UNIVERSITY OF CALIFORNIA PRESS
JOURNALS + DIGITAL PUBLISHING

An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates
Author(s): Asterios Zacharakis, Konstantinos Pastiadis and Joshua D. Reiss
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 31, No. 4 (April 2014), pp. 339-358
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/10.1525/mp.2014.31.4.339

AN INTERLANGUAGE STUDY OF MUSICAL TIMBRE SEMANTIC DIMENSIONS AND THEIR ACOUSTIC CORRELATES

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

A STUDY OF MUSICAL TIMBRE SEMANTICS WAS conducted with listeners from two different linguistic groups. In two separate experiments, native Greek and English speaking participants were asked to describe 23 musical instrument tones of variable pitch using a predefined vocabulary of 30 adjectives. The common experimental protocol facilitated the investigation of the influence of language on musical timbre semantics by allowing for direct comparisons between linguistic groups. Data reduction techniques applied to the data of each group revealed three salient semantic dimensions that shared common conceptual properties between linguistic groups namely: luminance, texture, and mass. The results supported universality of timbre semantics. A correlation analysis between physical characteristics and semantic dimensions associated: i) texture with the energy distribution of harmonic partials, ii) thickness (a term related to either mass or luminance) and brilliance with inharmonicity and spectral centroid variation, and iii) F_0 with mass or luminance depending on the linguistic group.

Received: November 12, 2012, accepted July 5, 2013.

Key words: musical timbre semantics, timbre spaces, verbal attribute magnitude estimation, CATPCA, acoustic correlates

TIMBRE IS A FUNDAMENTAL ATTRIBUTE OF auditory perception (Helmholtz, 1877/1954). Much recent work on timbre has investigated music materials (Grey, 1977; Grey & Gordon, 1978; Kendall & Carterette, 1993a; Krimphoff, McAdams, & Winsberg, 1994; Krumhansl, 1989; McAdams, Winsberg, Donnadieu, Soete, & Krimphoff, 1995). For an overview of studies on musical timbre, see Hajda (2007) and Donnadieu (2007). ANSI (1973) defines timbre as: "that attribute of auditory sensation in terms of which a subject can judge that two sounds similarly presented and having the same loudness, pitch and duration are dissimilar" (p. 56). However, the ANSI definition has also been the target of some criticism (Donnadieu, 2007; Pratt & Doak, 1976; Sankiewicz & Budzynski, 2007). Various researchers have shown that musical timbre has a multidimensional nature (Grey, 1977).

Several methods have been employed to investigate timbre perception. The most popular approach is to measure the perceived pairwise dissimilarity between sound stimuli (Caclin, McAdams, Smith, & Winsberg, 2005; Iverson & Krumhansl, 1993; Miller & Carterette, 1975; Plomp, 1970, 1976). The resulting dissimilarity matrices are then analyzed within the multidimensional scaling (MDS) framework so as to obtain a geometric configuration of the timbres under study that is generally known as timbre space (e.g., Grey, 1977). Timbre spaces are useful constructs that allow a physical interpretation of the major perceptual dimensions of timbre (i.e., dimensions that best explain the perceived dissimilarities between the stimuli). Previous studies on the perception of musical timbre have identified either three or four major perceptual dimensions for modeling timbres of monophonic acoustic instruments (Grey, 1977; Krimphoff, 1993; Krimphoff et al., 1994; McAdams et al., 1995). Another popular method is to perform a discrimination task whereby certain acoustic parameters are controlled by the researcher to directly investigate and quantify their perceptual significance (Grey, 1978; Hajda, 1998, 1999; McAdams et al., 1999).

While the above methods are particularly effective for exploring the perceptual properties of timbre, they are not designed to investigate its semantic dimensions. The mapping between a semantic and a perceptual timbre space must be complex and partial since not all perceivable attributes of sound can be adequately verbalized, and also because verbalization might be a product of conditioning. However, interest in timbre semantics has a long

Music Perception, VOLUME 31, ISSUE 4, PP. 339–358, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312. DOI: 10.1525/MP.2014.31.4.339

history (Helmholtz, 1877/1954; Lichte, 1941) as researchers tried to associate the meaning attributed to sound objects with their physical properties. Technological advances allow the potential development of a common, language-independent semantic framework for timbre description, which could be exploited for intuitive sound synthesis and sound processing.

As a result, a complementary approach that aims to bridge perception with semantics of timbre has been adopted by many researchers. The objective in this case is the elicitation of verbal descriptors, usually in the form of adjectives (Kendall & Carterette, 1993a, 1993b; von Bismarck, 1974a, 1974b). According to this method, sound objects are represented by a feature vector of semantic attributes rather than by their relative distances. This is based on the hypothesis that timbre can be adequately described by the use of semantic scales (Samoylenko, McAdams, & Nosulenko, 1996). The concept of using verbal attributes has also been applied for describing properties of musical instruments and characteristics of their performance (Barthet, Guillemain, Kronland-Martinet, & Ystad, 2010; Disley & Howard, 2004; Fritz, Blackwell, Cross, Woodhouse, & Moore, 2012; Nykänen, Johansson, Lundberg, & Berg, 2009; Saitis, Giordano, Fritz, & Scavone, 2012), polyphonic timbre (Alluri & Toiviainen, 2010), and acoustic assessment of concert halls (Lokki, Pätynen, Kuusinen, Vertanen, & Tervo, 2011). An overview of various methods that can be used for elicitation of verbal descriptions is given by Neher, Brookes, and Rumsey (2006).

The most widely applied methods for obtaining semantic descriptions of timbre are the semantic differential (Lichte, 1941; Osgood, Suci, & Tannenbaum, 1957; von Bismarck, 1974a) and one variant of this method, verbal attribute magnitude estimation (VAME) (e.g., Kendall & Carterette, 1993a, 1993b). Whereas with the semantic differential each sound is rated along scales whose endpoints are labeled by two opposing verbal attributes such as "bright-dull," with the VAME method the endpoints of the scales are labeled by an attribute and its negation ("not harsh-harsh"). These multidimensional data are then analyzed with dimension reduction techniques such as principal components analysis (PCA; Kendall, Carterette, & Hajda, 1999; Lokki et al., 2011; von Bismarck, 1974a) or factor analysis (FA; e.g. Alluri & Toiviainen, 2010), and by cluster analysis techniques (Disley, Howard, & Hunt, 2006; Kendall & Carterette, 1993a) in order to achieve the reduction of a large number of semantic descriptions to a smaller number of interpretable factors.

One of the most cited studies on verbal description of timbre was conducted by von Bismarck (1974a, 1974b) in German. He performed a semantic differential listening test featuring 30 verbal scales in order to rate the verbal attributes of 35 steady-state synthetic tones. The four dimensions identified by von Bismarck were labeled: full-empty, dull-sharp, colorful-colorless and compact-diffused. Throughout this document '-' will be used to indicate antonyms and '/' to indicate synonyms. Other related studies have also identified three or four semantic axes. Working with simple synthetic tones and English adjectives, Pratt and Doak (1976) proposed a 3D space featuring the dimensions: bright-dull, warm-cold, and rich-pure. Štěpánek's (2006) study in Czech and German revealed the following dimensions for violin and pipe organ sounds: gloomy-clear, harsh-delicate, full-narrow, and noisy/rustle. Moravec and Štěpánek's work (2003), also in Czech, acquired descriptors through a questionnaire for timbre description without the presentation of any stimuli. It also identified four semantic dimensions, namely: bright/clear-gloomy/dark, hard/sharp-delicate/soft, wide-narrow, and hot/hearty. Finally, Disley's (2006) study in English used strings, brass, woodwind and percussive stimuli from the MUMS sound library (Opolko & Wapnick, 2006) and uncovered four salient dimensions: bright/thin/harsh-dull/warm/gentle, pure/percussive-nasal, metallic-wooden and evolving.

The inconsistencies observed in the above studies could be potentially attributed to factors related to method, stimuli, level of musical background, or language. Štěpánek (2006) has proposed that semantic dimensions of timbre are dependent on pitch and instrument type, Krumhansl and Iverson (1992) have concluded that pitch and timbre are not perceived independently, and Handel and Erickson (2004) showed that pitch differences can confuse instrument identification. Additionally, Marozeau & de Cheveigné (2007) found that auditory brightness (as predicted by the spectral centroid) is affected by the fundamental frequency, a fact that was additionally supported by Schubert and Wolfe (2006) through a semantic description listening test. The above imply that the variety of stimuli and pitches used in the different studies, as well as the musical background of the participants, could be responsible for the diversity in identified semantic dimensions. Furthermore, the data acquisition (selection and number of verbal descriptors) and analysis approaches (PCA, FA, etc.) also varied among the aforementioned studies. Finally, language is another potential factor of influence on timbre semantics. It has been argued that people's thinking about objects (i.e., object description, memory of objects, gender assignment, etc.) is affected by grammatical differences across languages (Boroditsky, Schmidt, & Phillips, 2003). Additionally, it has been reported that the use of some descriptive adjectives differs even between UK and US English speakers (e.g., Disley & Howard, 2004). Therefore, more solid conclusions regarding the influence of language on semantic descriptions of timbre will require careful control of several factors.

This work constitutes an inter-language study that seeks to:

1) investigate a potential influence of language on timbre semantic description while making an informed choice regarding stimuli and method. More specifically, we argue in favor of the intuitive assumption that semantics of musical timbre must share some common ground across different languages. To this end, a large-scale listening experiment was designed and conducted so as to allow direct comparison between two different linguistic groups, one English and one Greek speaking. A combination of continuant and impulsive stimuli of both acoustic and synthetic nature that also varied in fundamental frequency (F_0) were rated through VAME in order to reach generalizable conclusions regarding timbre semantics;
2) investigate the existence of nonlinear relationships between the examined verbal attributes through the use of CATegorical PCA (CATPCA) optimal transformation. Factor analysis assumes linear relationships between the variables under study. This, however, is not always guaranteed to be the case when analyzing semantic variables. Accounting for such nonlinearities can enhance the robustness of the resulting semantic space;
3) identify the acoustic correlates of the salient semantic dimensions. This is also a matter of ambiguity among various studies (e.g., Disley et al., 2006; Ethington & Punch, 1994; Faure et al., 1996; von Bismarck, 1974b). The association of timbre semantics with certain physical characteristics of sound is highly desirable as it contributes towards a better understanding of timbre perception and facilitates the development of intuitive sound processing applications.

In the following section the experimental and analytical methods are described. Next, results are given for the within-linguistic group analyses. Semantic dimensions and the resulting timbre spaces are identified and discussed for each group. An inter-linguistic comparison and discussion follows. Finally, acoustic correlates are identified for the semantic dimensions corresponding to both groups. The paper concludes with a summary of the major findings.

Method

A listening test based on a modification of the verbal attribute magnitude estimation (VAME) method was designed and conducted. VAME was preferred for the purpose of this study because, unlike the semantic differential, it reduces potential biases associated with assumptions concerning synonym and antonym relationships between the verbal labels for the rating scales. As a trade off, VAME requires double the number of verbal variables for the same number of adjectives in comparison to the semantic differential.

The listeners were provided with a preselected list of 30 verbal descriptors (in their native language) and were asked to describe the timbral attributes of 23 sound stimuli by choosing the adjectives they believed were most salient for each stimulus. No limit was imposed on the number of adjectives that could be used by each participant for each description. The verbal descriptors provided were intended for the description of sound impression (Wake & Asahi, 1998) and were selected among adjectives that are commonly found in the musical timbre perception literature (Disley et al., 2006; Ethington & Punch, 1994; Faure et al., 1996; von Bismarck, 1974a, 1974b;). The collection of terms is given in Table 1. Once a listener chose a descriptor, he or she was asked to estimate a value that corresponded to the sound on a scale anchored by the full extent of the verbal attribute and its negation, such as "not brilliant-very brilliant." This rating was input using a horizontal slider with a hidden continuous scale ranging from 0 to 100. A source of criticism regarding the provision of a predefined vocabulary is that the set of verbal attributes does not always correspond to descriptors that the participants would choose spontaneously (Donnadieu, 2007). To alleviate such issues, we allowed our listeners to freely propose up to three additional adjectives of their own choice to describe each stimulus.

STIMULI AND APPARATUS

Aiming to promote ecological validity, a set of 23 sounds drawn from commonly used acoustic instruments, electric instruments, and synthesizers and with fundamental frequencies varying across three octaves was selected. The following 14 instrument tones came from the MUMS (McGill University Master Samples) library: *violin, sitar, trumpet, clarinet, piano* 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

TABLE 1. *Spearman Correlation Coefficients Between the 30 Equivalent Semantic Variables (Descriptors) of the Two Languages*

Descriptor	Correlation	Descriptor	Correlation
Brilliant (Λαμπερός)	.77**	Sharp (Οξύς)	.67**
Hollow (Υπόκωφος)	−.08	Rich (Πλούσιος)	.37
Clear (Καθαρός)	.54**	Bright (Φωτεινός)	.80**
Rough (Τραχύς)	.82**	Dense (Πυκνός)	.80**
Metallic (Μεταλλικός)	.81**	Full (Γεμάτος)	.70**
Warm (Ζεστός)	.73**	Nasal (Ένρινος)	.73**
Smooth (Μαλακός)	.85**	Soft (Απαλός)	.62**
Thick (Παχύς)	.80**	Dark (Σκοτεινός)	.60**
Rounded (Στρογγυλεμένος)	.86**	Compact (Συμπαγής)	.02
Harsh (Σκληρός)	.82**	Dirty (Βρώμικος)	.77**
Dull (Θαμπός)	.40	Empty (Άδειος)	.02
Thin (Λεπτός)	.78**	Messy (Τσαλακωμένος)	.52*
Shrill (Διαπεραστικός)	.85**	Light (Ελαφρύς)	.67**
Cold (Ψυχρός)	.50*	Dry (Ξερός)	.61**
Distinct (Ευδιάκριτος)	.52*	Deep (Βαθύς)	.85**

Note: $^{*}p < .05$, $^{**}p < .01$. The Greek equivalent terms as translated by a linguist appear in parentheses.

used along with a set of eight 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.

Musical timbre studies usually restrict the sound stimuli to a fixed fundamental frequency (F_0). The reason why we have chosen to relax this restriction was to stimulate a wider range of verbal descriptions, to enhance generalization of the findings and to also investigate the influence of F_0 on the semantic dimensions of musical timbre. Marozeau, de Cheveigné, McAdams, and Winsberg (2003) and Marozeau and de Cheveigné (2007) have investigated this influence as well. Furthermore, Alluri and Toiviainen (2010, 2012) have shown that listeners can consistently rate short musical excerpts of varying key and rhythm on semantic scales. Since the task of this experiment was the assignment of a value of a semantic descriptor rather than a strictly controlled pairwise comparison, the stimuli were not required to be of equal duration either. Durations ranged from 3 to 8 seconds depending on the nature of the instrument (continuant or impulsive). Nevertheless, sound samples were equalized in loudness in an informal listening test within the research team. The RMS playback level was set between 65 and 75 dB SPL (A-weighted). Eighty three percent (83%) of the Greek participants found that level comfortable for all stimuli and 78% reported that loudness was perceived as being constant across stimuli. For the English participants these values were 93% and 85%, respectively.

The listening test was conducted 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

A first linguistic group consisting of 41 native Greek speakers (age range = 19-55, mean age = 23.3, 13 male) and a second one consisting of 41 native UK English speakers (aged 17-61, mean age 29.6, 28 male) participated in the listening test. None of the listeners reported any hearing loss and they had been practicing music for 13.5 (Greek) and 18.8 (English) years on average, ranging from 5 to 35 (Greek) and from 4 to 45 (English). There was also a prerequisite that participants did not have sound related synaesthesia or absolute pitch, as such a condition could affect the results due to pitch variation within the stimulus set. Participants were students of the Department of Music Studies of the Aristotle University of Thessaloniki, researchers from the Centre for Digital Music at Queen Mary University of London, and students of the Royal College of Music and of the Music Department of Middlesex University in London.

PROCEDURE

Listeners became familiar with the timbral range of the experiment during an initial presentation of the stimulus set (random order). On each trial of the experimental phase, participants were presented with one sound stimulus. They could listen to it as many times as required before submitting their ratings. The sounds were presented in random order and listeners were advised to use as many of the provided terms as they felt were necessary for an accurate description of each different timbre, and also to take a break when they felt signs of fatigue. They were also offered the option to withdraw at any point. The overall listening test procedure, including instructions and breaks, lasted approximately 45 min.

CLUSTER ANALYSIS, FACTOR ANALYSIS, AND CATPCA TRANSFORMATION

We applied two statistical analysis techniques in order to reach conclusions regarding the salient semantic dimensions of timbre. Cluster analysis (Romesburg, 2004) indicated groups of semantically related verbal descriptors while Factor analysis (FA; Harman, 1976) uncovered the latent structure of our inter-correlated semantic variables.

Factor analysis was preferred over principal components analysis (PCA) for this study because it aims to explain the relationships between a set of variables by modeling their correlations. PCA, on the other hand, only achieves data reduction through maximization of the variance explained by the principal components (Fabrigar et al., 1999). FA achieves this by treating each measured variable as a linear combination of one or more common factors and one unique factor, while PCA does not differentiate between common and unique variance.

As already mentioned, an important element of the analysis in this study is the fact that it allowed for the possible existence of nonlinear relationships between the measured verbal attributes. That is, within the framework of factor analysis, we relaxed the constraint on strict linear relations between variables by anticipating necessary optimal transformations of the original variables along with data reduction. For this reason, we employed a hybrid approach of optimal transformation of variables (offered by a readily available technique, CATPCA) followed by a typical factor analysis of the transformed variables.

Analysis and Results

The magnitude ratings for each verbal descriptor and each musical timbre were averaged across the 41 participants in each of the language groups. Thirty-seven percent (37%) of the Greek participants inserted one or more extra verbal descriptor thus providing 31 additional terms. However, only eight of these terms where mentioned more than once, and only six were mentioned by more than one participant. Sixty-six percent (66%) of English participants used at least one extra term, thus providing 117 additional verbal descriptors. Thirty-three (33) of these terms were inserted more than once, and 27 were used by more than one participant. The extra terms are presented in Appendix C and discussed in subsection: *Interlinguistic relationships*.

STATISTICAL ANALYSIS

The analytic strategy used to reduce the large number of variables (30) was structured according to three basic steps. In the initial step, a hierarchical cluster analysis (centroid linkage) based on squared Euclidean distances was employed in order to reveal the major clusters and outliers among the adjectives. A preliminary factor analysis was then performed within each of the clusters to identify its salient adjectives and discard the rest. A final factor analysis on this reduced set of salient adjectives resulted in the major factors (i.e., semantic dimensions). For more details see Appendix A.

Nonlinear transformation of the variables. Preliminary analyses of the English group data showed that a simple rank ordering transformation resulted in a tighter clustering of the adjectives and explained a larger amount of variance with fewer dimensions compared to the untransformed case. This may be due to the presence of nonlinearities among the perceptual variables that have been more efficiently modeled by the non-metric approach. Based on this finding, here we apply an optimal spline ordinal transformation performed by the CATPCA module of the SPSS suite. This transformation has additionally contributed to addressing issues with strongly skewed data. Figure 1 shows two indicative nonlinear transformation plots obtained by the CATPCA optimization as an example of the shape of the transformations applied to the variables. The optimal nonlinear transformation has contributed to a more compact representation of the semantic variables (i.e., tighter clustering), which allowed our subsequent data reduction strategy. Additionally, FA on the transformed variables explained a higher amount of total variance, which was also concentrated on the first two factors compared to the untransformed case. This suggests that the transformation has accounted for existing nonlinearities between the variables and has yielded a more accurate representation of the semantic space. For details regarding the advantages of the nonlinear transformation see Appendix B.

INTRALINGUISTIC SEMANTIC DIMENSIONS

The transformed variables analyzed with the maximum likelihood algorithm resulted in a three-factor solution (eigenvalues ≥ 1) that explained the same amount of total variance (82%) in both linguistic groups (see Table B1 in Appendix B). Specifically for the Greek group, the

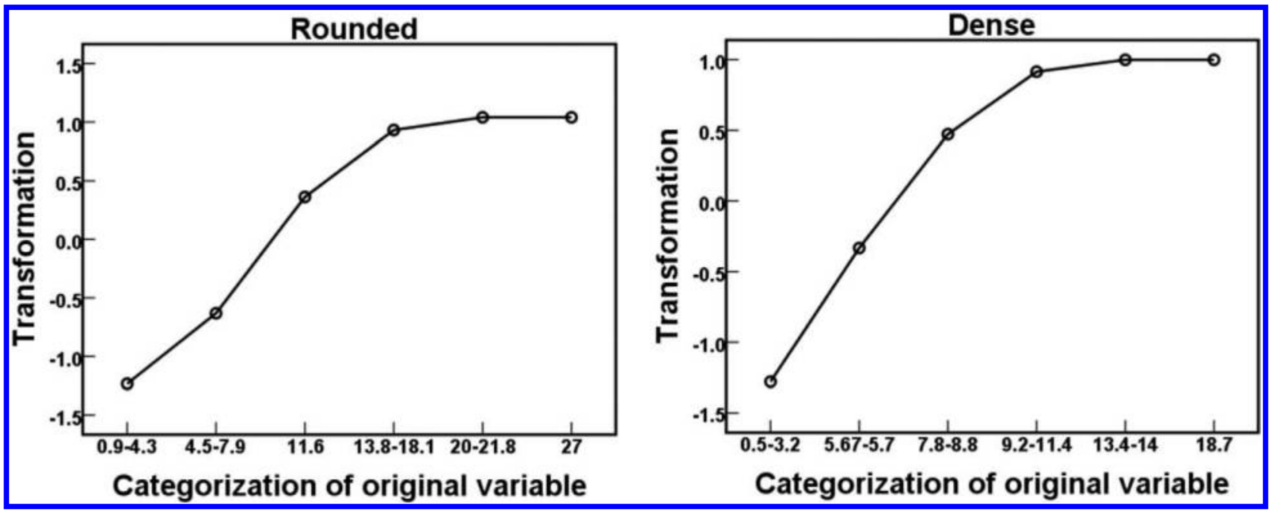


FIGURE 1. Indicative optimal nonlinear transformations of original variables. Rounded (Greek) on the left and Dense (English) on the right. The abscissa represents the categories in which the variable is separated (in this case six) and the ordinate represents the value that is assigned to each category by the algorithm.

first two factors explained a similar amount of variance (36.5% and 30.5%), while the third only explained 15% of the variance. For the English group almost half of the variance (48.7%) was contained in the first factor, while the second factor explained 27.3%, and the third factor only 5.9% of the total variance prior to rotation.

The emerging factors in FA are often computed as mutually orthogonal (Disley et al., 2006). Subsequently, they are subjected to a rotation to improve the interpretability of the solution by maximizing the already large factor loadings and minimizing the small ones. However, in several cases, the orthogonality of the factors constitutes a strict condition and therefore can impede the interpretability of the results. Consequently, we chose to relax the requirement of factor orthogonality by employing a non-orthogonal (oblique) rotation of the initial orthogonal solution, which allows for factors to be correlated. We have used the direct oblimin method, which (among others) is considered as a viable approach to the problem of oblique factor rotation (Harman, 1976).

The data reduction methodology gave the most representative verbal descriptors for this set of sounds. These adjectives, along with their factor loadings, appear in Table 2 for both Greek and English groups. Factor loadings are the regression coefficients (ranging from -1 to $+1$) between variables and factors. Their values indicate the relative contribution that a variable makes to a factor and are crucial for the labeling and interpretation of the factors. Only descriptors with factor loadings $\geq .75$ were considered significant in this work and will be used for factor interpretation (Comrey & Lee, 1992; Tabachnick & Fidell, 2001). Based on the above, a proposed labeling was applied by choosing a couple of terms that we believed would better capture the essence of each semantic dimension. According to this, factor 1 could be: *Depth-Brilliance* for Greek and *Brilliance/Sharpness* for English, factor 2: *Roundness-Harshness* for Greek and *Roughness/Harshness* for English, and factor 3: *Richness/Fullness* for Greek and *Thickness-Lightness* for English.

The correlation coefficients between the rotated factors together with the corresponding angles ($angle = cos^{-1}(r)$) are shown in Table 3. The very low correlation coefficients between factors for the Greek group imply the existence of a nearly orthogonal semantic space. However, for the English group, there appears to be a mild correlation between the first and the third (121.4°) and also between the first and the second dimensions (72.8°).

Figure 2 shows the positions of the stimuli in the common factor space based on the factor scores. The presentation consists of six 2D planes resulting from the 3D Euclidean semantic timbre spaces (although dimensions are not entirely orthogonal) for both Greek and English groups. The Euclidean representation is less accurate for the English group due to its higher inter-dimensional correlation. The different symbols for each sound represent classes of musical instruments according to von Hornbostel and Sachs (1914), and the filling of the symbols represents the type of excitation (black for continuant sounds and white for impulsive sounds).

As can be noticed by visual inspection of Figure 2, the musical sounds' position within the common factor

TABLE 2. *Pattern Matrix of the Greek and English Factor Loadings with Suggested Labeling After Oblimin Rotation.*

	Factors					
	Greek			*English*		
	1 (Depth-Brill.)	*2 (Round.-Harsh.)*	*3 (Rich./Full.)*	*1 (Brill./Sharp.)*	*2 (Rough./Harsh.)*	*3 (Thick.-Light.)*
Brilliant	*−.82**	*.19*	*.25*	*.99**	*−.22*	*−.01*
Deep	*.91**	*.23*	*.13*	*−.16*	*−.22*	*.74*
Rough	–	–	–	*−.27*	*.96**	*.08*
Soft	*−.38*	*.86**	*−.09*	*−.49*	*−.68*	*−.19*
Full	*.18*	*−.02*	*.84**	–	–	–
Rich	*−.32*	*.12*	*.97**	–	–	–
Harsh	*.00*	*-.93**	*−.18*	*.41*	*.77**	*−.02*
Rounded	*.12*	*.88**	*.20*	–	–	–
Thick	*.79**	*.16*	*.36*	*−.02*	*−.12*	*.93**
Thin	–	–	–	*.23*	*.44*	*−.65*
Warm	*.11*	*.91**	*.19*	*−.48*	*−.57*	*.22*
Dark	–	–	–	*−.38*	*.24*	*.70*
Sharp	*−.49*	*−.62*	*.13*	*.78**	*.06*	*−.04*
Messy	–	–	–	*−.23*	*.88**	*.20*
Light	*−.41*	*.74*	*−.43*	*−.20*	*−.21*	*-.89**
Shrill	*−.30*	*−.74*	*.14*	*.43*	*.42*	*−.31*
Dense	*.62*	*−.08*	*.54*	*−.02*	*−.29*	*.83**
Dull	*.62*	*.49*	*−.09*	*−.37*	*−.54*	*.25*
Bright	–	–	–	*.69*	*−.02*	*−.35*

Note: *Loadings $\geq$ .75.

TABLE 3. *Inter-dimension Correlations and Angles.*

Correlation coefficient	*Greek*	*English*
r12	*.14 (82.2°)*	*.30 (72.8°)*
r23	*−.01 (90.6°)*	*.07 (86.1°)*
r31	*.16 (80.7°)*	*−.52 (121.4°)*

space (factor scores) does not provide any clear indication of possible favored relations between the identified timbral descriptions (factor labels) and the traditionally accepted classification schemes of musical instruments. As expected, our findings further support the difficulty to identify a direct relation of musical timbre description with terms referring to broad categories of musical instruments' sounds (Campbell, Greated, & Myers, 2006).

INTERLINGUISTIC RELATIONSHIPS

Table 1 presents the Spearman correlation coefficients that indicate the agreement on the use of each adjective between the two different linguistic groups. Interestingly, most of the adjectives that feature a poor intergroup correlation (e.g., *compact*, *empty*, *hollow*, *distinct*, and *cold*) are also weakly correlated with the other adjectives within the linguistic groups. This is evident from the dendrograms B1a and B1b (Appendix B) and has resulted in the removal of most of them during the data reduction phase.

A correlation analysis was subsequently performed between the semantic dimensions. The Spearman correlation coefficient between first dimensions is $\rho(21) = -.66$, $p < .01$, between the second dimensions is $\rho(21) = -.78$, $p < .001$, and between the third dimensions is $\rho(21) = .55$, $p < .01$. Figure 3 demonstrates the above by showing the scatter plots for each corresponding dimension between the two languages. While the third dimensions are only mildly correlated, the third English dimension is highly correlated with the first Greek dimension, $\rho(21) = .81$, $p < .001$, and the first English dimension shows some correlation with the second Greek dimension, $\rho(21) = -.46$, $p < .05$. This shows that the terms *thickness* and *sharpness,* which are included in these different dimensions, are nevertheless commonly understood between the two linguistic groups. *Sharpness* as a synonym for *brilliance* also links this dimension with Greek *roundness-harshness*, and *thickness* strongly links the first Greek with the third English dimension. This is in agreement with the strong interlinguistic correlations for *sharpness* and *thickness* that are evident in Table 1. The correlations featured across the remaining non-corresponding dimensions were non-significant ($p > .05$).

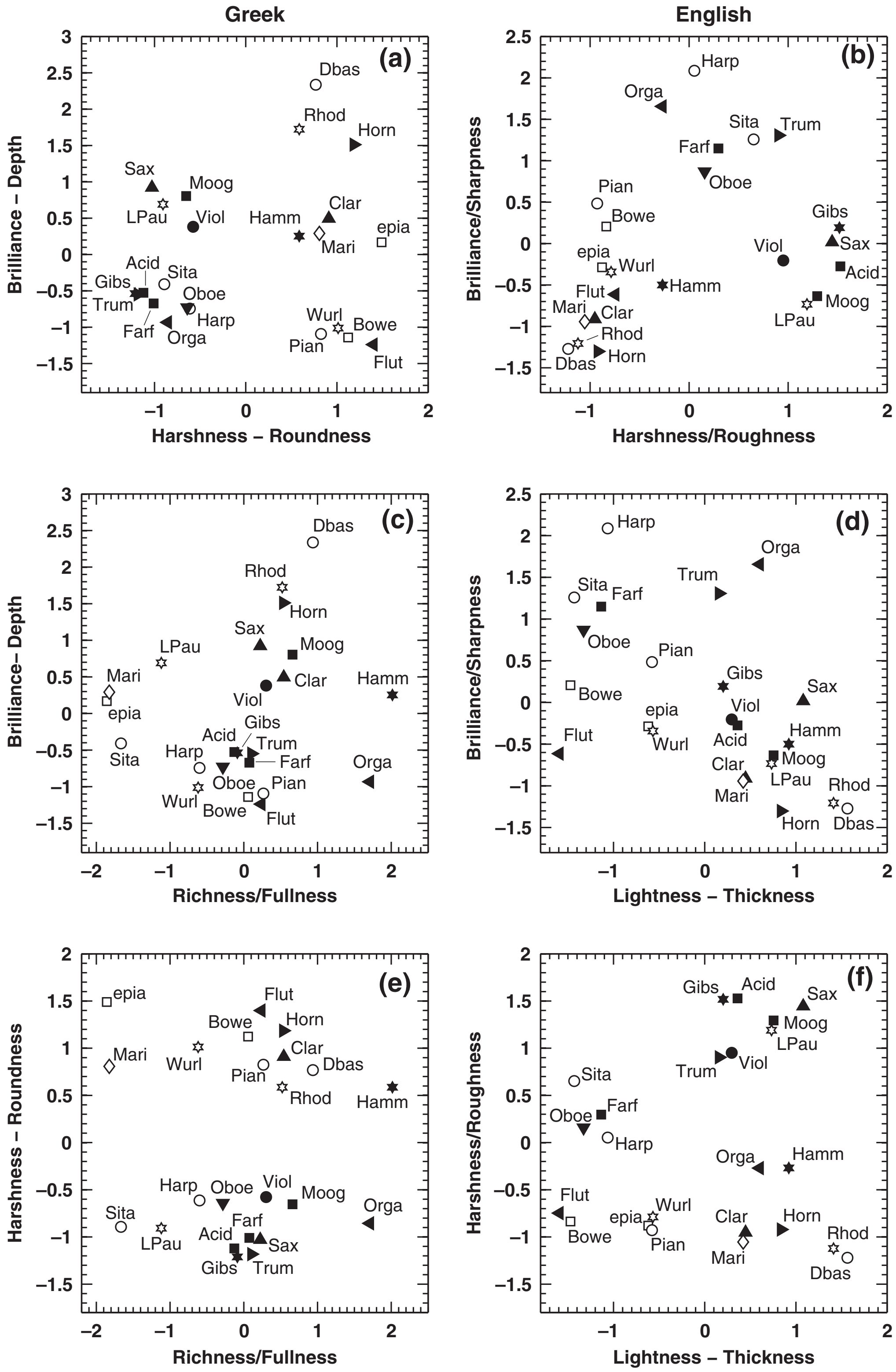


FIGURE 2. Six 2D planes of the Greek (left) and the English (right) 3D semantic timbre spaces. Black symbols: Continuant, white symbols: Impulsive, ▲: Single reed, ▼: Double reed, ◀: Aerophone, ▶: Lip reed, ●: Chordophone, ♦: Idiophone, ⋆: Electrophone, ■: Synthesizer

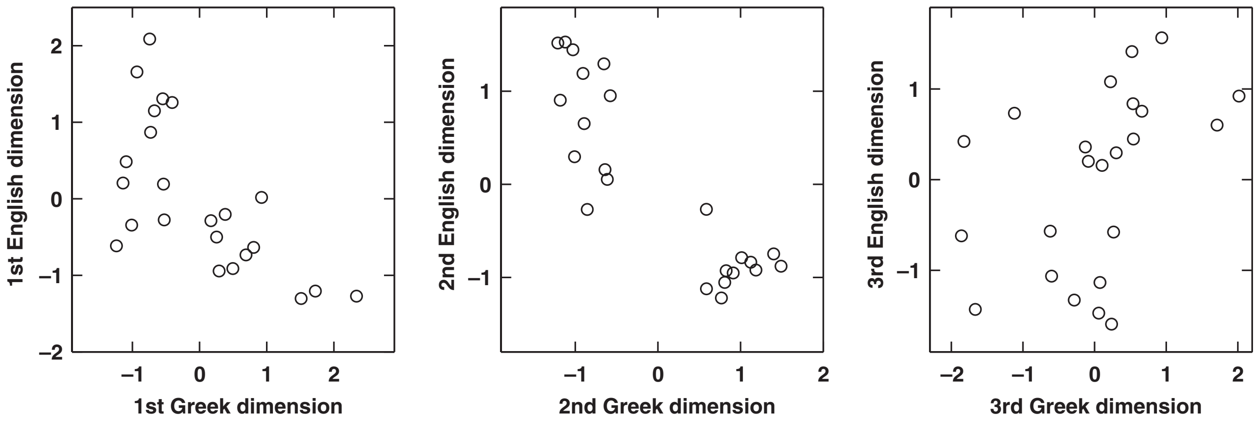


FIGURE 3. The scatter plots of the Greek and English semantic dimensions show that the 23 stimuli are similarly perceived on the corresponding dimensions. As expected from the correlation analysis, the relationship is stronger for the second dimensions and weaker for the third dimensions.

These results show that the three semantic dimensions feature an inherent one-to-one (first Greek-first English, second Greek-second English, third Greek-third English) conceptual relatedness (some differences between specific adjectives will be reported shortly). A two-sample Kolmogorov-Smirnov test showed no significant effect of language for any dimension (z = .15, p = 1.00 between first dimensions, z = .44, p = .99 between second dimensions and z = .59, p = .88 between third dimensions). The K-S test was preferred as several dimensions in each language group failed a Shapiro-Wilk normality test ($p < .05$).

Despite the evident similarities between the semantic spaces of the two linguistic populations, there are some differences that are also worth mentioning. The main difference concerns the terms loaded on the *brilliance* dimension for each language. The adjective *sharp* is grouped with *brilliant* in the English group but associated with *harsh* in the Greek group. This is evident both from inspection of Figure B1 and from Table 2. Additionally, it seems that *full* and *rich* form a separate group in the Greek population, whereas the same terms are more closely related to *thick*, *dense*, *deep*, etc. in the English population (see Figure B1). As a result, *rich* and *full* form a separate factor for Greek, but *thick* and *deep* load as opposites on the *brilliance* factor. The above produce a *brilliance* dimension that is enriched with unrelated terms for each of the two linguistic groups.

The extra terms provided by the listeners (see Table C1, Appendix C) generally fall into seven conceptual categories for both populations: 1) *properties of source* (wooden, glassy, synthetic, etc.), 2) *temporal evolution* (static, energetic, constant, etc.), 3) *emotional terms* (sinister, oppressive, suave, etc.), 4) *technical terms* (spectral, phasey, sinewave, etc.), 5) *sense of sight* (blurred, smoky, transparent, etc.), 6) *sense of touch* (raspy, gentle, blunt, etc.), and 7) *size of object* (large, majestic, heavy, etc.). These categories appeared to be more evident in the English group because of the larger number of extra terms given (117 extra terms in English compared to the 31 extra terms in Greek). The lack of terms in the last three categories can be explained by the fact that they were already well represented in the provided adjectives. The three largest categories in both linguistic groups were *properties of source*, *temporal evolution,* and *emotional terms*. The only predefined descriptor belonging to one of these three categories was *metallic.*

Discussion

The analysis presented in the previous section has identified three semantic dimensions that explain more than 80% of the variance in the descriptive data. These dimensions show high independence for the Greek group while the interdimensional correlation is moderate between some dimensions for the English participants.

The application of an optimal nonlinear transformation supported the existence of nonlinearities by providing a more compact representation of the data and explaining more variance in the first two dimensions for both groups. It can be argued that the transformation did not affect the qualitative interpretation of the semantic dimensions. However, the value of this approach lies in the output of a more accurate representation of the sound stimuli positions within the identified semantic timbre space. This is particularly significant for the search of acoustic correlates and for investigating the association of semantic with perceptual spaces.

As mentioned in the introduction, there exists evidence that language affects the way people think about objects.

Contrary to this, our work was partly motivated by an intuitive assumption that timbre semantics could feature a general agreement across languages. Although this assumption was not subjected to a thorough hypothesis-inference scrutiny (which would require careful control of several additional parameters and factors), we demonstrated that the three pairs of semantic dimensions for the two linguistic groups share common conceptual properties. This exploratory approach, supported by some preliminary inferential tests (K-S and Spearman correlation), provides strong indication that despite the differences in the use of individual descriptors, there exists a common semantic space for timbre between these two languages (at least for this stimulus set). In addition, it justifies further investigation of hypotheses regarding the universality of timbre semantics.

Therefore, we will propose an empirical labeling to express the common concept for each of the semantic dimensions. The dimension that shows the strongest agreement between the two groups is the one that describes whether a sound is perceived as smooth-and-round or rough-and-harsh. As these adjectives originate from tactile quality description we suggest the label *texture* for this dimension. The first dimensions for both linguistic groups have the adjective *brilliant* in common. This is a metaphor that comes from the domain of vision, we therefore suggest the label *luminance* for the description of this dimension. Finally, the third dimensions in both groups describe whether a sound is perceived as thick-dense-rich-and-full or light. We suggest *mass* as an appropriate general semantic label for this dimension.

These results seem to support Lichte (1941), who concluded that: " . . . complex tones have, in addition to pitch and loudness, at least three attributes. These are brightness, roughness, and one tentatively labeled fullness. The first two are probably more basic than the third" (abstract). There also seems to be some agreement regarding the number and naming of dimensions with some earlier studies (Alluri & Toiviainen, 2010; Moravec & Štěpánek, 2003; Pratt & Doak, 1976; Štěpánek, 2006; von Bismarck, 1974a). Taken as a whole, there appears evidence of language-independent verbalization of timbre descriptions.

In agreement with these studies, the boundaries between semantic dimensions are not always clearly defined. *Luminance* and *mass* dimensions are correlated with each other, particularly for the English group. Sounds that are described as *light* are more likely to also be described as *brilliant*, while sounds described as *thick* or *dense* are also described as *less brilliant*. Additionally, we provide some evidence that *luminance* is conceptually related to *texture* in the English language as suggested by the fact that *sharpness* (a term that is positioned in the *texture* cluster in Greek dendrograms B1a and B1c) is highly loaded (.778) on the *luminance* dimension. This last finding is not unexpected as Štěpánek (2006) has supported that *sharpness* is an auditory attribute that lies between *luminance* and *texture* (i.e., a sound object featuring both high *luminance* and high *texture* is described as *sharp*). However, the interpretation of specific differences (mainly some unrelated terms loaded on the *luminance* dimension) between the semantic dimensions of the two language populations would require a linguistic analysis which, although interesting per se, lies beyond the scope of this study.

The acquisition of extra terms from spontaneous descriptions suggests that future researchers on timbre semantics should consider including terms that belong to one additional semantic category: *temporal evolution*. Although the number of terms acquired for description of the *properties of source* and *emotions* is also considerably large, they should probably be avoided when studying the semantic description of sound impression (Wake & Asahi, 1998).

Finally, while it has been shown that same-family instruments tend to occupy similar regions in perceptual spaces resulting from pairwise dissimilarity ratings (Giordano & McAdams, 2010), this can not be supported by the semantic space structure of this work. As a possible explanation, it can be assumed that while perceptual spaces resulting from cognitive dissimilarity ratings and MDS analyses represent both sensory and semantically meaningful factors, verbal attribute studies can only capture the semantically charged portion of the MDS spaces. Consequently, the comparison of these semantic spaces with perceptual spaces resulting from a pairwise dissimilarity experiment using the same stimuli could be proven useful in testing the above hypothesis.

Acoustic Correlates of Semantic Dimensions

A large set of low-level features (see Table D1, Appendix D) was extracted from the experimental sound set as an initial attempt to identify acoustic correlates for the semantic dimensions that resulted from factor analysis. 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 = 44.1kHz) with an overlapping factor of 87.5%, the

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

	Component			
	1 *(Energy distribution of harmonic partials)*	*2 (Spectrotemporal)*	*3 (Spectrotemporal, Inharmonicity)*	*4 (Temporal, Spectrotemporal)*
SC_norm	.96*	−.03	.17	−.01
T3	.93*	−.13	.11	.05
SC_loud_cor	.88*	−.25	−.32	.06
SC_loud	.79*	−.20	−.49	.05
Spread	.79*	−.11	−.42	−.17
T2	−.73	.07	−.47	.20
Noisiness	.05	.91*	.25	−.21
Flux	−.20	.88*	.06	−.02
SC_std	−.34	.82*	.18	−.40
SC_var_loud	−.14	.39	.79*	−.13
Inharmonicity	.27	.30	.79*	−.14
OER	−.38	−.41	.65	−.34
Log_At_time	.01	.06	−.24	.83*
MCV	−.22	−.45	−.02	.76*
TC_norm	.15	−.57	−.21	.58

Note: *Loadings $\geq$ .75 are used for labeling the components. See Table D1 for the abbreviations.

zero padding factor was two 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 et al., 1997) instead of the amplitude of the harmonics or the FFT bins. Signals were gated so that regions with a low signal-to-noise ratio (SNR) were not included in the feature extraction.[1]

The problem of strongly correlated clusters of acoustic features needed to be addressed before proceeding to correlation analysis with the semantic dimensions. One approach would be to consider an acoustic feature as significantly associated with a dependent variable only when both their correlation and partial correlation were significant (Giordano, McAdams, Zatorre, Kriegeskorte, & Belin, 2012). However, while this approach avoids data reduction methods, it discards variance that is common between features. Thus, an exploitation of principal components analysis was favored similarly to Alluri and Toiviainen (2010), Giordano, Rocchesso, and McAdams (2010), and Peeters et al. (2011). To reduce high multicollinearity within the variable (feature) set, we initially inspected the Spearman coefficient correlation matrix and discarded strongly correlated features, $\rho(21) \geq .80$. We then rank-ordered the features and applied PCA to the reduced data set. Inspection of the anti-image correlation matrix[2] diagonal led to further removal of features whose individual Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) was less than .5 so as to achieve an acceptable overall KMO. The final solution consisted of four components (KMO = .67, Bartlett's test of sphericity $p < .001$) that explained 83.3% of the total variance. Table 4 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.

Features like the *normalized harmonic spectral centroid* (SC_norm), *tristimulus 3* (T3) (Pollard & Jansson, 1982), *SC_loud_cor* (corrected version of the spectral centroid in order to remove the influence of F_0, for an example, see Marozeau & de Cheveigné, 2007) and *harmonic spectral spread* (Spread) all represent spectral structure (i.e. distribution of energy among harmonic partials) rather than spectral content. Therefore, the first component is labeled: *energy distribution of harmonic partials.* The second component is related to *spectrotemporal* characteristics such as *noisiness*, *harmonic spectral flux* (Flux),

---

[1] In order to avoid the effect of the low signal-to-noise ratio (SNR) in the tail of the release (especially for percussive sounds) on the feature calculation, we cropped all our sounds to the point where the SNR dropped below 25 dB. The energy of the noise was calculated as the average energy of the last 10 frames of the signal (window: 1024, hop size: 128). Moreover, the sounds were also cropped in the beginning at the point where the SNR was above 1 dB so as to discard the initial silent gap before the onset. Special attention has been paid to avoid the introduction of any artifacts from this processing procedure.

[2] 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 5. *Spearman Correlation Coefficients Between Semantic Dimensions, the 4 Principal Components of the Audio Feature Set and F_0.*

		Energy distribution of harmonic partials	Spectrotemporal	Sectrotemporal, Inharmonicity	Temporal, Spectrotemporal	F_0
Greek	Depth/Thickness-Brilliance	$-.12$	$-.26$	$.68^{***}$	$.15$	$-.58^{**}$
	Roundness-Harshness	$-.75^{***}$	$.11$	$-.18$	$.04$	$.44^{*}$
	Richness/Fullness	$-.03$	$-.19$	$.03$	$.44^{*}$	$-.23$
English	Brilliance/Sharpness	$.62^{**}$	$.20$	$-.50^{*}$	$.07$	$.28$
	Harshness/Roughness	$.74^{***}$	$-.13$	$.01$	$-.04$	$-.18$
	Thickness-Lightness	$-.08$	$-.18$	$.70^{***}$	$.22$	$-.76^{***}$

and the *standard deviation of the harmonic spectral centroid* (SC_std). The third component is represented by both *spectral centroid variation* (SC_var_loud) calculated from Moore's specific loudness and *inharmonicity*. Finally, the fourth component is related to a temporal characteristic like *the logarithm of the attack time* (Log_At_time) and a spectrotemporal one like the temporal variation of the first nine harmonics (*Mean coefficient of variation,* MCV, Kendall & Carterette, 1993b).

Table 5 presents the Spearman correlation coefficients between the mutually orthogonal components and the semantic dimensions (factor scores) for both linguistic groups. F_0 has been also considered in the correlation analysis in order to reveal its potential influence on the semantic dimensions.

GREEK INTRAGROUP RESULTS

The *Luminance (Depth/Thickness-Brilliance)* dimension shows significant positive correlation, $\rho(21) = .68, p < .01$, with the 3rd principal component (SC variation and inharmonicity) and is also influenced by the fundamental frequency, $\rho(21) = -.58, p < .01$. The *Texture (Roundness-Harshness)* dimension shows a strong negative correlation, $\rho(21) = -.75$, $p < .001$, with the first component representing the energy distribution of harmonic partials. The *Mass (Richness/Fullness)* dimension does not exhibit strong correlations with any of the principal components.

ENGLISH INTRAGROUP RESULTS

The *Luminance (Brilliance/Sharpness)* dimension is correlated with the energy distribution of harmonic partials, $\rho(21) = .61, p < .01$, and is weakly correlated, $\rho(21) = -.50$, $p < .05$, with the third principal component (SC variation and inharmonicity). The *Texture (Harshness/Roughness)* dimension exhibits strong correlation, $\rho(21) = .74$, $p < .001$, with the energy distribution of harmonic partials. Finally, the *Mass (Thickness-Lightness)* dimension features strong correlation, $\rho(21) = .70$, $p < .001$, with the 3rd principal component (SC variation and inharmonicity) and is also heavily influenced by the fundamental frequency [$\rho(21) = -.76$, $p < .001$].

INTERLINGUISTIC COMPARISON AND DISCUSSION

The second part of this study examined possible relationships between the uncovered semantic dimensions and acoustic characteristics of the sound stimuli. The most important factor for the auditory perception of *texture* seems to be the *energy distribution of harmonic partials*. The correlations for both linguistic groups indicate that sounds with stronger high partials are more likely to be characterized as *rough* or *harsh* and the opposite as *round* or *soft*. This appears to support Faure et al. (1996), Howard and Tyrrell (1997), Barthet, Depalle, et al. (2010) and Barthet, Depalle, Kronland-Martinet, and Ystad (2011), who have generally associated higher spectral centroid values with *roughness* and *shrillness* and lower spectral centroid values with *softness*.

Luminance featured significant correlation with spectral structure only in the English group, but there is some evidence that the amount of inharmonicity influences auditory brilliance (i.e., more inharmonic sounds are perceived as less brilliant) in both groups. Additionally, sounds with a stronger spectral centroid fluctuation are also more likely to be perceived as less brilliant. There is some evidence that fundamental frequency is positively correlated with brilliance in the Greek group. The findings concerning *luminance* and *texture* seem to support Schubert and Wolfe (2006) whose empirical study has proposed that *simple SC* is a better correlate for perceptual brightness than the *normalized SC*. In other words, these results suggest that the distribution of energy, as expressed by the normalized SC, seems to be a better correlate of *texture*, whereas spectral content (also related with F_0) might predict *luminance* more efficiently.

Mass did not correlate significantly with any component in the Greek group. On the contrary, it exhibited two strong correlations in the English group. These correlations suggested that sounds with higher F_0 were perceived as lighter and also that auditory thickness and density increased with inharmonicity and with fluctuation of the spectral centroid. The latter is in some agreement with Terasawa's definition of density (Terasawa, 2009) as "the fluctuation of instantaneous intensity of a particular sound, both in terms of rapidity of change and degree of differentiation between sequential instantaneous intensities."

Overall, the combination of the Greek and English group findings suggest that *texture* is evidently affected by the energy distribution of the harmonic partials. The picture is not so clear for *luminance* and *mass* and future research on their acoustic correlates is mandated. However, there are indications that auditory thickness is enhanced by inharmonicity and SC fluctuation, whereas auditory brilliance is decreased. The influence of F_0 was more evident in the English group's perception of mass and less evident in the Greek group's perception of luminance, indicating that the effect of F_0 on timbre semantics needs to be further investigated.

Conclusion

This study investigated the underlying structure of musical timbre semantics through an analysis of verbal description of different timbres. Factor and cluster analyses were performed on semantic descriptors that were obtained from two linguistic groups (Greek and English) for musical instrument tones. The salient semantic dimensions for timbre description were identified and compared between the two linguistic groups. A correlation analysis between extracted acoustic descriptors and semantic dimensions indicated the prominent acoustic correlates. The major contributions of this work can be summarized as follows:

(1) The statistical analysis results suggested the existence of nonlinear relationships between the semantic variables. An optimal nonlinear transformation applied to the raw data accounted for such nonlinearities between the variables and resulted in a more efficient modeling of their underlying structure. This means that linear modeling of such data should be undertaken with care.

(2) While there did not seem to be consensus in the use of every descriptive adjective between the two linguistic groups (see Table 1), the three identified semantic dimensions exhibited a high degree of universality. These common semantic dimensions could be labeled as *luminance*, *texture,* and *mass*. This is an indication of language-independent description of musical timbre, at least between English and Greek.

(3) The strongest acoustic correlates identified for both linguistic groups were the following: i) the energy distribution of harmonic partials was associated with *texture*, ii) inharmonicity and variation of the SC were positively correlated with *thickness* and negatively correlated with *brilliance*, iii) F_0 affected English *mass* negatively and Greek *luminance* positively.

Future work should attempt to link this descriptive approach to other popular approaches in timbre perception research such as pairwise dissimilarity tests and MDS analysis. Thus, it will be possible to further examine the relationship between semantics and perception of musical timbre.

Author Note

Portions of this work were reported in Zacharakis et al. (2011, 2012).

The authors wish to thank Dr. Robin Hart together with Susan Sturrock for organizing the listening tests at the Royal College of Music, as well as Dr. John Dack for organizing the listening tests at Middlesex University. We would also like to express our gratitude to Prof. Stephen McAdams, Andy Simpson, Dr. Marcus Pearce, Dr. Mathieu Barthet, Dr. Michael Terrell, Prof. Georgios Papadelis, Dr. Bruno Giordano, Dr. Jeremy Marozeau, and one anonymous reviewer 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: a.zacharakis@qmul.ac.uk

References

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

→ Alluri, V., & Toiviainen, P. (2012). Effect of enculturation on the semantic and acoustic correlates of polyphonic timbre. *Music Perception*, *29*, 297-310.

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.

ANSI. (1973). *Psychoacoustical Terminology*, S3.20-1973, New York: American National Standards Institute.

Barthet, M., Depalle, P., Kronland-Martinet, R., & Ystad, S. (2010). Acoustical correlates of timbre and expressiveness in clarinet performance. *Music Perception*, *28*, 135-153.

Barthet, M., Depalle, P., Kronland-Martinet, R., & Ystad, S. (2011). Analysis-by-synthesis of timbre, timing, and dynamics in expressive clarinet performance. *Music Perception*, *28*, 265-278.

Barthet, M., Guillemain, P., Kronland-Martinet, R., & Ystad, S. (2010). From clarinet control to timbre perception. *Acta Acustica United with Acustica*, *96*, 678-689.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Genter & S. Goldin-Meadow (Eds.), *Language in mind*. Cambridge, MA: MIT Press.

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.

Campbell, M., Greated, C., & Myers, A. (2006). *Musical instruments: History, technology and performance of instruments of western music*. New York: Oxford University Press.

Comrey, A., & Lee, H. (1992). *A first course in factor analysis* (2nd ed.). New York: Psychology Press.

Disley, A. C., & Howard, D. M. (2004). Spectral correlates of timbral semantics relating to the pipe organ. *Speech, Music and Hearing*, *46*, 25-40.

Disley, A. C., Howard, D. M., & Hunt, A. D. (2006). Timbral description of musical instruments. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC 09)* (pp. 61-68), Bologna, Italy: Alma Mater Studorium.

Donnadieu, S. (2007). Mental representation of the timbre of complex sounds. In J. W. Beauchamp (Ed.), *Analysis, synthesis, and perception of musical sounds* (pp. 272-319). New York: Springer.

Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, *4*, 95-104.

Ethington, R., & Punch, B. (1994). Seawave: A system for musical timbre description. *Computer Music Journal*, *18*, 30-39.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

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.

Fritz, C., Blackwell, A. F., Cross, I., Woodhouse, J., & Moore, B. C. J. (2012). Exploring violin sound quality: Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties. *Journal of the Acoustical Society of America*, *131*, 783-794.

Giordano, B. L., & McAdams, S. (2010). Sound source mechanics and musical timbre perception: Evidence from previous studies. *Music Perception*, *28*, 157-170.

Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., & Belin, P. (2012). Abstract encoding of auditory objects in cortical activity patterns. Advance online publication. *Cerebral Cortex, doi:10.1093/cercor/bhs162*

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.

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

Grey, J. M. (1978). Timbre discrimination in musical patterns. *Journal of the Acoustical Society of America*, *64*, 467-472.

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.

Hajda, J. M. (1998). The effect of amplitude and centroid trajectories on the timbre of percussive and nonpercussive orchestral instruments. In P. K. Kuhl & L. A. Crum (Eds.), *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America* (Vol. 3, pp. 1887-1888). Seattle, WA: Acoustical Society of America.

Hajda, J. M. (1999). *The effect of time-variant acoustical properties on orchestral instrument timbres* (Unpublished doctoral dissertation). University of California, Los Angeles.

Hajda, J. M. (2007). The effect of dynamic acoustical features on musical timbre. In J. W. Beauchamp (Ed.), *Analysis, synthesis, and perception of musical sounds* (pp. 250-271). New York: Springer.

Handel, S., & Erickson, M. L. (2004). Sound source identification: The possible role of timbre transformations. *Music Perception*, *21*, 587-610.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago, IL: University of Chicago Press.

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

Howard, D., & Tyrrell, A. (1997). Psychoacoustically informed spectrography and timbre. *Organised Sound*, *2*, 65-76.

Hutcheson, G., & Sofroniou, G. (1999). *The multivariate social scientist*. London, UK: SAGE Publications Ltd.

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

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.

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.

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

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

Lokki, T., Pätynen, J., Kuusinen, A., Vertanen, H., & Tervo, S. (2011). Concert hall acoustics assessment with individually elicited attributes. *Journal of the Acoustical Society of America*, *130*, 835-849.

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., Beauchamp, J. W., & Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *Journal of the Acoustical Society of America*, *105*, 882–897.

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.

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.

Moravec, O., & Štěpánek J. (2003). Verbal description of musical sound timbre in Czech language. In A. Askenfelt, S. Felicette, E. Jansson, & J. Sundberg (Eds.), *Proceedings of the Stockholm Music Acoustics Conference (SMAC03)* (pp. 643-645). Stockholm, Sweden: University of Stockholm.

Neher, T., Brookes, T., & Rumsey, F. (2006). A hybrid technique for validating unidimensionality of perceived variation in a spatial auditory stimulus set. *Journal of the Audio Engineering Society*, *54*, 259-275.

Nykänen, A., Johansson Ö., Lundberg, J., & Berg, J. (2009). Modelling perceptual dimensions of saxophone sounds. *Acta Acustica United with Acustica*, *95*, 539-549.

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

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

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.

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: Sijthoff.

Plomp, R. (1976). Timbre of complex tones. In R. Plomp (Ed.), *Aspects of tone sensation: A psychophysical study* (pp. 85-110). London, UK: Academic Press.

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

Pratt, R., & Doak, P. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration*, *45*, 317-328.

Romesburg, C. (2004). *Cluster analysis for researchers*. Raleigh, NC: Lulu Press.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65.

Saitis, C., Giordano, B. L., Fritz, C., & Scavone, G. P. (2012). Perceptual evaluation of violins: A quantitative analysis of preference judgments by experienced players. *Journal of the Acoustical Society of America*, *132*, 4002-4012.

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

Sankiewicz, M., & Budzynski, G. (2007). Reflections on sound timbre definitions. *Archives of Acoustics*, *32*, 442-452.

Schubert, E., & Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta Acustica United with Acustica*, *92*, 820-825.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon.

Terasawa, H. (2009). *A hybrid model for timbre perception: Quantitative representations of sound colour and density* (Unpublished doctoral dissertation). Stanford University.

Štěpánek J. (2006). Musical sound timbre: Verbal descriptions and dimensions. In V.Verfaille (Ed.), *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)* (pp. 121-126). Montreal, Canada: McGill University.

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

von Bismarck, G. (1974b). Sharpness as an attribute of the timbre of steady sounds. *Acustica*, *30*, 159-172.

von Hornbostel, E. M., & Sachs, C. (1914). Systematik der Musikinstrumente: Ein Versuch [Classification of musical instruments] (A. Baines & K. Wachsmann, Trans.). *Galpin Society Journal*, *14*, 3-29.

Wake, S., & Asahi, T. (1998). Sound retrieval with intuitive verbal expressions. In *International Conference on Auditory Display (ICAD'98)* (pp. 1-5). Glasgow, UK: University of Glasgow.

Zacharakis, A., Pastiadis, K., Papadelis, G., & Reiss, J. D. (2011). An investigation of musical timbre: Uncovering salient semantic descriptors and perceptual dimensions. In A. Klapuri & C. Leider (Eds.), *Proceedings of the International Society of Music Information Retrieval Conference (ISMIR 12)* (pp. 807–812). Miami, FL: University of Miami.

Zacharakis, A., Pastiadis, K., Reiss, J. D., & Papadelis, G. (2012). Analysis of musical timbre semantics through metric and non-metric data reduction techniques. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC12) and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM 08)* (pp. 1177-1182). Thessaloniki, Greece: Aristotle University.

Appendix A

Analytic Procedure

MEASURE OF SALIENCE FOR EACH ADJECTIVE

Prior to applying statistical analysis techniques to the data of the two groups, the salience of our adjectives was tested using the following heuristic criterion that is based on the number of times that each adjective was selected by the participants:

$$F_oS(i) = \sum_{n=1}^{23} a_n(i) + \frac{\max\left(\overline{a(i)}\right)^k}{\sum_{n=1}^{23} a_n(i)} \tag{1}$$

where $F_oS(i)$ is the factor of salience for each adjective i, $a_n(i)$ is the number of times a certain adjective i has been chosen by all the participants for describing a particular sound sample n, and $a(i)$ is the $(1, 23)$ vector that contains the number of appearances corresponding to adjective i for the 23 sounds. The power k, to which the maximum number of appearances is raised, was set to 3 so as to maintain a balance between each of the two terms. This factor takes into account a combination of both the overall number of appearances and the maximum number of these appearances for each adjective. This is because even if an adjective has only a small number of overall appearances among all sound samples, a single high maximum at one particular sound can suggest that this adjective is still meaningful. The calculation of F_oS for all the adjectives revealed that no F_oS was less than the mean minus two standard deviations for both groups of listeners. Therefore, no adjective could be characterized as a non-significant outlier and none was discarded at that stage.

STEP-BY-STEP DATA REDUCTION METHODOLOGY

In factor analysis, a mild multicollinearity between variables (in this case verbal descriptors) is generally desirable and for this reason variables that either correlate very highly (extreme multicollinearity) or variables that are not correlated with the rest of the group are discarded prior to the analysis. The steps followed towards data reduction are summarized below:

- A hierarchical cluster analysis (centroid linkage) based on squared Euclidean distances over the verbal descriptors identified the major clusters and outliers among them. The outliers were adjectives that could not be grouped with other adjectives as they appeared to have many instances of low intercorrelation coefficients. As a consequence such variables were discarded based on an observation of the dendrogram.
- In order to further reduce the number of verbal descriptors, a preliminary factor analysis was performed within each cluster and a non-orthogonal

oblique rotation of the extracted factors was employed. The adjectives with extracted communalities < .6 were then discarded as the communality measures the percentage of variance in a given variable explained by all the factors jointly. This criterion ensured that only the verbal descriptors that were adequately explained by the model for each cluster were retained.

- The correlation matrix of the remaining adjectives was inspected and extremely multicollinear verbal descriptors were removed.
- The descriptors selected in the preliminary stage were then subjected to a factor analysis also applying oblique rotation. The descriptors featuring communalities < .6 were again discarded and the remaining set of descriptors was subjected to a final FA. The final data reduction step uses factor loadings as a criterion for labeling the major factors.

Appendix B
Nonlinear Optimal Transformation

Figures B1a and B1b show the dendrograms of the original adjectives and Figures B1c and B1d show the dendrograms of the transformed adjectives as resulting from the application of cluster analysis to both linguistic groups. In the original dendrograms, the absence of clearly defined clusters reflects the lack of cohesive groups among the adjectives. The transformed dendrograms, on the contrary, demonstrate a tighter clustering among the adjectives. The Average Silhouette Width Validity Index (ASWVI) (Rousseeuw, 1987) (readily available in the MATLAB Statistics Toolbox) is a measure of clustering validity that indicates how appropriate the assignment of points to clusters is. It ranges from -1 to 1, with 1 showing best assignment, 0 representing average, and -1 representing inappropriate assignment. In our case the ASWVI increased after the spline ordinal transform from .17 to .42 for the Greek data, and from -.02 to .37 for the English data. A similar pattern was also observed for other relevant indices (e.g. Dunn's index (Dunn, 1974)).

This means that the application of the spline ordinal transformation has led to a higher organization of the data that in turn resulted in a clearer formulation of clusters for both linguistic groups. It is important to note here that our analytic strategy (based on preliminary factor analyses within the identified clusters) could not have been applied to the Greek data without the transformation, due to inadequate clustering.

Subsequently, we applied our analytic strategy to the original and transformed data and compared the results. Table B1 shows the percentage of total and factorial variance prior to rotation that was explained by the final solution in the case of the original and spline ordinal transformed variables. Data from the Greek original variables are not depicted because, as noted above, the deployment of the data reduction methodology was prevented due to inadequate clustering.

TABLE B1. *Comparison of the Amount of Factor Variance Prior to Rotation Explained by Different Variable Transformations and FA Procedures*

Transformation/ method	Percentage of total variance	
	Greek	*English*
Original/PAF	. . .	**77.12**
	. . .	(42.77, 24.54, 9.8)
Spline Ordinal/ML	**82.3**	**82**
	(36.5, 30.5, 15.2)	(48.7, 27.3, 5.9)

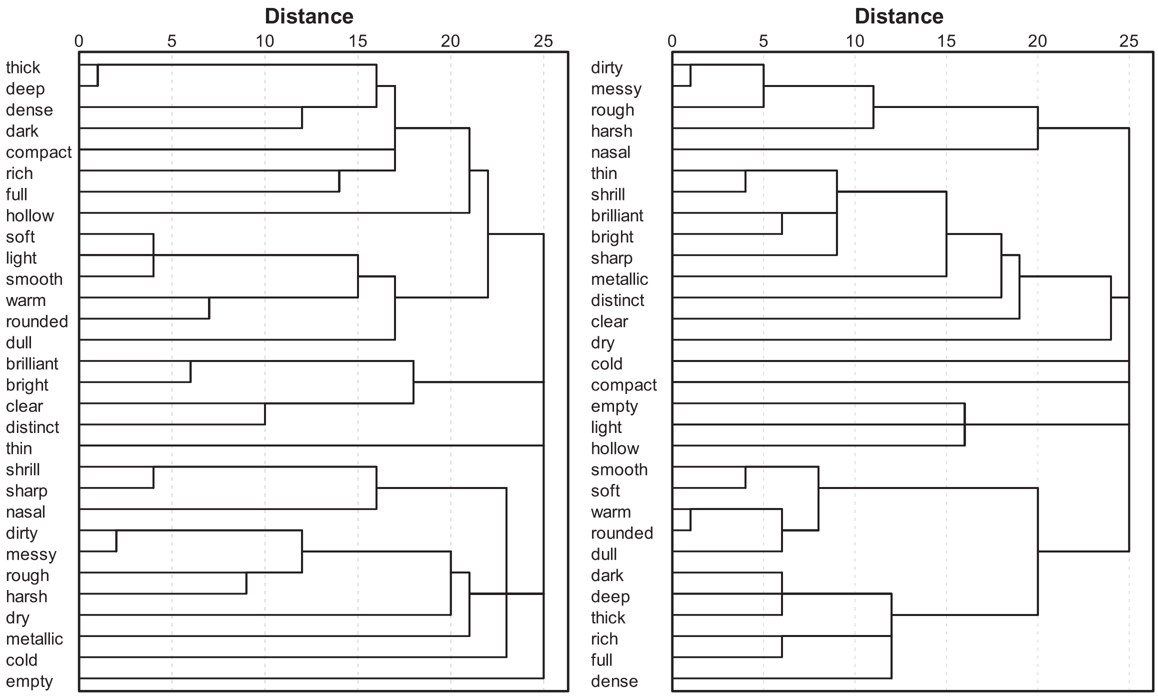
Note: Criterion used for deciding the number of factors: eigenvalues ≥ 1. Total variance is shown in bold and variance explained by each factor in parentheses. (ML: Maximum Likelihood algorithm, PAF: Principal Axis Factoring algorithm)

Table B1 highlights the fact that the spline ordinal transformation explained a larger proportion of total variance than the original case for the English group. Additionally, the spline ordinal transformation increased (by 8.7%) the variance explained by the first two dimensions of the English group. The higher concentration of accounted variance in the first two factors of the optimally transformed solution suggests increased correlations between the transformed variables (also evident from the dendrograms). This finding justifies the use of the optimal nonlinear approach, as the modeling of nonlinear relationships between variables led to greater explained variance by the use of fewer dimensions.

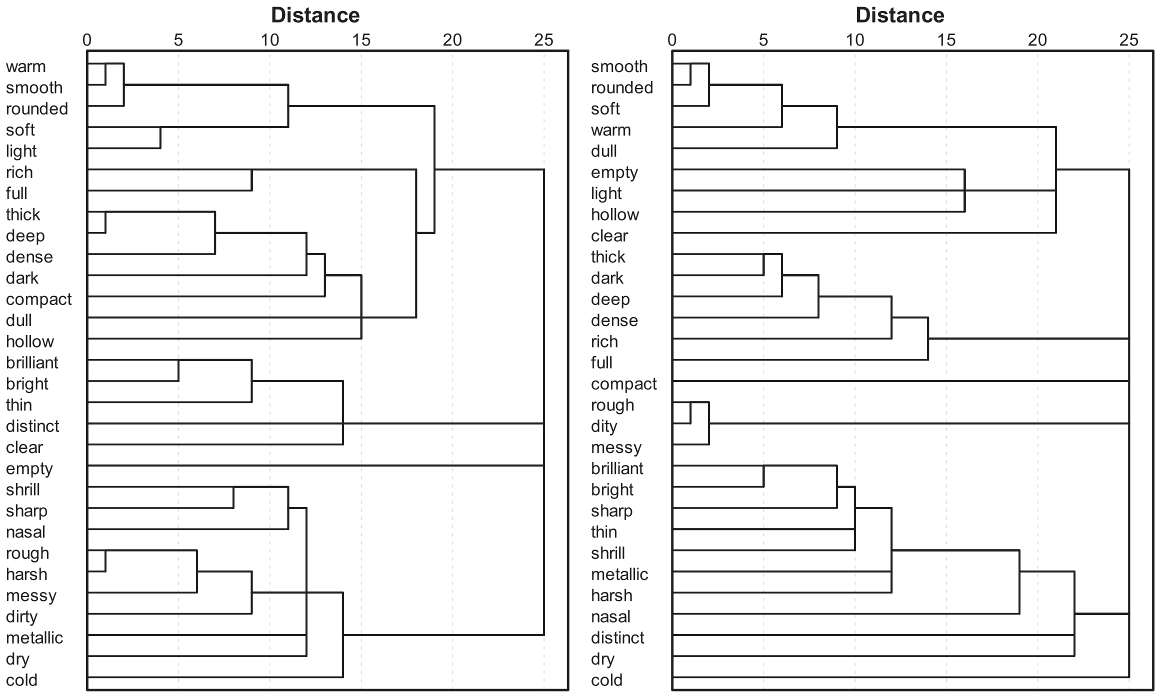
MAXIMUM LIKELIHOOD ALGORITHM FOR FACTOR ANALYSIS

Maximum likelihood (ML) was the preferred factor analysis algorithm. However, the original data featured extreme positive skewness for both linguistic groups, which violates the condition of multivariate normality in the data set that is assumed by ML. Thus, the original English group was analyzed using the principal axis factoring algorithm instead. The transformed data set were analyzed with ML, as the spline ordinal transformation improved the conditions for its application by reducing skewness.

Two goodness-of-fit measures confirmed the validity of our factor analysis model. The *Kaiser-Meyer-Olkin* criterion equaled .80 and .71 for the Greek and English-speaker dataset respectively, both of which are regarded as 'good' (Hutcheson & Sofroniou, 1999, p. 225). Bartlett's test of sphericity also showed statistical significance (p < .001 for both Greek and English speaker datasets), revealing that the correlation matrix was significantly different from the identity matrix (i.e., the variables were not perfectly independent).



(a) Greek original (b) English original



(c) Greek transformed (d) English transformed

FIGURE B1. Dendrograms of the Greek (left) and English (right) adjectives before (a), (b) and after (c), (d) the spline ordinal transformation.

Appendix C

Additional Descriptors Acquired by Free Verbalization

TABLE C1. *Collection of Descriptors From Free Verbalization*

	Semantic Category						
Group	source properties	temp. evolution	emotional terms	sight	touch	size	technical
English	wooden (6), elastic, glassy (4), scraping, synthetic (4), wet, percussive (2), breathy (3), plastic, electronic (2), real, buzzy, brassy, bassy, natural, twangy (2), reedy (3), steely, airy, unnatural (2), pianolike, organlike, desiccated, ethereal, artificial (3), farty, resonant (2), sterile, organic, futuristic, alien, pure (3), jingly, complex (5), distant, muffled, tinny (2)	*wavey (4), flat (5), energetic, rinsing, constant, fluctuating, unstable, oscillating, stable, vibrating (4), continuous, static, pulsating (2), phase-beating, wobbly, cycling (3), throbbing, varied, unsettled, evolving, spinning, consistent, moving (2), bouncy*	*sinister, confusing (2), oppressive, trivial, suave, intriguing, relentless, boring, interesting, ugly, keen, unattractive, annoying, brittle, disorientating, neutral, sickly, unpleasant (2), attractive, harmless*	*blurred, focused, transparent, diffused, fuzzy, smoky, golden, indistinct*	*raspy (2), gentle, blunt, textured, piercing (2), penetrating, grating, wooly*	*large, majestic, heavy, full-bodied, forceful, substantive, limited, superficial, shallow, 1-dimensional, 3-dimensional*	*spectral (2), phasey, sinewave, morphing, distorted, overtoney, vibrating (4), resonant (2), harmonic*
Greek	*spacey (3), (διαστημικός), muffled, (μπουκωμένος), (Indian), Ινδικός, fake (4), (ψεύτικος), electronic (3), (ηλεκτρονικός), noisy, (θορυβώδης)*	*abrupt, (απότομος), discontinuous, (ασυνεχής), vibrated, (βιμπράτο), unstable (3), (ασταθής)*	*sweet (4), unsure, (γλυκός), (αβέβαιος), hesitant, funny, (διστακτικός), (αστείος), relaxing, (χαλαρωτικός), psychedelic, (ψυχαιδελικός), befooling, emetic, (κοροϊδευτικός), (εμετικός), dizzying, hypotonic, (ζαλιστικός), (υποτονικός), nice, annoying (2), (συμπαθητικός), (ενοχλητικός), hair-rising, (ανατριχιαστικός), lacking vividness, (χωρίς ζωντάνια)*	*transparent, (διάφανος), indistinct, (δυσδιάκριτος)*	*squeaky, (τσιριχτός)*	*dynamic, (δυναμικός), intense, (έντονος), exaggerated, (υπεβολικός)*	*echo*

Note: *The number in parentheses represents the number of different participants that have used the term. The Greek terms (appearing in parentheses below the English equivalent) were translated into English by the authors.*

Appendix D

The Extracted Audio Features

TABLE D1. *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* F_0 *(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*