbre perception has been the application of multidimensional scaling (MDS) techniques to dissimilarity matrices obtained by pairwise dissimilarity ratings of sound stimuli. MDS creates a geometrical configuration of the timbres under study called *timbre space*, which has been used for identification of the salient perceptual dimensions. Plomp (1970) was the first to make use of this approach which has since been adopted by a plethora of researchers (e.g., Caclin, McAdams, Smith, & Winsberg, 2005; Grey, 1977; Iverson & Krumhansl, 1993; McAdams, Winsberg, Donnadieu, Soete, & Krimphoff, 1995).

The verbal description of sound quality and its association with the physical properties of sound has also intrigued researchers for a long time. Helmholtz (1877/1954, p.118-119) has made one of the first systematic attempts to associate semantic attributes with acoustic characteristics, and Lichte (1941) has broken down the timbre of complex tones into three independent semantic components, namely, *brightness*, *roughness,* and *fullness*. Efficient as MDS analysis may be for the identification of timbral perceptual dimensions, it is incapable of applying labels to them. The labeling of the dimensions in such cases often comes as a result of some speculative interpretation. However, applying an accurate semantic label to a perceptual dimension is highly desirable for intuitive human computer interaction applications on sound processing and synthesis.

When the major objective is to investigate verbal description of musical timbre, then methods like the semantic differential (e.g., von Bismarck, 1974; Lichte, 1941) and its variant called verbal attribute magnitude estimation (VAME; e.g., Kendall & Carterette, 1993a, 1993b) are usually employed instead of MDS. Both methods require the rating of perceptual objects along semantic scales. Kendall and Carterette (1993a, 1993b) and Kendall, Carterette, and Hajda (1999) attempted to exploit a combination of pairwise dissimilarity and verbal attribute ratings for isolated and dyad timbres. The timbre spaces that resulted from these two approaches were compared but their similarities were found to be rather limited. Faure, McAdams, and Nosulenko (1996) have also tried to bridge semantics with perception through a pairwise dissimilarity test and additional free verbal description of the perceptual differences and

Music Perception, VOLUME 32, ISSUE 4, PP. 394–412, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312. © 2015 BY THE REGENTS OF THE UNIVERSITY OF CALIFORNIA ALL RIGHTS RESERVED. PLEASE DIRECT ALL REQUESTS FOR PERMISSION TO PHOTOCOPY OR REPRODUCE ARTICLE CONTENT THROUGH THE UNIVERSITY OF CALIFORNIA PRESS'S RIGHTS AND PERMISSIONS WEBSITE, HTTP://WWW.UCPRESSJOURNALS.COM/REPRINTINFO.ASP. DOI: 10.1525/MP.2015.32.4.394

similarities. This study identified 22 semantic descriptors and associated them with perceptual dimensions and acoustic characteristics. The majority of the adjectives correlated to more than one perceptual dimension. Therefore, the value of musical timbre description by verbal means remained an open question.

Other studies have also addressed this issue from a different viewpoint. From a linguistics perspective, Samoylenko, McAdams, and Nosulenko (1996) found that verbal description of perceived timbral dissimilarities corresponded well with numerical dissimilarity ratings. Therefore, a relationship between timbre description and timbre dissimilarity was suggested, but as stated by the authors, a remaining question was whether this relationship held up at the level of timbre space dimensions. The subsequent work by Kendall et al. (1999) found only weak support for the relationships requested by Samoylenko et al. (1996).

Furthermore, timbre semantics have recently been investigated through a neuroscientific approach that offered new insight into the question of meaning conveyed by timbre. Painter and Koelsch (2011) carried out two EEG experiments that demonstrated the ability of musical timbre to carry extramusical meaning. More specifically, it has been demonstrated that prior listening to a sound can significantly influence the meaningful processing of a subsequent word or sound. Alluri and Toiviainen (2010) have also identified three salient perceptual dimensions for polyphonic timbre, namely *activity*, *brightness,* and *fullness*. In a subsequent study, Alluri et al. (2012) investigated the neural underpinnings of timbral and other features of a naturalistic musical stimulus. The acoustic parameters representing the basic perceptual timbre dimensions were identified and functional magnetic resonance imaging (fMRI) was utilized to localize parts of the brain that were responsible for processing each of these separate dimensions.

The above suggest that semantic description of musical timbre can provide significant information regarding perceptual representation of sound. However, this has not been adequately validated through comparison of pairwise dissimilarity rating and verbal description studies. In a previous interlanguage study between English and Greek speaking participants, we demonstrated robustness of musical timbre semantics for different languages (Zacharakis et al., 2014). This work focused merely on iconic musical meaning (see Koelsch, 2011); that is, timbral descriptions associating sounds with qualities of objects or qualities of abstract concepts. The participants were asked to rate the timbre of 23 musical tones choosing from a pool of 30 provided adjectives according to the VAME methodology. These 30 semantic variables were analyzed through factor analysis, which identified three salient semantic dimensions for each language. As is usually the case with the interpretation of such factors, the labeling of these dimensions resulted from the effort to integrate the number of specific adjectives that were highly loaded on each factor into one single concept. The respective dimensions of the two different languages were characterized by common conceptual properties and similar sound positioning. Therefore the suggested labels were *luminance*, *texture,* and *mass* for both English and Greek. *Luminance* was adopted to unify concepts such as brilliance and depth, *texture* encapsulated the terms rough, harsh, rounded, warm, soft, and messy, and *mass* was used to represent concepts such as fullness, richness, lightness, and thickness.

The motivation behind the present study was to extend the work of Samoylenko et al. (1996) by investigating the relationship between semantics of musical timbre and its underlying perceptual dimensions. Since, our previous work investigated timbral semantics between two different linguistic groups, one additional point of interest would be to examine whether pairwise dissimilarity ratings of timbre are affected by the first language of the participants. To this end, similarly to our previous experimental set up, a pairwise dissimilarity listening test using the same sound stimulus set as in Zacharakis et al. (2014) was conducted to participants belonging to two separate linguistic populations, Greek and English. The data were subsequently analyzed through MDS analysis, resulting in two perceptual timbre spaces. This enabled the comparison of four timbre spaces, i.e., one semantic and one perceptual for each of the two linguistic groups. Unlike other related studies (e.g., Elliott, Hamilton, & Theunissen, 2012; Faure et al., 1996) the participants in our work were different for each of the four separate listening tests.

The major objectives of this work are:

1) to examine the influence of native language on timbre perception. A potentially strong relationship between the two perceptual spaces will indicate a minimal effect of language on dissimilarity ratings of timbre and vice versa.
2) to assess the value of musical timbre description (i.e., the extent to which semantics can account for what is actually being perceived as measured by dissimilarity ratings) by comparing the perceptual with the semantic spaces.

Method

The perceptual timbre spaces resulted from MDS analyses that were separately applied to the pairwise dissimilarity ratings of the two linguistic groups (33 Greek and 20 English speakers). The semantic spaces were the outcome of separate factor analyses applied to the data of the verbal attribute magnitude estimation (VAME) listening tests undertaken by 41 English and 41 Greek speakers (presented in Zacharakis et al., 2014).

In the pairwise dissimilarity listening test, participants were asked to compare all the pairs among the 24 sound stimuli using the free magnitude estimation method (Bensmaïa & Hollins, 2005; Cho, Kim, & Casali, 2002; DeCarlo & Cross, 1990; Green et al., 1996; Marks, 1980; Stevens, 1971; Yoshioka et al., 2007). Therefore, they rated the perceptual distances of 300 pairs (same pairs included) by freely typing in a number of their choice to represent dissimilarity of each pair (i.e., an unbounded scale) with 0 indicating a same pair. The free magnitude estimation method was favored over bounded magnitude estimation as the latter introduces the following two issues during a rating procedure. Participants, not being in a position to anticipate upcoming dissimilarities, may never utilize the available range of the scale in case an even larger dissimilarity shows up later in the test. On the other hand, they may prematurely select the scale's maximum when their maximum rating should normally be appointed to an upcoming pair, thus clipping their intended response. Details on considerations for MDS data preparation are presented in subsequent paragraphs.

STIMULI AND APPARATUS

The VAME listening test consisted of 23 stimuli while one additional cello tone was included in the pairwise dissimilarity test. The 24 sounds within the initial set varied in fundamental frequency (range of three octaves). The following 15 instrument tones came from the McGill University Master Samples (MUMS) library (Opolko & Wapnick, 2006): *violin, sitar, trumpet, clarinet, piano, cello* each at A3 (220 Hz), *Les Paul Gibson guitar, baritone saxophone B flat* each at A2 (110 Hz), *double bass pizzicato* at A1 (55 Hz), *oboe* at A4 (440 Hz), *Gibson guitar, pipe organ, marimba, harpsichord* each at G3 (196 Hz), and *French horn* at A#3 (233 Hz). A *flute* recording at A4 was also used along with a set of 8 synthesizer and electromechanical instrument sounds: *Acid, Hammond, Moog, Rhodes piano* each at A2, *electric piano (rhodes), Wurlitzer, Farfisa* each at A3, and *Bowedpad* at A4.

In contrast to the VAME test where the sounds varied in both duration (from 3 to 8 s) and pitch, these two variables needed to be equalized as much as possible for the pairwise dissimilarity test. To this end, only the first 1.3 s of each sound were retained with an exponential fade out applied to the last 113 ms (i.e., 5000 samples). Furthermore, the five sound samples at G3 and A#3 were all pitch shifted to A3 so that the whole sound set consisted of merely chroma class 'A' (ranging from A1 to A4). This mild modification has not affected the timbral quality of the sounds. This was highlighted by an extra pairwise dissimilarity listening test that was performed just on these five stimuli (marimba, harpsichord, pipe organ, French horn and Gibson guitar) and their pitch shifted versions. The MDS analysis presented in Appendix A showed that a 2D space is adequate to model the relationships within this stimulus set and that the original and pitch shifted versions of each stimulus occupied the same positions. Finally, Krumhansl and Iverson (1992) have stated that even though pitch and timbre are not perceived independently this does not imply that a comparison of timbres with different pitches is impossible. Marozeau, de Cheveigné, McAdams, and Winsberg (2003) and Marozeau and de Cheveigné (2007) have also shown that listeners were able to ignore pitch differences and focus merely on timbre for a range of up to at least 1.5 octave.

The sound samples were loudness equalized in an informal listening test within the research team. One sound from the stimulus set was initially picked up as a reference and was set at a convenient listening level. Then the rest of the stimuli were equalized in loudness according to this reference by the first author. The equalized set was in turn evaluated by the rest of the authors. The resulting RMS playback level was finally measured and found to be between 65 and 75 dB SPL (A-weighted, slow response) for all stimuli. All the participants found this level comfortable for all stimuli and reported that loudness was perceived as being constant across stimuli in a subsequent questionnaire-based evaluation.

The listening test was conducted under controlled conditions in acoustically isolated listening rooms. Sound stimuli were presented through the use of a laptop computer, with an M-Audio (Fast Track Pro USB) external audio interface, and a pair of Sennheiser HD60 ovation circumaural headphones.

PARTICIPANTS

Thirty-three native Greek speakers (age range = 19-50, mean age = 24, 19 female) and 20 native English speakers (age range = 21-40, mean age = 30, 6 female) participated in the listening test. None of the participants reported any hearing loss or absolute pitch and they had been practicing music for 13.2 (Greeks) and 16.1 (English) years on average, ranging from 6 to 25 (Greek) and 8 to 30 (English). The absence of absolute pitch from the group of our participants was a prerequisite as such a condition could affect the results due to pitch variation within the stimulus set. The Greek-speaking participants were students in the Department of Music Studies of the Aristotle University of Thessaloniki and the English-speaking participants were research students from the Centre for Digital Music at Queen Mary University of London.

PROCEDURE

Listeners became familiar with the timbral range of the experiment during an initial presentation of the stimulus set (random order). This was followed by a brief training stage where listeners rated five selected pairs of stimuli. For the main part of the experiment participants were allowed to listen to each pair of sounds as many times as needed prior to submitting their rating. The pairs were presented in random order and listeners were advised to base their ratings merely on timbral differences ignoring differences in pitch and to retain a consistent rating strategy throughout the experiment. Participants were prompted to take one break at the completion of the first third and a second one at the completion of the second third of the overall experiment. They were also offered the option to withdraw at any point. In total, the listening test sessions, including instructions and breaks, lasted around one hour for most of the participants.

NON-METRIC MDS

Multidimensional scaling (MDS) is a series of data analysis techniques that are used to transform distance matrices into N-dimensional spatial configurations of the objects under study. MDS originates from psychometrics and was developed to enable the interpretation of people's pairwise dissimilarity judgments over a set of perceptual objects (Kruskal, 1964a; Shepard, 1962a). It is particularly popular in timbre perception research as the spatial representation of a group of sound objects enables the investigation of the underlying perceptual dimensions.

Figure 1 shows the ratio of maximum-to-mean rating for all the participants in the two groups. The maximum ratio (appearing in the English group) is on the order of 3. This is not deemed an outlying value and therefore no participant was discarded at this point. In order to accurately profile each participant's contribution to the MDS solution we incorporated the weighted individual differences scaling (INDSCAL) algorithm. INDSCAL computes weights that represent the importance attributed to each perceptual dimension by each participant and then uses these weights to reconstruct an "average" perceptual space. However, since there was variation of the mean among participants, we have chosen to analyze the dissimilarities using a non-metric MDS approach (Kruskal, 1964b; Shepard, 1962b) as offered by the SPSS PROXSCAL (proximity scaling) algorithm (Meulman & Heiser, 2008). PROXSCAL applies an ordinal (rank order) transformation to the raw dissimilarities within each participant responses. Additionally, the non-metric MDS approach has been proven robust to the presence of

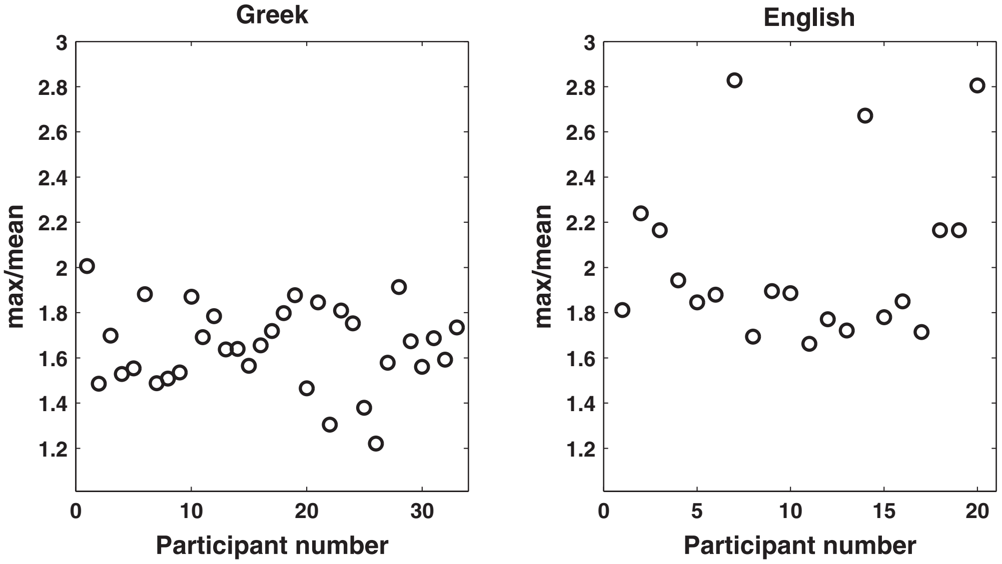

FIGURE 1. Scatter plots of the maximum-to-mean ratios for the pairwise dissimilarity ratings of Greek (left) and English (right) participants.

monotonic transformations or random error in the data (Shepard, 1966; Young, 1970).

Analysis and Results

The main objective of this study was to examine the influence of language on musical timbre and identify potential relations between semantics and perception. The block diagram in Figure 2 presents the schema of our investigation, namely the two perceptual spaces (English and Greek MDS spaces), the two semantic spaces (English and Greek semantic spaces as identified through factor analysis in Zacharakis et al., 2014), and their mutual relationships. These relationships are labeled as X_rX_r using the language initials ($X = G$ for Greek and $X = E$ for English) and a referral according to the nature of each space ($r = p$ for perceptual vs. $r = s$ for semantic). As shown in Figure 2, both heterologous intralanguage (i.e., E_pE_s and G_pG_s) and heterologous interlanguage (i.e., E_pG_s and G_pE_s) relations are examined. This is because only after comparing the two cases can we reach a solid conclusion regarding the effect of language. This section is structured in three subsections. The first examines configurational similarity between the timbre spaces under study while the second investigates relationships at the dimension level. Finally, the third subsection looks at the acoustic correlates of the perceptual dimensions.

Before proceeding to the main body of the analysis we examined the internal consistency of the responses within each linguistic group. Cronbach's alpha was .96 among Greek and .94 among English participants indicating high interparticipant reliability. This was also supported by the very similar weights attributed by each participant to each of the MDS dimensions as shown in Figure 3. As weights signify the importance attributed to each dimension by each subject, their tight clustering across all three facets of dimensions implies that judgments were based on similar criteria among participants. Additionally, the concentration of weights in the center of each diagram indicates that all three perceptual dimensions are of equivalent importance.

In the main body of the analysis, the dissimilarity ratings within each linguistic group were analyzed through non-metric (ordinal) MDS with dimension weighting (INDSCAL within SPSS PROXSCAL algorithm). Table 1 shows two measures-of-fit (S-Stress[1] and D.A.F.[2]) along with their improvement for each added dimension. The optimal dimensionality was deemed to be three as the improvement of the measures-of-fit from a 3D to a 4D space solution was minimal for both groups. All MDS solutions attained stress values lower than the expected values for random data (Spence, 1979). Both measures-of-fit for the non-metric approach were better than those of the metric approach for the same dimensionality.

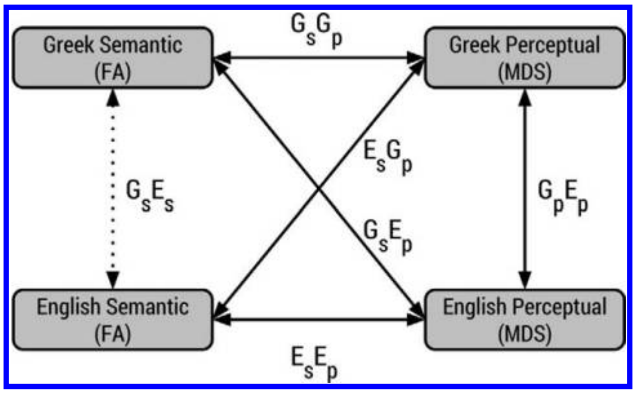

FIGURE 2. Investigation schema consisting of Greek and English semantic and perceptual spaces and their mutual relationships. The labeling is according to language and nature of the space (e.g., G_s stands for Greek semantic, E_p for English perceptual, etc.). The dotted line combining the semantic spaces signifies a relationship that was mainly examined in our previous work. FA and MDS denote semantic space configuration that are driven by Factor Analysis and Multidimensional Scaling, respectively.

CONFIGURATIONAL SIMILARITY BETWEEN TIMBRE SPACES

In this section the relationships between the semantic and/or perceptual spaces were investigated in terms of their configurational similarity, wherein the examined sounds represented the objects of the configurations.[3] The configurational similarity reflects the similarity of the solid shapes defined by the swarms of objects within the spaces. Any global form of similarity between spaces should also take into account the orientation of the swarms relative to the axes and the scales of the spaces.

An orthogonal Procrustes transformation (i.e., only allowing combinations of uniform scaling, translation and rotation) (Borg & Groenen, 2005; Schönemann & Carroll, 1970) was applied within each semantic-to-perceptual comparison.[4] The semantic space configuration was always the one transformed so that it best fitted

---

[1] S-Stress is a measure of misfit. The lower the value (to a minimum of 0) the better the fit.

[2] D.A.F.: Dispersion Accounted For is a measure of fit. The higher the value (to a maximum of 1) the better the fit.

[3] The extra cello tone was removed from the perceptual spaces to enable direct comparison with the semantic spaces.

[4] The oblique coordinates of the semantic spaces were changed into an equivalent orthogonal system.

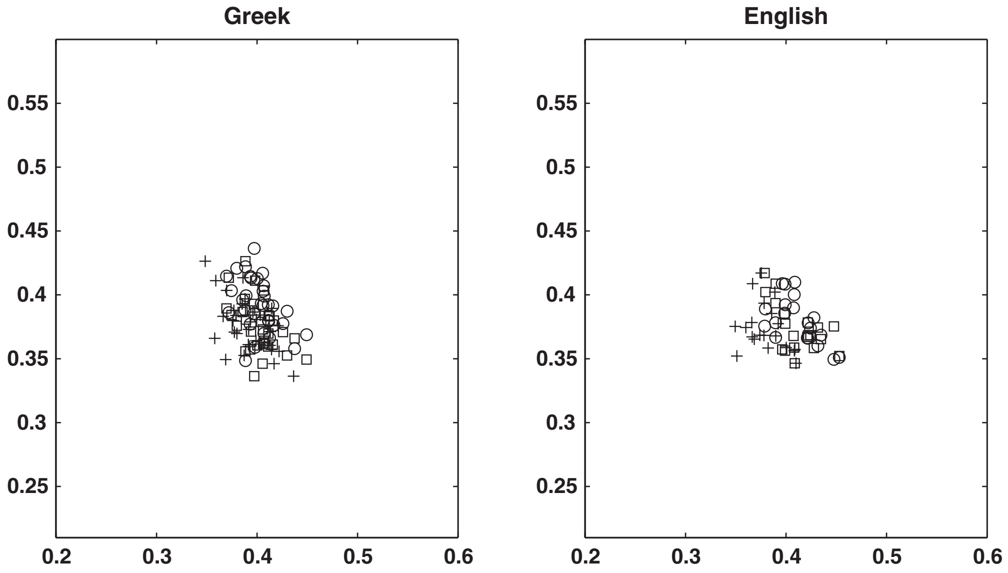

FIGURE 3. Weights attributed to each dimension by the participants of the two linguistic groups as identified by the INDSCAL algorithm. x-axis and y-axis represent different dimensions for each type of marker. ○: weights between 1st (x-axis) and 2nd (y-axis) dimension, □: weights between 1st (x-axis) and 3rd (y-axis) dimension, +: weights between 2nd (x-axis) and 3rd (y-axis) dimension. The tight clustering of all weights around the diagonal for both languages indicates that participants not only based their judgments on similar criteria but they also attributed equal importance to each of the perceptual dimensions regardless of language.

TABLE 1. *Measures-of-fit and Their Improvement for Different MDS Dimensionalities for the Greek and English Groups.*

	Greek				English			
Dimensionality	**S-Stress**	**Improv.**	**D.A.F.**	**Improv.**	**S-Stress**	**Improv.**	**D.A.F.**	**Improv.**
1D	*.34*	–	*.81*	–	*.36*	–	*.81*	–
2D	*.19*	*.15*	*.92*	*.11*	*.19*	*.17*	*.92*	*.11*
3D	*.12*	*.07*	*.95*	*.03*	*.13*	*.06*	*.95*	*.03*
4D	*.10*	*.02*	*.97*	*.02*	*.10*	*.03*	*.97*	*.02*

the perceptual space, which was unaltered in all cases. No transformation was applied to the perceptual spaces even for their direct comparison in order to preserve the perceptually meaningful configuration produced by the INDSCAL algorithm.

Subsequently, configurational similarity between spaces was judged by two indices computed from the distances between the objects; Tucker's congruence coefficient (Borg & Groenen, 2005; Tucker, 1951) and the m^2 statistic for Procrustes analysis (Gower, 1971, 1975; Gower & Dijksterhuis, 2004). The m^2 resembles a measure of alienation $1\text{-}r^2$ (where r is the correlation coefficient between the sequences of within-space distances of the two examined spaces) (Peres-Neto & Jackson, 2001). The exploitation of both indices was mandated by the fact that no single measure of configurational similarity is globally adequate to depict the relationship between two examined spaces (Borg & Groenen, 2005; Borg & Leutner, 1985).

As a guideline, for the congruence coefficient, values larger than .92 are considered good/fair, and values larger than .95 practically show equality between configurations (Lorenzo-Seva & ten Berge, 2006). Significance of the congruence coefficient between the two configurations was tested using a bootstrap analysis method (Monte Carlo estimate of its expected value under chance conditions) (Cutzu & Edelman, 1996; Efron & Tibshirani, 1993). For the m^2 statistic, values $< .75$ (based on recommendations for $r^2 > .25$ as described in Ellis, 2010) signify a large effect size. Statistical significance is tested using an approach that employs a large number of random permutations of the original data and is suited to Procrustes analysis under the name of PROcrustean randomization TEST

(PROTEST; Jackson, 1995; Legendre & Legendre, 1998; Peres-Neto, 2000; Peres-Neto & Jackson, 2001). The statistical significance of r^2 (derived from m^2) has only been investigated in a small number of studies (Andrews & Inglehart, 1978; Borg & Leutner, 1985; Langeheine, 1982), which showed that critical values for r^2 varied with dimensionality of configurations and with number of objects.

Table 2 summarizes the values of the congruence coefficient and m^2 for the relationships of configurations between all examined spaces. The configurations of the English and Greek perceptual (MDS) spaces show a high degree of similarity (congruence coefficient = .98, well above the statistical significance of $p = .05$, and m^2 highly significant). The similarity between the semantic configurations of the two languages (G_sE_s) is lower than the similarity of the perceptual spaces, but nevertheless remains fair (congruence coefficient = .93, $m^2 = 0.41$, both highly significant). In general, the homologous configurations between the two languages show the highest degree of similarity among all examined intralinguistic and interlinguistic relationships. All heterologous configurations also remain fair with minor differences between intralinguistic and interlinguistic relationships.

The strong configurational similarity that was quantified by the similarity metrics is also evident by visual inspection of Figure 4. Detailed commenting on the dimensions of timbre spaces will follow in the next subsection.

RELATIONSHIPS AT THE DIMENSION LEVEL - SEMANTIC INTERPRETATION OF PERCEPTUAL DIMENSIONS

In this section we investigate one-to-one comparisons between semantic and perceptual dimensions. This serves the purpose of semantic interpretation of the unlabeled MDS dimensions. Orthogonal Procrustes transformations may introduce data transformations (e.g., rotation of configurations), which result in altered interpretation of semantic dimensions and are also inadmissible if applied to MDS solutions with dimension weighting. Thus, the investigation of relationships between spaces (and not merely configurations) was based on the original dimensions and unrotated data.

Table 3 presents the Pearson correlation of dimensions between the perceptual (MDS) spaces of the two languages. As the strong one by one correlations suggest, there is not a mere configurational similarity (based on the congruence coefficient and the m^2) but an almost complete coincidence of the two perceptual spaces.

Table 4 presents the Pearson correlation of dimensions between the semantic spaces of the two languages. Such relations have also been reported in detail in a previous work (Zacharakis et al., 2014) where we investigated timbre semantics and their relationships between Greek and English. There appear statistically significant and strong intercorrelations between dimensions of the semantic spaces.[5] The strongest correlation is observed between the dimensions labeled as *texture*. The *luminance* dimensions are also strongly correlated, while the

TABLE 2. *Congruence Coefficients, m^2 and r for the Mutual Relationships Across Timbre Spaces as Described in the Schema of Figure 2.*

Relationship	**Congruence coefficient (expected value, *SD*)***	**m^2**	**r**
G_sG_p	.93 (.85, .02)	.55**	.67
E_sE_p	.92 (.83, .02)	.63**	.60
G_sE_p	.93 (.84, .02)	.61**	.63
E_sG_p	.94 (.83, .02)	.46**	.73
G_pE_p	.98 (.87, .01)	.17**	.91
G_sE_s	.93 (.84, .02)	.41**	.77

**expected chance value, estimated by bootstrap with 10000 runs.*
***p = .001, PROTEST significance testing.*

TABLE 3. *Correlation Matrix (Pearson's r) Between the English and Greek Perceptual Dimensions.*

Perceptual dimensions	**1st English**	**2nd English**	**3rd English**
1st Greek	.89**	−.01	.04
2nd Greek	−.18	.95**	.20
3rd Greek	−.12	−.04	.85**

***Correlation is significant at the .01 level (2-tailed).*
**Correlation is significant at the .05 level (2-tailed).*

TABLE 4. *Correlation Matrix (Pearson's r) Between the English and Greek Semantic Dimensions (from Data Presented in Zacharakis et al., 2014).*

Greek semantic dimensions	**English semantic dimensions**		
	Luminance	**Texture**	**Mass**
Luminance	−.77**	.08	.79**
Texture	−.54**	−.85**	−.07
Mass	−.23	.04	.43*

***Correlation is significant at the .01 level (2-tailed).*
**Correlation is significant at the .05 level (2-tailed).*

---

[5] The reported negative correlations come from the factor analysis solution, which in some cases yielded reversed axes between the two languages without, however, altering their interpretation. For more details please see Table 2 from Zacharakis et al. (2014).

TABLE 5. *Pearson Correlation Coefficients Among Perceptual and Semantic Dimensions For the Two Languages.*

Semantic dimensions		Greek perceptual dimensions			English perceptual dimensions		
		1st	2nd	3rd	1st	2nd	3rd
Greek	Luminance	.22	−.10	.74**	.09	−.13	.58**
	Texture	−.59**	.54*	.51*	−.73**	.42*	.45**
	Mass	−4e-3	−.33	.15	−.09	−.33	−.17
English	Luminance	−.02	−.11	−.83**	.23	−.03	−.60**
	Texture	.67**	−.56**	−.38	.74**	−.51*	−.32
	Mass	.47*	−.21	.63**	.31	−.18	.36

***Correlation is significant at the .01 level (2-tailed).*
**Correlation is significant at the .05 level (2-tailed).*

respective *mass* dimensions are only mildly related. However, there is a strong intercorrelation between the Greek *luminance* and English *mass* dimensions and a less strong effect between the Greek *texture* and English *luminance*. Although such findings support an inherent interlinguistic correspondence and a common understanding of specific notions in timbre semantics (Zacharakis et al., 2014), they ultimately confirm just a fair configurational similarity between the two semantic spaces, in comparison to the strong similarity between perceptual spaces as demonstrated above.

Table 5 presents the relationships between all heterologous (semantic vs. perceptual) pairs of spaces. The most prominent relationships among them are commented below. Greek *luminance* is related to the third Greek and English perceptual dimensions. Greek *texture* is mostly related to the first Greek and English perceptual dimensions, and, to a lesser degree, to all other perceptual dimensions. English *luminance* and *texture* also demonstrate similar relationships. However, the Greek *mass* presents no important association with any perceptual dimension, while the English *mass* is only correlated to the third Greek perceptual dimension. Thus, in general, the *mass* related semantic dimensions do not seem to have a clear (if any) perceptual counterpart. These findings regarding the strength of relationships between semantic and perceptual dimensions also confirm the just fair configurational similarities between heterologous timbre spaces that were reported in the previous subsection.

Multiple regression was subsequently performed in order to quantify the predictive potential of semantics upon perceptual dimensions, and thus provide a means for their "semantic labeling." The modeling was intra-language. That is, only combinations of same-language semantic and perceptual dimensions (GsGp and EsEp) were considered. The regression models with perceptual dimensions as dependent variables and semantic dimensions as predictors were constructed from subsets of predictors entering the regression equation at once (forced entry; Field, 2013). To facilitate substantive regression modeling, for each prediction equation (i.e., perceptual dimension as the dependent variable) the selected model was constructed from that subset of predictors (among all possible subsets) according to a triple heuristic criterion: the model should maintain a corrected Akaike Information Criterion (*AICc*; Hurvich & Tsai 1989) value up to 20% above the minimum, an adjusted R^2 (Field, 2013; Rao & Wu, 2001) value down to 20% below the maximum, and should also contain those semantic dimensions that showed statistically significant correlations with the predicted perceptual dimension. Both the AICc and the adjusted R^2 metrics are based on optimization of regression accuracy and penalization of model complexity. Well-fitted models have lower *AICc* values but higher adjusted R^2 values. Tables 6 presents the results of the multiple regression analysis.

The accuracy of predictions is acceptable (criterion: adjusted R^2 values $>.25$, again supporting the above reported fair configurational similarities. The best results are obtained in the cases of the third Greek perceptual dimension (*AICc* = -20.65 and adjusted $R^2 = .69$) and the first English (*AICc* = -13.73 and adjusted $R^2 = .58$). The least accurate fit was observed for the second perceptual dimension in both languages.

The regression models are in agreement with the correlations between dimensions, in terms of regression coefficient values and statistical significance. *Luminance* is the best predictor for the third dimensions. The first perceptual dimension in English is adequately modeled by *texture* which also affects the second dimension, but to a lesser extent. Although in Greek the contribution of *texture* on the first perceptual dimension is much less pronounced (*texture* also contributes to the second and third Greek perceptual dimensions), we claim that the

TABLE 6. *Multiple Regression Models Using Perceptual Dimensions as Dependent Variables and Semantic Dimensions of the Same Language as Predictors (i.e., Greek Semantics Predict Greek Perception and English Semantics Predict English Perception).*

		Semantic Predictors													
		intercept			**Luminance**			**Texture**			**Mass**				
	Perceptual dimensions	*B*	*S.E.*	β	*B*	*S.E.*	β	*B*	*S.E.*	β	*B*	*S.E.*	β	**Adj.-R^2**	**AIC**
Greek	**1st**	9e-3	.17	–	.33	.18	**.31**	−.67*	.18	**−.63**	–	–	–	**.39**	**−5.2**
	2nd	.05	.17	–	–	–	–	.55*	.18	**.53**	−.34	.18	**−.33**	**.34**	**−4.6**
	3rd	−.02	.12	–	.71*	.13	**.68**	.44*	.13	**.42**	–	–	–	**.68**	**−20.65**
English	**1st**	−.02	.12	–	–	–	–	.76*	.15	**.72**	.28	.15	**.26**	**.58**	**−13.73**
	2nd	.06	.19	–	–	–	–	−.50*	.19	**−.49**	−.15	.19	**−.15**	**.20**	**−3.02**
	3rd	−.02	.18	–	−.63*	.18	**−.60**	–	–	–	–	–	–	**.33**	**−5.01**

**Statistically significant at the .05 level (2-tailed)*

Greek first perceptual dimension may also reflect *texture*. Had the Greek data been the only available, it might not be acceptable to make such a claim. However, *texture* shows a clear association with the first English perceptual dimension, and the correlation between the first perceptual dimensions is very strong, $r(22) = .89$, $p < .01$. Additionally, *texture* is the semantic dimension featuring the highest interlanguage agreement (see Table 4). Thus, within the spirit of a pursued unification, *texture* is proposed as the optimal interpretation of the first perceptual dimensions. *Mass* does not appear to be a valid predictor for any perceptual dimension. Consequently, the perceptual dimensions can be loosely attributed the following semantic labeling: first perceptual dimension-*texture*, third perceptual dimension-*luminance*. The second dimension in both languages cannot be clearly "labeled" as it only relates mildly to *texture*.

Figure 4 shows the three 2D planes of the perceptual (MDS) spaces for both languages. The spatial configurations of the sounds could be commented under the prism of semantic-to-perceptual relationships. The *texture* related dimension (first) features the Moog and Acid synthesizers on the positive extreme (rough) and Bowedpad with flute on the negative extreme (smooth). At the same time the influence of F_0 on this dimension is evident as the higher F_0s within the set are generally positioned on the negative end and F_0 generally decreases towards the positive end (with the exception of the double bass pizzicato). The *luminance* related dimension (third) features double bass and Rhodes piano on the positive extreme (dull) and Farfisa, pipe organ, harpsichord, and sitar on the negative extreme (brilliant). Nevertheless, the second perceptual dimension, which was not strongly related to any semantic identifier, appears to express the notions of "percussiveness" or "transience" since impulsive sounds (e.g., marimba, piano, etc.) mostly occupy the positive half-planes (along the second perceptual dimension) while continuant instruments mostly reside on the negative ones.

ACOUSTIC CORRELATES OF PERCEPTUAL DIMENSIONS

A large set of low-level features (see Table B1, Appendix B) was extracted from the experimental sound set as an initial attempt to identify acoustic correlates for the perceptual dimensions obtained by MDS analysis. Identically to Zacharakis et al. (2014), the selection of acoustic features was based on the existing literature (e.g., Peeters, 2004; Peeters, Giordano, Susini, Misdariis, & McAdams, 2011), and they were calculated using the spectral modeling synthesis (SMS) MATLAB platform (Amatriain, Bonada, Loscos, & Serra, 2002). The window length applied was 4,096 samples ($f_S = 44.1$kHz) with an overlapping factor of 87.5%, the zero padding factor was 2, and 50 harmonic partials were extracted for all sounds. A variation of some basic features was also extracted using the instantaneous specific loudness of the ERB bands as calculated by Moore's loudness model (Moore, Glasberg, & Thomas, 1997) instead of the amplitude of the harmonics or the FFT bins. Finally, the mean, median, standard deviation, range, skewness and kurtosis of each acoustic descriptor were additionally computed in an effort to capture elements of the time-variant behavior of the sounds.

High multicollinearity within our acoustic features set was addressed applying the same procedure as in Zacharakis et al. (2014). The dimensionality of the feature set was reduced by means of principal components analysis (PCA) (see also Alluri & Toiviainen, 2010; Giordano, Rocchesso, & McAdams, 2010; Peeters et al., 2011). The Spearman coefficient correlation matrix was initially inspected, and when strongly correlated feature pairs [$\rho(22) \geq .80$] were identified, one of

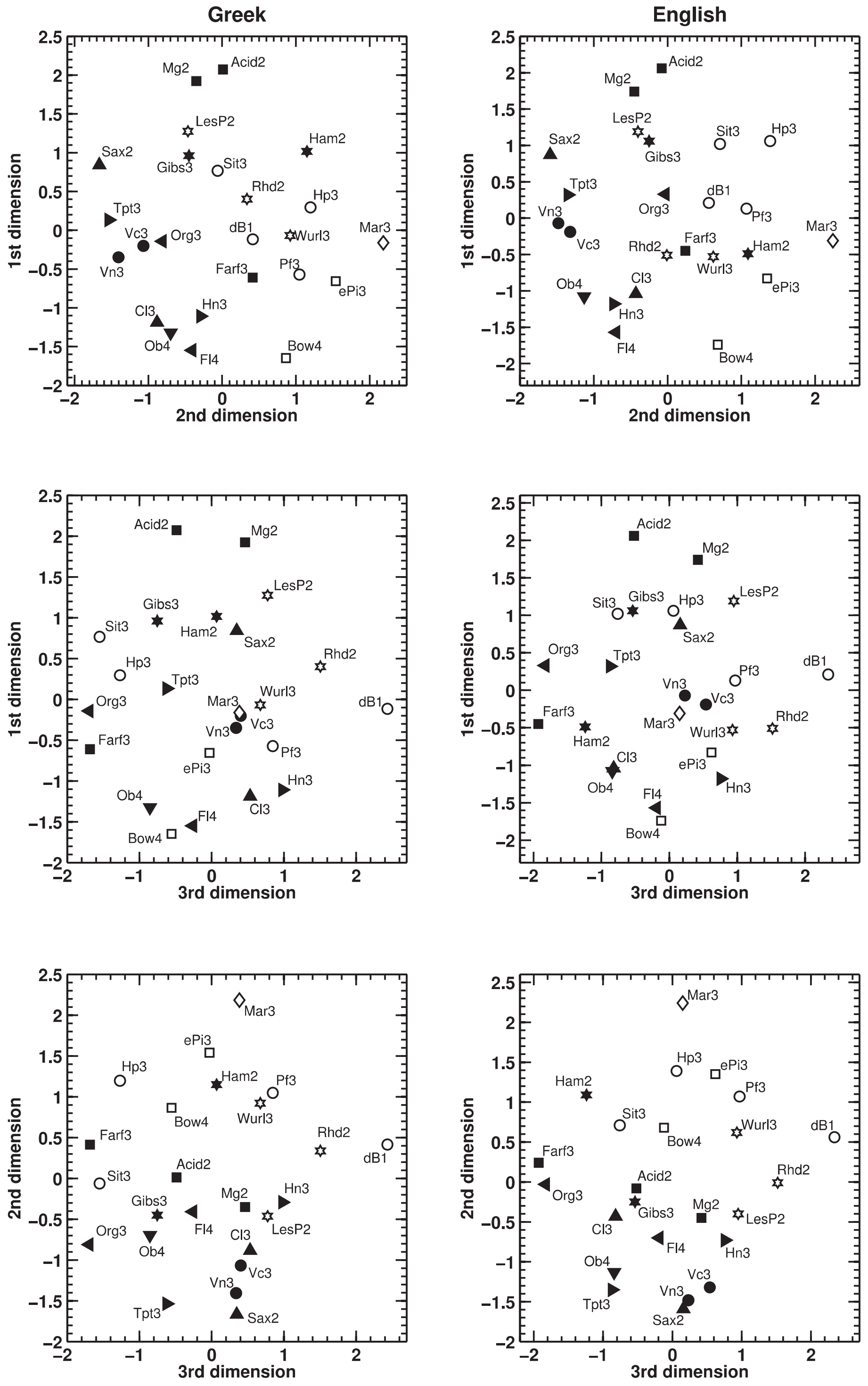

FIGURE 4. 2D planes of the perceptual (MDS) spaces for Greek (left) and (English). Black symbols: Continuant, white symbols: Impulsive, ▲: Single reed, ▼: Double reed, ◀: Aerophone, ▶: Lip reed, ●: Chordophone, ♦: Idiophone, ★: Electrophone, ■: Synthesizer. Abbreviations of instrument names, Acid: Acid, Bow: Bowedpad, Cl: clarinet, dB: double bass pizzicato, ePi: electric piano (rhodes), Farf: Farfisa, Fl: flute, Gibs: Gibson guitar, Ham: Hammond, Hn: French horn, Hp: Harpsichord, LesP: Les Paul Gibson guitar, Mar: marimba, Mg: Moog, Ob: oboe, Org: pipe organ, Pf: piano, Rhd: Rhodes piano, Sax: saxophone, Sit: sitar, Tpt: trumpet, Vc: *cello, Vn:* violin, Wurl: Wurlitzer. The number next to each instrument represents its $F_{0,}$ 1 for 55 Hz, 2 for 110 Hz, 3 for 220 Hz and 4 for 440 Hz.

them was discarded.[6] We then rank ordered the features and applied PCA to the reduced data set. Inspection of the anti-image correlation matrix[7] diagonal led to further removal of features whose individual Kaiser-Meyer-Olkin measure of sampling adequacy (*KMO*) was less than .50 so as to achieve an acceptable overall *KMO*. The final solution consisted of 4 components ($KMO = .64$, Bartlett's test of sphericity, $p < .001$) that explained 83.2% of the total variance. Table 7 shows the loadings of the features on the four components after orthogonal varimax rotation. The components are labeled based on the acoustic correlates that are highly loaded on each one. As shown in Table 7 and discussed below, the groupings of the acoustic features within principal components afford a qualitative perspective that is in close relevance with the existing literature. This organization of acoustic features facilitates a cohesive representation of the physical ground of timbre and its relationship to semantics and perception.

Features like the *normalized harmonic spectral centroid* (SC_norm), *tristimulus 3* (T3) (Pollard & Jansson, 1982), and *SC_loud_cor* (corrected version of the spectral centroid calculated from Moore's specific loudness in order to remove the influence of F_0; for an example, see Marozeau & de Cheveigné, 2007) all represent spectral structure (i.e., distribution of energy among harmonic partials) rather than spectral content (i.e., concentration of energy on frequency bands). Therefore, the first component is labeled: *energy distribution of harmonic partials (EDHP)*. The second component is represented by both *odd even ratio* (OER) and *inharmonicity* and we call it *spectral detail (SDT)*. The third component is related to two *spectrotemporal variation (STV)* characteristics such as *noisiness*, *harmonic spectral Flux* (Flux), and the *standard deviation of the harmonic spectral centroid* (SC_std). Finally, the fourth component is related to temporal characteristics such as *the logarithm of the attack time* (Log_At_time), *temporal centroid* (TC), and a *spectrotemporal* one (temporal variation of the first nine harmonics described by the *Mean coefficient of variation, MCV*, Kendall & Carterette, 1993b), thus we call it *temporal/spectrotemporal (T/STV)*. The small differences between the loadings on principal components reported in this work compared to the ones in Zacharakis et al. (2014) come from the fact that the stimulus set in this case was enhanced with an additional cello tone.

TABLE 7. *Loadings of the Audio Features on the First 4 Principal Components as a Result of PCA with Varimax Rotation.*

	Component			
	1 EDHP	2 SDT	3 STV	4 T/STV
T3	***.96***	*.06*	*−.02*	*.04*
SC_norm	***.94***	*.04*	*.05*	*−.04*
T2	***−.93***	*.16*	*.09*	*.14*
SC_loud_cor	***.85***	*.50*	*.03*	*.01*
Spread	***.73***	*.45*	*−.01*	*−.02*
SC_loud	***.70***	*.65*	*.01*	*.07*
OER	*−.17*	***−.77***	*−.03*	*−.15*
Inharmonicity	*.15*	***−.71***	*.43*	*−.35*
Noisiness	*.24*	*.08*	***.87***	*−.15*
Flux	*−.05*	*−.14*	***.82***	*.04*
SC_std	*−.30*	*.25*	***.72***	*.22*
Log_At_time	*.08*	*−.04*	*.23*	***.88***
MCV	*−.22*	*−.44*	*−.02*	***.76***
TC	*.24*	*−.47*	*−.13*	***.74***
SC_var_loud	*−.62*	*−.61*	*−.07*	*−.15*

Note: Loadings ≥ .7 are presented in bold and used for labeling the components. See Table B1 for the abbreviations. EDHP = energy distribution of harmonic partials; SDT = spectral detail; STV = spectrotemporal variation; T/STV = temporal/spectrotemporal variation).

Acoustical interpretation of perceptual dimensions. The relationships between acoustic components and perceptual dimensions were examined in the same way as the relationships between perceptual and semantic dimensions. Such an investigation may facilitate a subsequent interpretation and labeling of the perceptual dimensions in terms of acoustical signal properties.

Pearson's correlation coefficients among the principal acoustic components and the perceptual dimensions for each linguistic group were initially calculated. We have also included F_0 in the correlation analysis to examine its influence on the formulation of the spaces. Table 8 presents the obtained correlation coefficients.

All statistically significant correlations show medium ($r > .30$) and large effects ($r > .50$) (Ellis, 2010). As expected by the high similarity between the Greek and English perceptual spaces the major acoustic correlates for each space are also almost similar.

The first perceptual dimension for both languages is strongly correlated to F_0, $r(22) = -.74$, $p = .001$ for Greek and $r(22) = -.63$, $p = .001$ for English, and to EDHP, $r(22) = .64$, $p = .001$ for Greek and $r(22) = .73$, $p = .001$ for English. The second dimension in both languages shows strong correlations to T/STV, $r(22) = -.62$, $p = .001$ for Greek and $r(22) = -.65$, $p = .001$ for

---

[6] The principles by which the rejection of multicollinear features was performed was to always prefer the mean statistic over the rest and to favor features that are commonly found in the literature over less common descriptors.

[7] The anti-image correlation matrix contains measures of sampling adequacy for each variable along the diagonal and the negatives of the partial correlation on the off-diagonals.

TABLE 8. *Pearson's Correlation Coefficients Between Acoustic Components and Perceptual Dimensions for Both Linguistic Groups.*

	Greek perceptual dimensions			**English perceptual dimensions**		
Acoustic Features	1^{st}	2^{nd}	3^{rd}	1^{st}	2^{nd}	3^{rd}
F_0	$-.74^{**}$	$-6e\text{-}3$	$-.44^{*}$	$-.63^{**}$	$-.08$	$-.32$
EDHP	$.64^{**}$	$-.60^{**}$	$-.41^{*}$	$.73^{**}$	$-.44^{*}$	$-.45^{*}$
SDT	$-.36$	$-.14$	$-.66^{**}$	$-.21$	$-.13$	$-.61^{**}$
STV	$.11$	$.19$	$-.10$	$.20$	$.26$	$.01$
T/STV	$-.40$	$-.62^{**}$	$.12$	$-.40$	$-.65^{**}$	$.02$

***Correlation is significant at the .01 level (2-tailed).*

**Correlation is significant at the .05 level (2-tailed).*

Note: EDHP = energy distribution of harmonic partials; SDT = spectral detail; STV = spectrotemporal variation; T/STV = temporal/spectrotemporal variation; F_0 *= fundamental frequency.*

English, and a strong and medium correlation with EDHP in Greek and English, respectively, $r(22) = -.60$, $p = .001$ for Greek and $r(22) = -.44$, $p = .001$ for English. The third dimension maintains a strong relation with SDT, $r(22) = -.66$, $p = .001$ for Greek and $r(22) = -.61$, $p = .001$ for English, and a medium one with EDHP, $r(22) = -.41$, $p = .001$ for Greek and $r(22) = -.45$, $p = .001$ for English. An additional moderate relation with F_0 is also observed for the third Greek dimension, $r(22) = -.44$, $p = .001$. STV did not feature any correlation with any of the perceptual dimensions.

Multiple regression analysis was again performed to examine the prediction of each perceptual dimension from the available acoustic components. The employed criteria for the construction of the regression models were the same as in the prediction of perceptual dimensions by semantic dimensions. The results of the analysis are given in Table 9.

The accuracy of predictions is high (all adjusted $R^2 > .53$), which allows the interpretation and labeling of perceptual dimensions in terms of acoustic descriptors. The regression models are in good agreement with the correlations between perceptual dimensions and acoustic features. An additional statistically significant contribution of F_0 is detected for the third English perceptual dimension.

In total, the correlation and the multiple regression analyses show that the first perceptual dimension for both languages is predicted by a combination of F_0 and the energy distribution of harmonic partials (EDHP). However, the importance of these two predictors is reversed between Greek and English having F_0 as the prominent predictor in Greek and EDHP as the prominent predictor in English. The third dimension is also adequately modeled by EDHP and spectral detail (SDT), and to a lesser degree by F_0. Finally, the second dimension is again influenced by EDHP combined with the temporal/spectrotemporal (T/SPV) acoustic component. STV had no predictive influence.

Discussion

The results of the previous section have shed some light on the relationship between semantics and perception of musical timbre and have also provided some insight into the potential influence of native language on timbre description and perception. Configurational similarity measures and measures of correlation between dimensions revealed that the Greek and English MDS spaces were almost identical, implying robustness of musical timbre perception across these two different linguistic populations. All the remaining relationships among perceptual and semantic spaces were found to be fair. This suggests that verbal descriptions — regardless of language — were capable of reflecting a substantial amount of perceptual relations among the musical timbres under study (accepting that perception of timbre is adequately represented through pairwise dissimilarity ratings).

However, as one could expect, the configurational similarity does not necessarily imply some form of overall similarity between two spaces. Rather, the overall similarity may also affect the degree of the configurational similarity. Therefore, complementary analyses that ultimately target the one-to-one relationships between dimensions rather than merely shape similarity is desirable.

In this spirit, subsequent investigation revealed that the observed configurational similarities were also accompanied by analogous relationships between the dimensions of semantic and perceptual spaces. Correlation and regression analyses have shown that the first MDS dimension could represent auditory texture while the third MDS dimension could represent auditory luminance. The second MDS dimension, however, was not adequately related with any of the semantic dimensions, as auditory mass failed to consistently correlate with any perceptual dimension. This may either imply that the second MDS dimension completely lacks

TABLE 9. *Multiple Regression Models for Perceptual Dimensions as Dependent Variables and Acoustic Descriptors as Predictors.*

				Acoustic Features																
Perceptual dimensions		**intercept**		**F_0**			**EDHP**			**SDT**			**STV**			**T/STV**			**Adj.-R2**	**AIC**
		B	S.E.	B	S.E.	β	B	S.E.	β	B	S.E.	β	B	S.E.	β	B	S.E.	β		
Greek	**1st**	1.26*	.29	−6e-3*	1e-3	**−.59**	−.46*	.13	**.44**	–	–	–	–	–	–	–	–	–	.70	−20.03
	2nd	−.02	.12	–	–	–	−.66*	.13	**−.61**	–	–	–	–	–	–	−.62*	.12	**−.62**	.71	−15.47
	3rd	.80	.37	−4e-3*	2e-3	**−.39**	−.57*	.14	**−.56**	−.45*	.16	**−.44**	–	–	–	–	–	–	.65	−11.58
English	**1st**	.93*	.30	−4e-3*	1e-3	**−.43**	.60*	.13	**.59**	–	–	–	–	–	–	–	–	–	.68	−17.16
	2nd	−2e-3	.14	–	–	–	−.45*	.14	**−.45**	–	–	–	–	–	–	−.67*	.15	**−.63**	.54	−7.62
	3rd	.51	.42	−2e-3*	2e-3	**−.25**	−.56*	.16	**−.55**	−.48*	.18	**−.47**	–	–	–	–	–	–	.56	−10.82

***Correlation is significant at the .01 level (2-tailed).*

**Correlation is significant at the .05 level (2-tailed).*

Note: EDHP = energy distribution of harmonic partials; SFS = spectral fine structure; STV = spectrotemporal variation; T/STV = temporal/spectrotemporal variation; F_0 = fundamental frequency.

a semantic "charge" or that its semantic interpretation should be sought among additional appropriate descriptors. This, in turn, suggests the potential expansion of our previous results (Zacharakis et al., 2014) so as to include semantic terms able to express such effects. At this point, it also has to be highlighted that *mass* carried the smallest amount of semantic information for both languages and was the least clearly understood semantic dimension between them (see Zacharakis et al., 2014). Therefore, a further evaluation of its potency as a semantic dimension of musical timbre is mandated.

The fit between semantic and perceptual dimensions was acceptable and an interpretational improvement was observed in comparison to previous studies (e.g., Kendall & Carterette, 1993a, 1993b; Kendall et al., 1999). Therefore, these results realize the prospects set by Samoylenko et al. (1996) regarding wider associations between semantics and perception of timbre. Furthermore, the fact that heterologous interlanguage relations both at configurational and dimensional levels were of very similar strength to the intralanguage ones further supports the notion that the process of capturing perception through semantic description is not seriously affected by language.

However, differences between the perceptual and semantic spaces do exist and could be explained by two possible scenarios that require further investigation. The first would be that timbre perception cannot be completely captured by mere semantic description as some perceived aspects of sound are impossible to describe. To this end, the probability that representation of timbre through sensory modalities such as sight or touch may be complementary to semantic description might be worth examining. The second possibility would be that the inclusion of additional semantic scales, upon which a rating is made (e.g., replacing *mass* with a descriptor of non-stationarity), may result in increased common information between semantics and perception. Apart from artistic applications (e.g., contributing to musical creativity and interaction), pursuing such research directions could also improve music appreciation for specific groups of listeners (e.g., hearing impaired individuals).

The results of our analysis extend our previous findings (Zacharakis et al., 2014) regarding acoustic correlates of timbral semantic dimensions, with the consideration of their effect on perceptual dimensions. The acoustic features correlate highly with perceptual dimensions and allow their accurate prediction. The energy distribution of harmonic partials (EDHP), which was generally present in most perceptual dimensions for both languages, seems to mostly affect the texture-related English perceptual dimension (first), and to a lesser extent the remaining ones. The picture is slightly different for the Greek data where the first MDS dimension is better predicted by F_0 than by EDHP. It appears that this could account for the small percentage of uncommon variance between the first perceptual dimensions. However, the effects are of similar order, which leaves small space for rigorous interpretation. The fact that EDHP (rather than F_0) was the acoustic descriptor that had been previously associated with *texture* (Zacharakis et al., 2014) might account for Greek *texture* appearing "shifted" towards the second MDS dimension in comparison to the English *texture*. The association of auditory texture with EDHP is also confirmed by the fact that both are similarly correlated with most of the MDS dimensions. Furthermore, we now have more evidence that auditory luminance (i.e., third MDS dimension) is related to SDT (i.e., OER and inharmonicity). The semantically "unlabeled" second perceptual dimension seems to be influenced by temporal and spectrotemporal

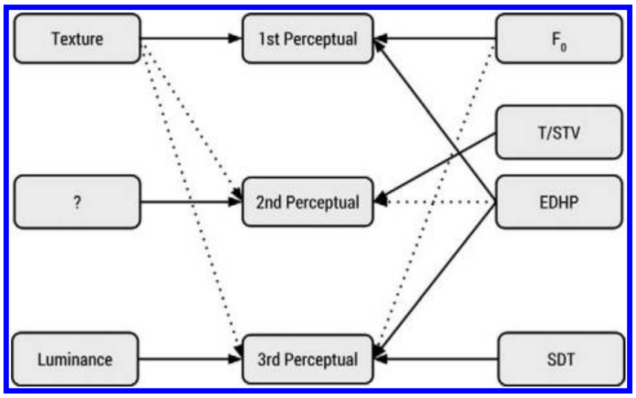

FIGURE 5. A proposal for a unified framework among timbre perception, semantics and acoustics. Solid lines represent primary effects and dotted lines represent weaker associations.

acoustic characteristics. This could imply that the provided semantic descriptions have mainly been able to capture static/stationary phenomena (i.e., spectral content and detail) but were deprived of the potential to represent dynamic aspects of the sounds such as the attack time and temporal centroid. The summary of the identified relationships among perceptual dimensions, semantics and acoustic correlates is given in Figure 5.

Whereas these results agree with the existing literature regarding acoustic correlates of perceptual dimensions, there seems to be some discrepancy concerning the semantic interpretation. Most studies (Grey & Gordon, 1978; Iverson & Krumhansl, 1993; Kendall et al., 1999; Krimphoff, 1993; Krumhansl, 1989; Lakatos, 2000; McAdams et al., 1995;) have identified the spectral centroid and some measure of impulsiveness (e.g., logarithm of the attack time) as the acoustic correlates of the first and second MDS dimensions, respectively. On the other hand, the third dimension of a typical 3D perceptual space appears more controversial, as some works have linked it to spectral variation (Kendall & Carterette, 1993b; Kendall et al., 1999; Krimphoff, McAdams, & Winsberg, 1994; McAdams, 1999), while others link it to spectral fine structure (Krumhansl, 1989; Krimphoff, 1993b, McAdams, 1999). Our findings do not contradict the literature in this respect. However, when it comes to semantics, the typically reported univocal relation between auditory brightness and spectral content was not so evident in our work. Rather, the energy distribution of the partials was mostly associated with *texture*, while *luminance* was related to both energy distribution of the partials and spectral detail.

Finally, an observation regarding the possible effect of different F_0 values to the listeners' judgments has to be made; participants were capable of providing meaningful judgments of timbral dissimilarity even within an F_0 range of three octaves. Although F_0 was correlated with the first MDS dimensions (supporting the findings of Marozeau et al., 2003, and Marozeau & de Cheveigné 2007), F_0 variation was by no means dominating over the other perceptual dimensions as has been the case with previously reported results on simple synthetic stimuli (Miller & Carterette, 1975). A possible explanation could be that the inherent timbral diversity stemming from complex musical timbres like the ones used in our investigation (combination of natural and synthetic instruments) prevailed over the effect of F_0 variation.

Conclusion

This study investigated the relationship of timbral semantics with timbre perception and the extent to which this relationship was influenced by native language. Two semantic timbre spaces obtained using verbal attribute magnitude estimation listening experiments were compared with two perceptual timbre spaces from pairwise dissimilarity rating experiments. All four timbre spaces concerned the same stimulus set and each type of experiment was performed by separate groups of English- and Greek-speaking participants. Additionally, acoustic features were employed to provide physical interpretation and labeling of the perceptual dimensions. Thus, the present work was an attempt to provide an interlinguistic unifying framework of musical timbre description in terms of semantic, perceptual, and acoustical consideration.

The main findings can be summarized as follows:

1) The strong similarity between the two perceptual spaces suggests that native language has no effect on timbre perception, at least for the two linguistic cultures tested.
2) The fair configurational similarity between semantic and perceptual spaces combined with some significant correlations between semantic and perceptual dimensions shows that verbal description of timbral qualities can indeed capture some aspects of the perceptual structure within a set of timbres. It also provides a partial basis for a "semantic labeling" of perceptual dimensions.
3) An interpretation of perceptual spaces by means of acoustic terms showed that the energy distribution of harmonic partials (EDHP) is the most prominent acoustic correlate. It is associated with all three perceptual dimensions for both languages. The first

perceptual dimension receives the strongest association of EDHP and a major contribution of F_0. The third perceptual dimension is also affected by EDHP and F_0 together with spectral detail (SDT). Finally, the second perceptual dimension is adequately modeled by the combination of energy distribution of harmonic partials (EDHP) and temporal/spectrotemporal variation (T/STV).

The configurational similarity between the semantic and perceptual spaces, even though not very strong, indicates that the relationships between sounds within one space closely resemble the respective relationships in the other. This holds for both languages. The overall similarity (configurational and dimensional) between semantic and perceptual spaces allows for a hypothesis that there may exist a substantial latent influence of timbre semantics on pairwise dissimilarity judgments. That is, the perceived dissimilarity between a pair of different timbres might be influenced by combined evaluations over several latent semantic constructs such as auditory texture and luminance. This hypothesis could be further investigated by combining an MDS with a structural equation modeling (confirmatory factor analysis) approach.

An important implication of this result is that changes within one space may produce similar alterations to the other. The processing of a musical sound based on semantic terms would produce a predictable and similar shift of its location in the perceptual space. It would also affect the sound's perceived relationships with other timbral objects in a similar manner. As an application example, a collection of appropriate semantic scales linked with a specialized lexicon of signal processing operations could be used to drive certain types of perceived sound transformations.

Author Note

The authors wish to thank Prof. Stephen McAdams, Dr. Andy Simpson, Dr. Michael Terrell, Prof. Steven Handel, Prof. Petri Toiviainen, and two anonymous reviewers for their very helpful comments on previous versions of this manuscript. Finally, we would like to thank all the people who participated in the listening experiments.

Correspondence concerning this article should be addressed to Asterios Zacharakis, Aristotle University of Thessaloniki, School of Fine Arts, Music Department, 54124, University Campus, Thessaloniki, Greece, E-mail: aszachar@mus.auth.gr

References

Alluri, V., & Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception*, *27*, 223-242.

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, *59,* 3677-3689.

Amatriain, X., Bonada, J., Loscos, A., & Serra, X. (2002). Spectral processing. In Udo Zölzer (Ed.), *DAFX - Digital Audio Effects* (pp. 373-438). Chichester, England: John Wiley and Sons Ltd.

Andrews, F. M., & Inglehart, R. F. (1978). The structure of subjective well-being in nine Western societies. *Social Indicators Research*, *6*, 73-90.

Bensmaïa, S., & Hollins, M. (2005). Pacinian representations of fine surface texture. *Perception and Psychophysics*, *67*, 842-854.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). New York: Springer.

Borg, I., & Leutner, D. (1985). Measuring the similarity of MDS configurations, *Multivariate Behavioral Research*, *20*, 325-334.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, *118*, 471-482.

Cho, G., Kim, C., & Casali, J. G. (2002). Sensory evaluation of fabric touch by free modulus magnitude estimation. *Fibers and Polymers*, *3*, 169-173.

Cutzu, F., & Edelman, S. (1996). Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Sciences*, *93*, 12046-12050.

DeCarlo, L. T., & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*, *119*, 375-396.

Efron, B., & Tibshirani, R. (1993). An introduction to the bootstrap. London, UK: Chapman and Hall.

Elliot, T. M., Hamilton, L. S., & Theunissen, F. E. (2012). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *Journal of the Acoustical Society of America*, *133*, 389-404.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results.* Cambridge, UK: Cambridge University Press.

Faure, A., McAdams, S., & Nosulenko, V. (1996). Verbal correlates of perceptual dimensions of timbre. In B. Pennycook

& E. Costa-Giomi (Eds.), *Proceedings of the 4th International Conference on Music Perception and Cognition (ICMPC 04)* (pp. 79-84). Montreal, Canada: McGill University.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, UK: Sage Publications Ltd.

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 462-479.

Gower, J. C. (1971). Statistical methods for comparing different multivariate analyses of the same data. In J. R. Hodson, D. G. Kendall, & P. Tautu (Eds.), *Mathematics in the archaeological and historical sciences* (pp. 138-149). Edinburgh, Scotland: Edinburgh University Press.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, *40*, 33-51.

Gower, J. C., & Dijksterhuis, G. B. (2004). Procrustes problems. Oxford, UK: Oxford University Press.

Green, B. G., Dalton, P., Cowart, B., Shaffer G., Rankin, K., & Higgins, J. (1996). Evaluating the 'Labeled Magnitude Scale' for measuring sensations of taste and smell. *Chemical Senses*, *21*, 323-334.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, *61*, 1270-1277.

Grey, J. M., & Gordon J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, *63*, 1493-1500.

Helmholtz, H. L. F. von (1954). *On the sensations of tone* (A. J. Ellis, Trans.). New York: Dover. (Original work published 1877)

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

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, *94*, 2595-2603.

Jackson, D. A. (1995). PROTEST: A PROcrustean randomization TEST of community environment concordance. *Ecoscience Sainte-Foy*, *2*, 297-303.

Kendall, R. A., & Carterette, E. C. (1993a). Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's adjectives. *Music Perception*, *10*, 445-468.

Kendall, R. A., & Carterette, E. C. (1993b). Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's Orchestration. *Music Perception*, *10*, 469-502.

Kendall, R. A., & Carterette, E. C. (1996). Difference thresholds for timbre related to spectral centroid. In B. Pennycook & E. Costa-Giomi (Eds.), *Proceedings of the 4th International Conference on Music Perception and Cognition ICMPC 04* (pp. 91-95). Montreal, Canada: McGill University.

Kendall, R. A., Carterette, E. C., & Hajda, J. M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, *16*, 327-364.

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews, 8*, 89-105.

Krimphoff, J. (1993). *Analyse acoustique et perception du timbre* [Acoustic analysis and perception of timbre] Unpublished master's thesis). Université du Maine, Le Mans, France.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II : Analyses acoustiques et quantification psychophysique. [Characterization of the timbre of complex sounds. 2. Acoustic analysis and psychophysical quantification]. *Journal de Physique*, *4*, 625-628.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music: Proceedings of the Marcus Wallenberg symposium* (pp. 43-53). Lund, Sweden: Excerpta Medica, Amsterdam.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 739-751.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrica, 29*, 1-28.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrica, 29*, 115-130.

Langeheine, R. (1982). Statistical evaluation of measures of fit in the Langeheine-Borg Procrustean individual differences scaling. *Psychometrika*, *47*, 427-442.

Lakatos, L. (2000). A common perceptual space for harmonic and percussive timbres. *Perception and Psychophysics*, *62*, 1426-1439.

Legendre, P., & Legendre, L. (1998). *Numerical ecology* (92nd English ed.). Amsterdam: Elsevier Science.

Lichte, W. (1941). Attributes of complex tones. *Journal of Experimental Psychology*, *28*, 455-480.

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*, 57-64.

Marks, L. E. (1980). The effect binaural summation of loudness: Noise and two-tone complexes. *Perception and Psychophysics*, *27*, 489-498.

Marozeau, J., & de Cheveigné, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *Journal of the Acoustical Society of America*, *121*, 383-387.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, *114*, 2946-2957.

McAdams, S. (1999). Perspectives on the contribution of timbre to musical structure. *Computer Music Journal*, *23*, 85-102.

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*, 177-192.

Meulman, J. J., & Heiser, W. J. (2008). *Categorical principal component analysis (CATPCA).* Chicago, IL: SPSS Inc.

Miller, J. R., & Carterette, E. C. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America*, *58*, 711-720.

Moore, B. C. J., Glasberg, B. R., & Thomas, B. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, *45*, 224-240.

Opolko, F., & Wapnick, J. (2006). *McGill university master samples collection on DVD.* Montreal, Canada: McGill University.

Painter, J. G., & Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? An ERP study on the processing of meaning in music. *Psychophysiology*, *48*, 645-655.

Peeters, G. (2004). *A large set of audio features for sound description (similarity and classification) in CUIDADO project.* CUIDADO 1st Project Report. Paris, France: IRCAM.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting acoustic descriptors from musical signals. *Journal of the Acoustical Society Of America*, *130*, 2902-2916.

Peres-Neto, P. R. (2000). *Documentation for program PROTEST.EXE.* Toronto, Canada: University of Toronto.

Peres-Neto, P. R., & D. A. Jackson. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia 129*, 169-178.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 397-414). Leiden, Netherlands: Sijthoff.

Pollard, H., & Jansson, E. (1982). A tristimulus method for the specification of musical timbre. *Acustica*, *51*, 162-171.

Rao, C. R., & Wu, Y. (2001). On model selection. IMS *Lecture Notes-Monograph Series, 38,* 1-57.

Samoylenko, E., McAdams, S., & Nosulenko, V. (1996). Systematic analysis of verbalizations produced in comparing musical timbres. *International Journal of Psychology*, *31*, 255-278.

Schönemann, P. H., & Carroll, R. M. (1970). Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, *35*, 245-255.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function: I. *Psychometrica*, *27*, 125-140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function: II. *Psychometrica, 27*, 219-246.

Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology, 3*, 287-315.

Spence, I. (1979). A simple approximation for random rankings stress values. *Multivariate Behavioral Research*, *14*, 355-365.

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review*, *78*(5), 426-450.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* [Personnel Research Section Report No. 984]. Washington, DC: Department of the Army.

von Bismarck, G. (1974). Timbre of steady tones: A factorial investigation of its verbal attributes. *Acustica*, *30*, 146-159.

Yoshioka, T., Bensmaïa, S. J., Craig, J. C., & Hsiao, S. S. (2007). Texture perception through direct and indirect touch: An analysis of perceptual space for tactile textures in two modes of exploration. *Somatosensory and Motor Research*, *24*, 53-70.

Young, F. W. (1970). Nonmetric multidimaneional scaling: Recovery of metric information. *Psychometrica, 35*, 455-473.

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2014). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception, 31,* 337-356.

Appendix A

ASSESSMENT OF THE INFLUENCE OF SMALL PITCH DIFFERENCES ON TIMBRE PERCEPTION

In order to assess whether the mild pitch shifting that we applied to a few of our sound stimuli had not affected their timbral quality, we performed an extra pairwise dissimilarity listening test merely on these five stimuli (marimba, harpsichord, pipe organ and Gibson guitar at G3 and French horn at A#3) and their pitch-shifted versions.

Twenty-seven participants (age range = 20-23, mean age = 21.4, 20 female) took part in the listening test. None of the participants reported any hearing loss or absolute pitch, and they had been practicing music for 12.9 years on average, ranging from 6 to 19. The absence of absolute pitch from the group of our participants was again a prerequisite as such a condition could affect the results due to pitch variation within the stimulus set. All participants were students in the Department of Music Studies of the Aristotle University of Thessaloniki and were provided with course bonus for their participation.

The procedure and instructions given were identical to the ones of the main listening test. The overall experiment time, including instructions, lasted around thirty minutes on average.

The pairwise dissimilarity data were analyzed through the same MDS analysis (INDSCAL) as the one applied to the data of the main experiment. Table A1 shows the measures-of-fit for three dimensionalities. The minimal improvement of the measures-of-fit between the 2D and 3D solutions suggests that a 2D space is adequate to model the relationships within this stimulus set.

As shown in Figure A1, the original and pitch shifted versions of each stimulus occupy almost the same position in the two-dimensional timbre space. Therefore, it seems that the timbre of these five sounds was unaffected by the pitch shifting.

TABLE A1. *Measures-of-fit and Their Improvement for Different MDS Dimensionalities Concerning the Pairwise Dissimilarity Experiment That Examined the Influence of Mild Pitch Shifting on Timbre.*

Dimensionality	***S-Stress***	***Improv.***	***D.A.F.***	***Improv.***
1D	*.21*	–	*.90*	–
2D	*.08*	*.13*	*.97*	*.07*
3D	*.04*	*.04*	*.99*	*.02*

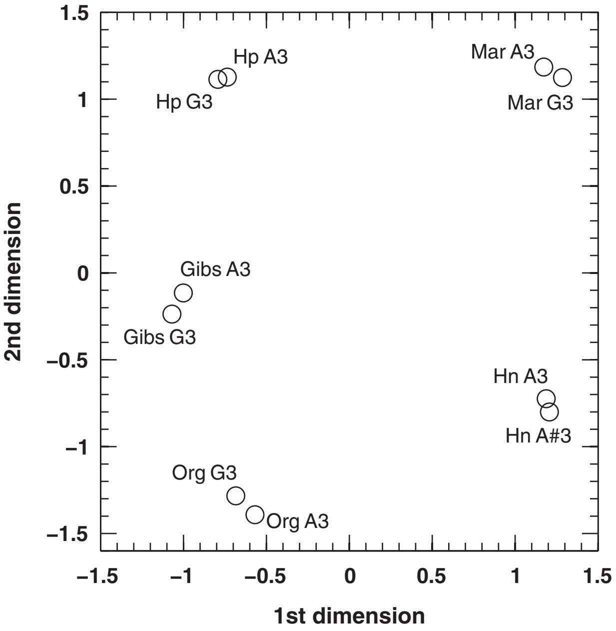

FIGURE A1. The 2-dimensional timbre space shows the positions of the original and pitch shifted version of each stimulus.

Appendix B

THE EXTRACTED AUDIO FEATURES

TABLE B1. Abbreviations and Definitions of the Significant Audio Features.

Category	Feature	Abbreviation	Explanation
Spectral Content	Harmonic Spectral Centroid	SC	Barycenter of the harmonic spectrum (Peeters et al., 2011)
	Spectral Centroid (loudness model)	SC_loud	SC of the specific loudness (Moore et al., 1997)
Energy distribution of harmonic partials	Normalized Harmonic Spectral Centroid	SC_norm	Normalized barycenter of the harmonic spectrum
	Tristimulus 1, 2, and 3	T1, T2, T3	Relative amplitudes of the 1st, the 2nd to the 4th and the 5th to the rest of the harmonics (Pollard & Jansson, 1982)
	Harmonic Spectral Spread	Spread	Spread of the harmonic spectrum around its mean value (Peeters et al., 2011)
	SC (loudness model) corrected	SC_loud_cor	SC of the specific loudness corrected for F0 (Moore at al., 1997; Marozeau & de Cheveigné, 2007)
Spectrotemporal	Harmonic Spectral Flux (or variation)	Flux	Amount of variation of the harmonic spectrum over time (Krimphoff, 1993)
	Mean Coefficient of Variation	MCV	Variation of the first 9 harmonics over time (Kendall & Carterette, 1993b)
	SC standard deviation	SC_std	SC standard deviation over time
	SC variation	SC_var	SC_std/SC_mean (Krimphoff, 1993)
	SC variation (loudness)	SC_var_loud	SC variation of the specific loudness
	Noisiness	Noisiness	Ratio of the noise energy to the total energy (Peeters et al., 2011)
Spectral fine structure	Harmonic Spectral Irregularity	Sp_Irreg	Measure of the harmonic spectrum fine structure (Kendall & Carterette, 1996)
	Odd Even Ratio	OER	Ratio of the energy contained in odd versus even harmonics (Peeters et al., 2011)
Harmonic series	Inharmonicity	Inharmonicity	Measure of the degree to which partials depart from whole multiples of the fundamental frequency (Peeters et al., 2011)
Temporal	Log of attack time	Log_At_time	Logarithm of the rise time (Peeters et al., 2011)
	Temporal Centroid	TC	Barycenter of the energy envelope (Peeters et al., 2011)
	Normalized Temporal Centroid	TC_norm	TC/duration