

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Alluri, Vinoo; Toiviainen, Petri

Title: Effect of Enculturation on the Semantic and Acoustic Correlates of Polyphonic Timbre

Year: 2012

Version:

Please cite the original version:

Alluri, V., & Toiviainen, P. (2012). Effect of Enculturation on the Semantic and Acoustic Correlates of Polyphonic Timbre. *Music Perception*, 29(3), 297-310.
<https://doi.org/10.1525/mp.2012.29.3.297>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

EFFECT OF ENCULTURATION ON THE SEMANTIC AND ACOUSTIC CORRELATES OF POLYPHONIC TIMBRE

VINOO ALLURI & PETRI TOIVIAINEN
University of Jyväskylä, Jyväskylä, Finland

POLYPHONIC TIMBRE PERCEPTION WAS INVESTIGATED IN A cross-cultural context wherein Indian and Western nonmusicians rated short Indian and Western popular music excerpts (1.5 s, $n = 200$) on eight bipolar scales. Intrinsic dimensionality estimation revealed a higher number of perceptual dimensions in the timbre space for music from one's own culture. Factor analyses of Indian and Western participants' ratings resulted in highly similar factor solutions. The acoustic features that predicted the perceptual dimensions were similar across the two participant groups. Furthermore, both the perceptual dimensions and their acoustic correlates matched closely with the results of a previous study performed using Western musicians as participants. Regression analyses revealed relatively well performing models for the perceptual dimensions. The models displayed relatively high cross-validation performance. The findings suggest the presence of universal patterns in polyphonic timbre perception while demonstrating the increase of dimensionality of timbre space as a result of enculturation.

Received September 17, 2010, accepted July 17, 2011.

Key words: polyphonic timbre, cross-cultural, perceptual dimensions, differentiation of dimensions, acoustic feature analysis

MUSIC IS KNOWN TO BE AN IMPORTANT FACET OF ALL human cultures (Merriam, 1964). To date, the majority of research on music perception and cognition has focused on Western music and listeners. To study music perception and cognition as a general phenomenon, there is a need to broaden the scope of research to accommodate a more global palette of both music and listeners. A wide range of music exists in cultures all around the world that differ from each other in various ways. Nevertheless, a fundamental question is

often posed in the field of comparative musicology about the universals in music structure and its perception.

Carterette and Kendall (1999) outline some structural elements of music that are common across cultures, such as the use of stable reference pitch and reference pulse, and the division of octave into scale steps, for instance. The authors contend that the process of music cognition in fact shares universals, and advocate the notion that the differences in musical elements across cultures are but an "elaboration of a few universals." The question then arises as to whether the presence of structural universals indicates the presence of universals in music perception. Despite the diversity in music across cultures, does music perception share fundamental biological sensory processing mechanisms? Evidence from infant studies indicates the presence of universals in the processing of some music principles (see Trehub, 2000, for an overview), such as categorization of successive events based on similarity in pitch, timbre, or loudness, interval categorization based on consonance or dissonance, melody contour equivalence regardless of pitch transposition, and rhythm equivalence despite tempo changes. Harwood (1976) delineates a few elements of auditory perception that are found to be similar across cultures, such as categorical pitch perception, judgments of octave equivalence, auditory stream segregation, and melodic contour processing. Results of some perceptual studies support the view that there exist cross-cultural regularities in the perception of certain structural elements of music such as beat (Toiviainen & Eerola, 2003), and intervals (Smith & Williams, 1999), as well as in the elicitation of melodic expectations (Krumhansl et al., 2000; Eerola, Louhivuori, & Lebaka, 2009; Castellano, Bharucha, & Krumhansl, 1984) and affect identification (Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunaga, 2004; Gregory & Varney, 1996; Meyer, Palmer, & Mazo, 1998; Adachi, Trehub, & Abe, 2004; Zacharopoulou & Kyriakidou, 2009). For example, Castellano, Bharucha, and Krumhansl (1984) reported marked similarities between Indian and Western participants' probe tone ratings in a North-Indian melodic expectancy setting. Similarly, Balkwill and Thompson (1999) found that listeners were

successfully able to decode musically expressed emotions in an unfamiliar tonal system.

Nevertheless, in addition to commonalities, some of these studies highlight differences in music perception and cognition, and associate them to one's culture-specific knowledge. Krumhansl et al. (2000), for instance, found that schematic knowledge about Western music influenced melodic expectations evoked by Sami yoiks more in Western listeners than they did in Sami listeners, whereas veridical (style specific) knowledge had a slightly larger influence in Sami listeners than in Western listeners. On similar lines, Curtis and Bharucha (2009) contend that musical expectancies, which can shape affective responses, are largely molded by cultural background. Balkwill et al. (2004) hinted at the existence of possible complex cultural influences that may influence emotional judgements of music from an unfamiliar culture, although this remains to be investigated. Furthermore, Demorest, Morrison, Beken, and Jungbluth (2008) contended that the universality in the perception of musical properties suggested by several studies may be attributed to their major reliance on music following rules of the Western diatonic system. They supported the notion that inclusion of participants and stimuli from several distinct cultures would provide grounds to identify and establish "universal" musical properties and their perception.

In addition to behavioral studies, several cross-cultural studies in the field of neuroscience have investigated the issue of universality in music processing. In an fMRI study carried out by Morrison, Demorest, Aylward, Cramer, and Maravilla (2003), no overall differences in the activation patterns were found for Western musicians while listening to familiar (Western) and unfamiliar (Chinese) music, suggesting the involvement of common (universal) neural mechanisms in the auditory analysis of both musical styles. Nevertheless, the musicians in comparison with the nonmusicians showed significant activations in the right superior temporal gyrus for both kinds of music, suggesting music expertise as a factor that modulates music perception and processing. In a cross-cultural ERP study on the perception of musical phrases (Nan, Knösche, & Friederici, 2006), the authors reported no differences between German and Chinese musicians while detecting familiar and unfamiliar, phrased and unphrased excerpts, although they found differences in the reaction to music from one's own culture in the earlier time windows (100 - 450 ms) of music processing. In a following fMRI study in which the participant pool was more homogeneous (i.e., only female German classical musicians; Nan, Knösche, Zysset, & Friederici, 2008), some significant differences were found in parts of the motor system and

the ventromedial prefrontal cortex that processed and evaluated familiar (Western) and unfamiliar (Chinese) music, hence for the first time illustrating effects of musical style acculturation on brain processing.

In summary, evidence from both behavioral and neural studies suggests macro level similarities across cultural backgrounds, with differences being observed on a more micro level. These differences could be attributed to one of the mechanisms involved in perceptual learning, that is, *differentiation* (Goldstone, 1998).

Goldstone (1998) discusses differentiation as a mechanism enabling perceptual learning, wherein prior exposure or training allows for the dissociation of perceptual elements from a single stimulus. A subcategory of this kind of perceptual learning is *differentiation of dimensions*, which refers to separation of perceptual dimensions and increase in the number thereof as a result of prior exposure or training. For instance, in the sensory modality of vision, Burns and Shepp (1988) found that color experts were able to discriminate changes in all three perceptual dimensions of color, namely hue, value, and chroma, whereas the nonexperts were unable to selectively perceive changes in hue. Similarly, in the olfactory sensory modality, studies have suggested that familiarity or experience with an odor heightens the intensity of its perception (Ayabe-Kanamure et al., 1998; Li, Luxenberg, Parish, & Gottfried, 2006). Moreover, on a neural level, it has been shown in animal models and humans alike that both associative conditioning and stimulus exposure alter receptive fields of sensory neurons in most sensory modalities (Brattico, Tervaniemi, & Picton, 2003; Fletcher & Wilson, 2002; Gibson, 1953; Gilbert, Sigman, & Crist, 2001).

Differentiation of dimensions as a form of perceptual learning has not been widely studied in musical contexts to date. Due to the multidimensional nature of timbre, one could postulate such a mechanism to be involved in its perception. In particular, one could hypothesize that prior exposure through enculturation may increase the dimensionality of perceptual timbre space for music from one's own culture.

While the perception of many musical features has been investigated in a cross-cultural setting, there remain gaps in knowledge about the perception of certain elements of music, in particular polyphonic timbre. Polyphonic timbre, or the "global sound" of a piece of music (Aucouturier, 2006), has been found to be a significant component especially in studies that involve perceptual tasks such as genre and song identification (Gjerdingen & Perrott, 2008; Krumhansl, 2010) or emotional affect attribution (Bigand, Vieillard, Maudrell, Marozeau, & Dacquet, 2005;

Peretz, Gagnon, & Bouchard, 1998). Additionally, it has been demonstrated to be an important element for computational categorization according to genre, style, mood, and emotions (see Alluri & Toiviainen, 2010, for an overview).

Alluri and Toiviainen (2010) explored the perceptual components of polyphonic timbre and their corresponding acoustical features. The first part comprised an exploration into semantic associations of polyphonic timbre. As a result, they found similarities between semantic associations of monophonic and polyphonic timbre. Following this, to investigate the acoustic correlates, short excerpts of Indian popular music (each 1.5 s long) were rated on eight bipolar scales, namely “Colorful-Colorless,” “Warm-Cold,” “Dark-Bright,” “Empty-Full,” “Soft-Hard,” “Strong-Weak,” “High Energy-Low Energy,” and “Acoustic-Synthetic.”¹ Factor analysis revealed three perceptual dimensions representing “Activity,” “Brightness,” and “Fullness.” Furthermore, relatively high correlations were observed between certain acoustic features and these perceptual dimensions, in particular for acoustic features that quantify spectrotemporal modulations, that is, the sub-band fluxes. These findings suggest that they play a vital role in the perception of polyphonic timbre. These perceptual dimensions were discussed in light of their counterparts in the monophonic timbre domain. The Activity dimension is comparable to the perceptual dimension in monophonic timbre spaces representing spectral fluctuations (Grey, 1977; McAdams, Winsberg, de Soete, & Krimphoff, 1995). Similarly, perceptual brightness, commonly associated with the spectral centroid, has been repeatedly found as one of the dimensions in monophonic timbre spaces (Beauchamp, 1982; Grey, 1977). However, perceptual fullness, which has been previously suggested as a perceptual quality of timbre (Helmholtz, 1885/1954), has never been reported in perceptual studies on monophonic timbre. Finally, the perceptual dimension of Activity could be predicted to a high degree of accuracy ($R^2 = .70$) using linear regression on acoustic features extracted from the sound. A limitation of this study was the use of only one style of music (Indian popular music) and only one type of listeners (Western), which may not allow for generalization of the results. To overcome this limitation it would be necessary to extend the experimental setting to accommodate a wider range of listeners and musical material.

¹ The scales were chosen from the first part of the investigation by Alluri and Toiviainen (2010) wherein a series of studies investigated methodically the appropriateness of several bipolar scales as semantic differentials of polyphonic timbre.

In the present study we investigate the perception of polyphonic timbre in a cross-cultural setting with the aim to identify eventual universal patterns and culture-dependent differences. To this end, we employ an interdisciplinary approach that involves experimental psychology, computational analysis of audio, and statistical modelling. Participants of Indian and Western origin rated two sets of stimuli comprising Indian and Western popular music using the semantic differential approach. From the obtained ratings, we estimated perceptual acuity employing a new approach based on intrinsic dimensionality estimation. Factor analyses were subsequently performed on the obtained ratings to examine the intrinsic dimensionality of the polyphonic timbre space and thereby identify the underlying perceptual dimensions. Finally, regression analyses were performed to investigate how well the perceptual dimensions can be predicted from the acoustic features. This process is explained in detail in the following sections.

Method

STIMULI

We selected two sets of stimuli comprising Indian and Western music. Each stimulus set consisted of one hundred musical excerpts with a duration of 1.5 s. The Indian popular music set was the same as that used in a previous polyphonic timbre experiment (Alluri & Toiviainen, 2010). The Western music excerpts were randomly chosen from the film soundtrack set used by Eerola and Vuoskoski (2010). Both stimulus sets encompassed a wide range of genres such as pop, rock, disco, and electronic, and contained various instrument combinations such as piano, violin, percussive instruments, and guitar. All the excerpts were converted to mono files in wav-format (44.1kHz, 16 bit) and were equalized in level by RMS value normalization. Conversion of the stimuli from stereo to mono was required as computational acoustic feature extraction in the field of Music Information Retrieval is performed on mono files. The stimulus sets can be found at <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/PolyphonicTimbreStimuli>.

PARTICIPANTS

A total of 150 nonmusicians of Indian and Western origin rated the music excerpts. The Indian participants were born and raised in India. Of the Western participants, all of whom were students at the University of Jyväskylä, the majority (72%) were born and raised in Europe or North America and the remaining participants

were of Non-Indian origin. Henceforth, the groups will be referred to by a two-letter combination where the first letter represents the stimulus set they rated and the second their origin (II – Indian music, Indian listeners; IW - Indian music, Western listeners; WI - Western music, Indian listeners; WW - Western music, Western listeners). Participants were assigned to these groups at random.

Group II comprised 43 participants (22 men, age $M = 25.5$, $SD = 6.9$), group IW 38 participants (15 men, age $M = 24.2$, $SD = 3.6$), group WI 33 participants (20 men, age $M = 23.5$, $SD = 3.8$), and group WW 36 participants (10 men, age $M = 24$, $SD = 3.7$). A series of t tests revealed no significant differences for age between any of the listener groups.

The majority (89%) of the participants reported listening to music as a hobby (music listening hours per week: Group II $M = 25.47$, $SD = 6.83$, group IW $M = 24.19$, $SD = 3.64$, group WI $M = 23.47$, $SD = 3.79$, group WW $M = 23.97$, $SD = 3.68$). Again, no significant differences for frequency of music listening were found between any of the groups.

Based on these results we concluded that the groups did not significantly differ from each other in terms of their age or their frequency of music listening. None of the participants reported having received any formal music education. Almost all Westerners reported having very little or no familiarity with Indian popular music and high familiarity with Western music. Almost all Indian participants reported high familiarity with Indian music and varying levels of familiarity with Western music. None reported any hearing problems. All the participants were fluent in English.

PROCEDURE

The listening experiment took place in a silent room in Hyderabad, India (for groups II and WI) and Jyväskylä, Finland (for groups IW and WW). Participants were given written instructions before the experiment. To present the stimuli and obtain the ratings, we used a graphical computer interface that displayed the eight rating scales “Colorful-Colorless,” “Warm-Cold,” “Dark-Bright,” “Empty-Full,” “Soft-Hard,” “Strong-Weak,” “High Energy-Low Energy,” and “Acoustic-Synthetic.” Each scale was divided into nine levels from which participants could choose the level that best described the music excerpt presented. Due to the short duration of the stimuli and high number of perceptual scales to be rated, participants had the provision to listen to each excerpt as many times as they wished. Prior to the actual experiment, participants were allowed to familiarize themselves with the working of

the interface. The music examples were presented via headphones in random order for each participant.

Results

PERCEPTUAL DATA ANALYSIS

The behavioral data were initially checked for inconsistencies and outliers. For each scale, one to two participants were eliminated owing to their mean intersubject correlation being 2 SD s below the overall mean intersubject correlation.²

The ratings were analyzed for each stimulus set separately. High agreement between the participants’ ratings, in terms of Cronbach’s alpha, was observed in all listener groups (Group II $.96 \geq \alpha \geq .74$ for all scales, except for “Warm—Cold” with $\alpha = .67$; IW $.97 \geq \alpha \geq .87$ for all scales; WI $.96 \geq \alpha \geq .71$ for all scales & WW $.97 \geq \alpha \geq .90$ for all scales). Table 1 displays Cronbach alphas for each of the perceptual scale for all groups.

These findings suggest the presence of fairly consistent opinions among listeners with respect to these scales. Therefore, for subsequent analysis, the individual ratings for each scale within each group were averaged across all participants.

Next, to investigate consensus in the associations of the perceptual qualities of the stimuli, we looked at the correlations of the rating scales between listener groups within the same stimulus sets. Figure 1 summarizes the intergroup correlations for each scale and each stimulus set.

As can be seen, all rating scales display relatively high correlations across listener groups for each stimulus set.

² Van Selst and Jolicoeur (1994) suggest plus or minus 2 or 3 SD s as common thresholds for outlier detection. To make the analysis compatible with that in Alluri and Toiviainen (2010), we opted for the minus 2 SD criterion.

TABLE 1. Cronbach Alphas for Each of the Bipolar Scales

	Cronbach alpha			
	Group II	Group IW	Group WI	Group WW
Colorless - Colorful	.91	.94	.92	.96
Warm - Cold	.67	.93	.84	.96
Dark - Bright	.91	.96	.95	.97
Empty - Full	.81	.87	.85	.90
Soft - Hard	.96	.97	.96	.97
Strong - Weak	.84	.93	.88	.94
High Energy - Low Energy	.92	.95	.94	.97
Acoustic - Synthetic	.74	.95	.71	.96

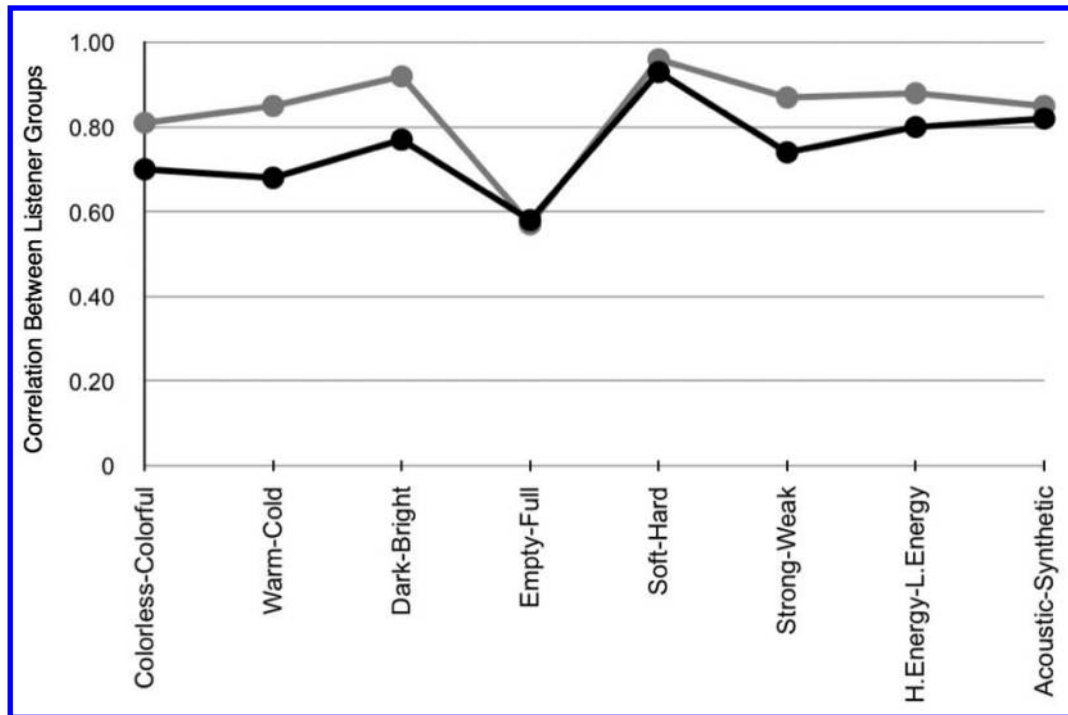


FIGURE 1. Intergroup correlations for all scales and both stimulus sets (all correlations are significant at $p < .001$).

This suggests that the semantic associations of timbre are not largely affected by listeners' cultural background.

Next, to investigate the interdependencies of the rating scales, and thus the underlying structure of the perceptual ratings, we performed intrinsic dimension estimation for all the data sets.

INTRINSIC DIMENSIONALITY ESTIMATION

The intrinsic dimensionality of a data set can be described as an estimate of the number of "independent" variables required to represent it. In order to evaluate the intrinsic dimensionality of our data sets we employed one of the most commonly used estimation methods, which is based on the Eigenvalues satisfying the Kaiser criterion (1960).³ As a result, two intrinsic dimensions were observed for the sets IW and WI, whereas II and WW displayed intrinsic dimensionality of three. In what is to follow, we investigate the two-factor and three-factor solutions for all groups.

Two-factor solutions. Table 2 summarizes the two-factor solutions obtained from the ratings using varimax rotation.

As can be seen, the factor loadings display a consistent overall pattern. The first factor regularly has high

loadings from the scales "Colorless-Colorful," "Warm-Cold," and "Dark-Bright." The factor-loading pattern is similar to that of the Brightness dimension found by Alluri and Toiviainen (2010). This factor thus appears to describe the perceptual brightness of the musical excerpt. Additionally, the scale "Empty-Full" has high loadings on the first factor for all groups except WI.

The scales "Strong-Weak," "Soft-Hard," and "High Energy-Low Energy" have high loadings on the second factor for all groups. Again, a similar factor-loading pattern was observed for the Activity dimension in Alluri and Toiviainen (2010) and hence this factor can be interpreted to represent the overall activity in the stimuli.

Three-factor solutions. Table 3 summarizes the three-factor solutions obtained from the ratings using varimax rotation.

As can be seen, the loadings for the first two factors display patterns similar to those in their respective two-factor solutions. For groups IW and WI, in line with the observation of the data sets displaying an intrinsic dimensionality of two, the third factor does not show any clear patterns and explains a small proportion of the variance. On the other hand, for groups II and WW, the third factor displays high loadings from the scale "Acoustic-Synthetic." Additionally, for group II, the scale "Soft-Hard" has the highest loadings on this

³ The eigenvalues were obtained as a result of factor analysis using varimax rotation.

TABLE 2. Two-Factor Solutions for Each of the Groups in Terms of Factor Loadings and Variances Explained

Stimulus Set	Indian				Western			
	Indian		Western		Indian		Western	
Listener Type	II		IW		WI		WW	
Group	II		IW		WI		WW	
Factor Number (Variance explained)	Factor 1 (48%)	Factor 2 (35%)	Factor 1 (50%)	Factor 2 (38%)	Factor 1 (43%)	Factor 2 (42%)	Factor 1 (53%)	Factor 2 (35%)
Colorless - Colorful	.98	.05	.95	.07	.93	.29	.97	-.12
Warm - Cold	-.82	.24	-.94	.19	-.95	-.07	-.94	.23
Dark - Bright	.93	.02	.91	.08	.96	.08	.94	-.02
Empty - Full	.82	.40	.82	.46	.50	.80	.90	-.15
Soft - Hard	-.13	.91	-.38	.91	-.40	.86	-.43	.89
Strong - Weak	-.52	-.79	-.03	-.98	.02	-.96	-.03	-.95
High Energy - Low Energy	-.53	-.85	-.41	-.88	-.22	-.96	-.54	-.82
Acoustic - Synthetic	-.33	.61	-.64	.43	-.52	.26	-.52	.56

Note: Factor loadings having absolute values above .65 are displayed in boldface.

factor. Interestingly, the scale “Empty-Full” loads highly on the Brightness dimension for the Western listener groups IW and WW, and loads highly onto the Activity dimension for the Indian listener groups, II and WI. Next, we investigated the relationship between the obtained perceptual dimensions based on the previous factor analyses, and acoustical features of the stimuli.

ACOUSTIC DATA ANALYSIS

Feature selection has always been an important step in computational modeling of music. For the present study, acoustic features used in previous timbre related studies, and those that are easily interpretable from a perceptual point of view, were chosen (Alluri & Toiviainen, 2010; Aucouturier, 2006; Aucouturier & Pachet, 2003; McAdams et al., 1995; Tzanetakis & Cook, 2002). A total of sixteen features, namely the zero-crossing rate, spectral centroid, high energy-low energy ratio, roughness, entropy, spectral roll-off, and sub-band flux (10 coefficients) were extracted using the MIRToolbox (Lartillot & Toiviainen, 2007).

A complete description of the feature selection and extraction procedures can be found in Alluri and Toiviainen (2010).

Modelling Perceptual Dimensions with Acoustic Features

CORRELATION ANALYSES

Next, to investigate the acoustic correlates of the perceptual dimensions, we performed correlation analyses between factor scores and the computationally extracted acoustic features. Due to the similarities observed in the loading patterns between the two-factor and three-factor solutions, we focused on the two-factor solutions to allow for comparisons across all groups. A Lilliefors test was used to check for the normality of the distribution of all the perceptual data and the acoustic features. As a result, the Brightness scores for Group IW, and six out of the sixteen acoustical features were transformed using the Box-Cox transformation (Box & Cox, 1964). Multivariate outliers were screened and

TABLE 3. Three-Factor Solutions for Each of the Groups in Terms of Factor Loadings and Variances Explained

Stimulus	Indian						Western					
	Indian			Western			Indian			Western		
Listener Type	Indian			Western			Indian			Western		
Group	II			IW			WI			WW		
Factor Number (Variance explained)	Factor 1 (37%)	Factor 2 (25%)	Factor 3 (24%)	Factor 1 (49%)	Factor 2 (38%)	Factor 3 (5%)	Factor 1 (41%)	Factor 2 (40%)	Factor 3 (8%)	Factor 1 (46%)	Factor 2 (33%)	Factor 3 (16%)
Colorless - Colorful	.87	.41	-.13	.94	.07	-.14	.28	.90	-.23	.90	-.03	-.38
Warm - Cold	-.75	-.20	.37	-.91	.18	.21	-.05	-.93	.20	-.91	.15	.32
Dark - Bright	.96	.21	-.05	.98	.08	.18	.04	.97	-.14	.99	.01	-.08
Empty - Full	.63	.71	.05	.76	.47	-.41	.83	.48	-.21	.73	.29	-.58
Soft - Hard	-.12	.36	.88	-.37	.91	.10	.83	-.29	.38	-.41	.83	.35
Strong - Weak	-.32	-.82	-.43	-.02	-.98	.04	-.96	.01	-.07	.04	-.97	-.04
High Energy - Low Energy	-.41	-.69	-.57	-.41	-.89	.03	-.93	-.29	-.16	-.51	-.84	-.03
Acoustic - Synthetic	-.25	.06	.68	-.58	.43	.33	.20	-.39	.55	-.34	.41	.74

Note: Factor loadings having absolute values above .65 are displayed in boldface.

constrained to $\pm 2 SD$. Figure 2 displays the correlations between the perceptual dimensions and each of the acoustic features. The correlation pattern for each group and for each perceptual dimension will subsequently be referred to as the acoustic profile of the respective group and the respective dimension.

We observed some moderately high correlations between the perceptual dimensions and some acoustic features. For instance, Brightness correlated moderately with the zero-crossing rate ($.42 \geq r \geq .28$, all $p < .01$), Sub-Band No. 9 flux (6,400 -12,800 Hz) ($.48 \geq r \geq .28$, all $p < .01$ except for WW, $r(98) = .08$, *ns*) and spectral centroid ($.41 \geq r \geq .21$, all $p < .05$). Activity correlated highly with spectral entropy ($.75 \geq r \geq .64$, all $p < .001$), Sub-Band No. 7 flux (1,600 -3,200 Hz) ($.68 \geq r \geq .55$, all $p < .001$), Sub-Band No. 8 flux (3,200 -6,400 Hz) ($.75 \geq r \geq .56$, all $p < .001$) and Sub-Band No. 9 flux (6,400 -12,800 Hz) ($.74 \geq r \geq .52$, all $p < .001$).

The correlations suggest that perceptually bright stimuli possess relatively more high frequency energy (zero-crossing rate, spectral centroid) than less bright stimuli.

In addition, Bright stimuli tended to have more fluctuation in the higher frequency regions (Sub-Bands No. 7-10, 1,600 -22,050 Hz) and less fluctuation in the lower frequency regions (Sub-Bands No. 1-2, 0-100 Hz).

As can be further seen, the stimuli with higher Activity scores tend to be associated with higher flux in the high end of the spectrum, in particular at frequencies in the range of 1600 Hz – 12800 Hz (Sub-Bands No. 7-9), more energy in the higher frequency bands when compared to the lower frequency bands (high energy-low energy ratio), and a more even spectral distribution (entropy).

Overall, the acoustic features correlated more highly with Activity than with Brightness. Nevertheless, it is noteworthy that for both perceptual dimensions the acoustic profiles are similar across the groups. Table 4 displays the correlation values between the acoustic profiles of all the groups for both perceptual dimensions.

As can be seen, the correlations between the acoustic profiles are higher for Activity than for Brightness. This finding suggests that all the groups relied on similar

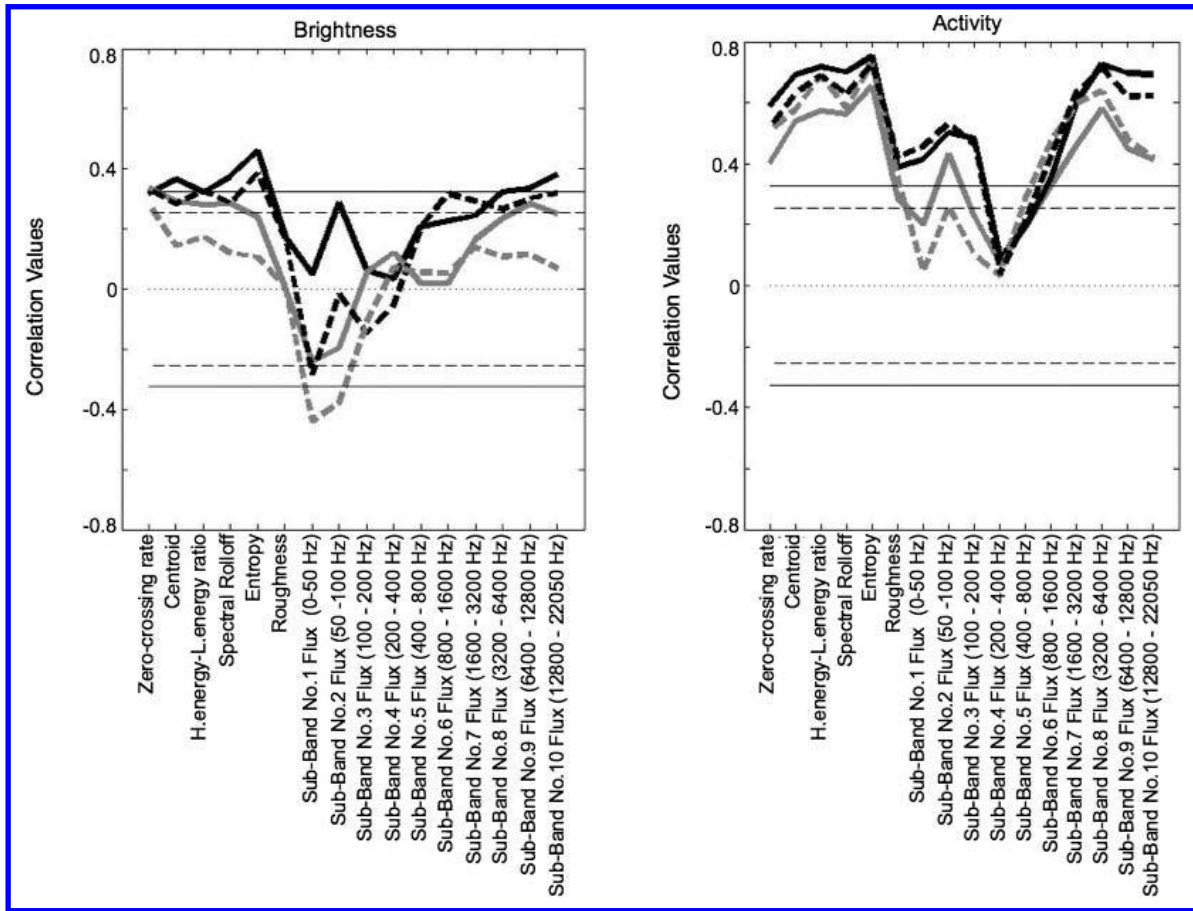


FIGURE 2. Acoustic profiles for each perceptual dimension and group (correlations with absolute values higher than those indicated with the dashed and solid lines are significant at the levels $p < .01$ and $p < .001$, respectively).

TABLE 4. Correlations Between the Acoustic Profiles Across Groups

	Brightness			Activity		
	II	IW	WI	II	IW	WI
IW	.87***			.91***		
WI	.59*	.72**		.82***	.68**	
WW	.52*	.81***	.88***	.79***	.63**	.97***

* $p < .05$ ** $p < .01$ *** $p < .001$

acoustic properties of the stimulus to obtain a mental representation of Activity. Thus, it can be considered that the perception of Activity is less dependent on stimulus set and listener type than is Brightness.

REGRESSION ANALYSES

As a next step, we performed two kinds of regression analyses, step-wise and principal components regression (PCR), to investigate the degree to which the perceptual

dimensions could be predicted by the acoustic features. Step-wise regression, while being explicit regarding the features that comprise the model created, has been criticized for being prone to overfitting. PCR, on the other hand, is more robust against overfitting but is more implicit. Hence, both modelling approaches were employed in order to increase the reliability of the results.

Step-wise regression analyses were performed with an inclusion criterion of $p < .05$ and an exclusion criterion of $p < .10$. Tables 5 and 6 summarize the results obtained employing stepwise regression.

As can be seen, the models can explain a larger proportion of the variance in the Activity dimension than in the Brightness dimension. A maximum of 25% of the variance in the Brightness dimension can be explained by one or two acoustic features. The models of the Activity dimension explain a relatively high proportion of the variance ($.58 < R^2 < .72$) by two to three acoustic features.

Next, to reduce unwanted effects due to eventual collinearity between the features, a principal component

TABLE 5. Beta-Coefficients as a Result of Stepwise Regression for Brightness Dimension

	Brightness			
	II $R^2 = .25$	IW $R^2 = .16$	WI $R^2 = .10$	WW $R^2 = .15$
Acoustic Features	β Coefficients			
Zero-crossing Rate	–	.41	.33	.32
Sub-Band No. 1 Flux	–	–	–	-.42
Sub-Band No. 2 Flux	-.23	–	–	–
Sub-Band No. 9 Flux	-.43	–	–	–

analysis was performed on the features. The first 6 PCs were chosen conforming to the maximal predictor-observation ratio as suggested by Green (1991) and were subsequently used as independent variables in a linear regression analysis.

The models for Brightness were able to explain 22%, $F(7, 92) = 5.70, p < .001$, of the variance for group II, 15%, $F(7, 92) = 3.96, p < .01$, for group IW, 21%, $F(7, 92) = 5.39, p < .001$, for group WI, and 23%, $F(7, 92) = 5.97, p < .001$, for group WW respectively.

Models for Activity performed much better by explaining 60%, $F(7, 92) = 25.87, p < .001$, of the variance for group II, 67%, $F(7, 92) = 31.29, p < .001$, for group IW, 70%, $F(7, 92) = 38.87, p < .001$, for group WI, and 69%, $F(7, 92) = 37.57, p < .001$, for group WW. Overall, the results of the principal component regression proved to be quite similar in terms of the variance explained to those obtained using step-wise regression.

CROSS-VALIDATION OF MODELS

A crucial step in estimating the success or validity of any model is to assess its predictive power for different data sets. Hence, in order to evaluate how well each of the Brightness and Activity models generalized to independent data sets, we performed cross-validation. To this end, we used each of the aforementioned stepwise regression models and PC regression models to predict the respective perceptual dimensions of the other data sets (e.g., II with WW, II with IW, WW with WI, etc.). Table 7 summarizes the step-wise regression cross-validation results and Table 8 displays the PC regression cross-validation results obtained henceforth. The first column indicates the regression models that are used to predict the corresponding perceptual dimensions of the data sets indicated in the second row.

The regression models for Activity perform well, as can be observed from the high correlation values between the predicted and observed ratings. Brightness

TABLE 6. Beta-Coefficients as a Result of Stepwise Regression for Activity Dimension

	Activity			
	II $R^2 = .58$	IW $R^2 = .64$	WI $R^2 = .65$	WW $R^2 = .72$
Acoustic Features	β Coefficients			
Entropy	.42	.45	.66	–
Sub-Band No. 2 Flux	.37	.19	.31	.41
Sub-Band No. 6 Flux	–	–	–	.30
Sub-Band No. 7 Flux	.28	.40	–	–
Sub-Band No. 8 Flux	–	–	–	.51

models, on the other hand, exhibit lower correlations and greater variability in the correlation values especially across stimulus sets. This is in line with the observation that the acoustic profiles tend to have lower correlation values across data sets for the Brightness dimension than for the Activity dimension (see Table 4). These findings suggest that the perceptual dimension of Activity can be robustly modeled and is generalizable.

Discussion

The main motivation of this study was to investigate the similarities and differences in polyphonic timbre perception in a cross-cultural setting. To this end, we explored the effect of enculturation on the dimensionality of the polyphonic timbre space and the configuration of these dimensions. In what follows we discuss our main findings.

Participants displayed a high degree of consistency in their ratings both within and between groups. This is reflected in the high values of the Cronbach's alphas and the high correlations between the ratings. This finding thus suggests that there exist commonalities in semantic associations to polyphonic timbre despite cultural differences. Intergroup correlation analysis for each of the rating scales revealed higher consensus for the Western stimuli than for the Indian stimuli (see Figure 1). This finding can be attributed to the asymmetry in the degree of familiarity of the participants with the stimulus sets. In particular, most of the participants of group WI reported familiarity with Western music whereas almost all participants of group IW reported as being as highly unfamiliar with Indian music. Ideally, a cross-cultural setting would involve homogeneous groups wherein each participant group would be extremely familiar with music from one culture and completely ignorant with that of the other culture. However, the global prevalence of Western music makes it increasingly challenging to meet such a requirement.

TABLE 7. Cross-Validation Results of Stepwise Regression Models

	Brightness				Activity			
	II	IW	WI	WW	II	IW	WI	WW
II	.51***	.31**	.13	-.09	.77***	.78***	.82***	.83***
IW	.40***	.41***	.33***	.28**	.74***	.81***	.80***	.80***
WI	.40***	.41***	.33***	.28**	.74***	.74***	.80***	.80***
WW	.18	.43***	.36***	.50***	.75***	.75***	.82***	.85***

Note: The first column indicates the regression models that are used to predict the corresponding perceptual dimensions of the data sets indicated in the second row. The values represent the correlations between the actual and predicted values. * $p < .05$ ** $p < .01$ *** $p < .001$

TABLE 8. Cross-Validation Results of PC Regression Models

	Brightness				Activity			
	II	IW	WI	WW	II	IW	WI	WW
II	.52***	.32**	.12	-.04	.79***	.78***	.83***	.84***
IW	.36***	.45***	.29**	.29**	.76***	.82***	.82***	.83***
WI	.11	.26**	.51***	.47***	.77***	.78***	.85***	.83***
WW	.02	.27**	.45***	.53***	.78***	.79***	.83***	.84***

Note: The first column indicates the regression models that are used to predict the corresponding perceptual dimensions of the data sets indicated in the second row. The values represent the correlations between the actual and predicted values.

* $p < .05$ ** $p < .01$ *** $p < .001$

We next examined the dimensionality of the perceptual timbre spaces by estimating the intrinsic dimensionality of the data sets. We found that participants' high level of familiarity with the musical style was associated with increased intrinsic dimensionality, from two dimensions to three, of their respective timbre spaces. This result suggests that, in line with our hypothesis, prior exposure through enculturation may enhance differentiation of perceptual dimensions for music and thus causing an increase in the dimensionality of perceptual timbre space for music of one's own culture. This finding is also in line with the ideas put forward by Hannon and Trainor (2007), wherein the authors reported enculturation as a factor that influences and shapes auditory perceptual capabilities.

Subsequently, we compared the underlying structure of these intrinsic dimensions. To this end, we investigated the loading patterns of the two-factor and three-factor solutions of the data sets. The two-factor solutions displayed highly similar loading patterns across participant groups, suggesting similarities in semantic associations with polyphonic timbre. Furthermore, both of these perceptual dimensions, Brightness and Activity, matched closely with the results of the study by Alluri and Toiviainen (2010), who used musically trained Western participants.

Following this, we investigated the three-factor solutions for the data sets in order to examine in detail the eventual culture-dependent differences in the semantic

associations with polyphonic timbre. For the groups familiar with the stimulus sets, that is, II and WW, the variance explained by the third factor is remarkably larger than for groups IW and WI. This finding is in line with the observed differences in the intrinsic dimensionalities of the listener groups, further substantiating the role of enculturation in increasing the dimensionality of the perceptual timbre space. Investigating the factor loadings in detail revealed three main findings. First, the scale "Acoustic-Synthetic" was found to dissociate from the Activity factor and load onto the third factor as a result of cultural familiarity (i.e., for groups II and WW). This suggests that the perceptual dimension of "Acoustic-Synthetic" gets differentiated from the perceptual dimension of Activity as a result of perceptual learning via exposure to a musical style. Second, the scale "Soft-Hard" was found to load on to the third factor for group II only. The reason for this finding is not obvious, although it could be attributed to differences in the covariance structures of the acoustic features between the stimulus sets. Third, the scale "Empty-Full" loaded onto the Activity dimension for the Indian participants while it loaded onto the Brightness dimension for the Western participants. This finding suggests a culture-dependent difference in the semantic association of this perceptual scale.

In general, these results suggest that while the overall configurations of the two main perceptual dimensions of Brightness and Activity are similar, there exist minor differences in the factor structure for the third perceptual dimension, thereby providing additional support to our initial hypothesis that similarities are found across cultural backgrounds at a macro level, with differences being observed on a micro level. As hypothesized, we observed an increase in the dimensionality of perceptual timbre space for music from one's own culture. This lends support to the notion of prior exposure via enculturation affecting perceptual learning resulting in finer differentiation of dimensions.

Correlation analyses between acoustic features and perceptual dimensions revealed acoustic profiles that were similar across ethnic backgrounds and stimulus types. This similarity was more evident for Activity than

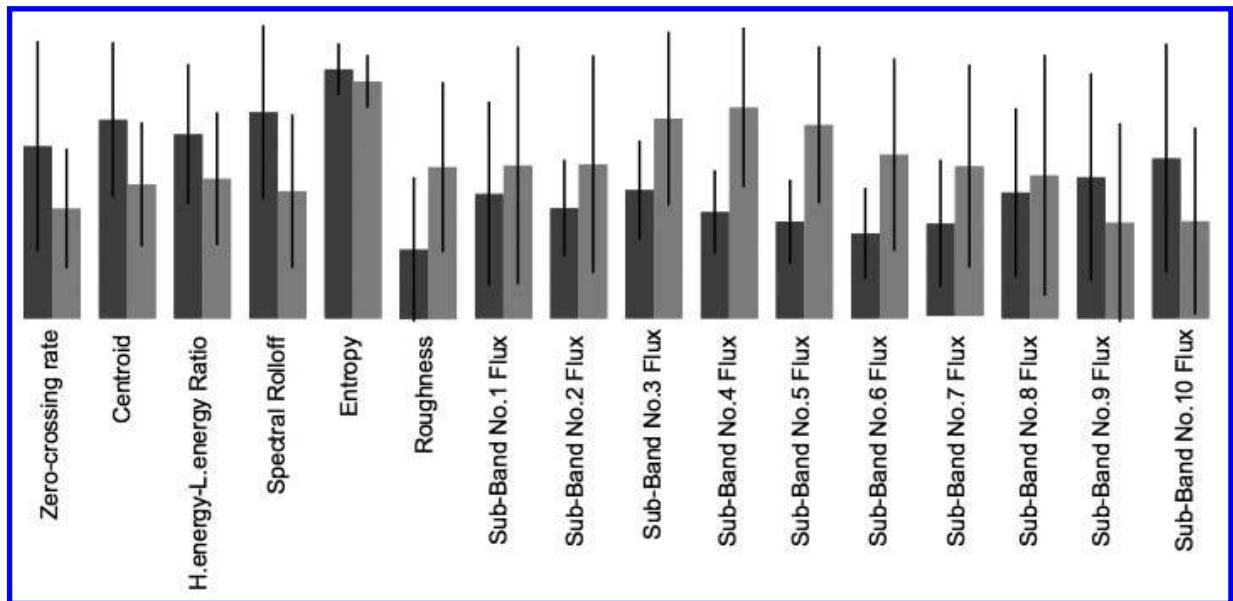


FIGURE 3. Means and standard deviations of acoustic features for both stimulus sets. Each feature is measured on a different scale and for many features the units are arbitrary. Hence, the ordinates (vertical axes) are omitted to facilitate comparison.

for Brightness. For Brightness, we observed greater consistency within stimulus sets than between them. This finding may be attributed to the differences in the distribution of acoustic features between stimulus sets.⁴ Figure 3 displays the means and standard deviations for each feature and stimulus set. As can be seen, features that correlate highly with Activity were found to have similar distributions across stimulus sets whereas this did not hold true for the Brightness dimension.

Nevertheless, the significant correlations observed between the acoustic profiles (see Table 4) suggest that the perception of polyphonic timbre with regard to its acoustic correlates involves common underlying mechanisms, specifically for the Activity dimension. Again, it is interesting to observe that the acoustic profiles are similar to those found in the study by Alluri and Toiviainen (2010), suggesting that the perception of these dimensions regarding their acoustic correlates is not largely altered by music training.

The high correlations between the sub-band fluxes and the perceptual dimensions imply that spectrotemporal modulations are important in the perception of polyphonic timbre, again concurring with the findings of Alluri and Toiviainen (2010). In addition, perceptual studies on monophonic timbre have previously reported spectral flux

as one of the acoustic correlates (Krumhansl, 1989; McAdams et al., 1995; Misdariis, Smith, Pressnitzer, Susini, & McAdams, 1998), although this finding has lacked consensus (Caclin, McAdams, Smith, & Winsberg, 2005). The present result is interesting in light of the proposition made by Carterette and Kendall (1999) regarding the significance of contrast involved in underlying mechanisms associated with perception. They postulated that detecting change or contrast is a fundamental process involved in human perception. Spectrotemporal modulations capture this very aspect of contrast in polyphonic timbre. Another commonly occurring acoustic correlate of monophonic timbre is log-attack time. However, in a polyphonic mixture, this feature is problematic to estimate computationally and might be perceptually less relevant due to the superposition and interleaving of various timbres.

In regression analyses, the models performed better for Activity than for Brightness. Contrary to Activity, the models of Brightness using PC regression show a slight improvement over those created using step-wise regression. This outcome may suggest that the Brightness dimension is more complex than Activity in the sense that it may rely on a conglomerate of acoustical features. For instance, selective attention to certain elements, such as the presence of a few notes in the high registers, may render a stimulus perceptually bright although the acoustic features that have been previously associated with perceptual brightness in the monophonic timbre domain, such as spectral centroid,

⁴ A series of *t* tests revealed significant ($p < .05$) differences for all the features between the stimulus sets with the exception of Sub-Band No. 8 flux.

might fail to capture this in a polyphonic mixture. In addition, it has also been argued that the spectral envelope is equally important as spectral centroid in distinguishing monophonic timbres (Hall & Beauchamp, 2009). Therefore, models of Brightness could be improved by using a larger combination of features.

To assess the generalizability of the regression models, we performed a cross-validation across the data sets. The high success observed in the predictability of the Activity dimension indicates the robustness of the present modeling approach. Furthermore, this finding suggests that the perceptual dimension of Activity possesses a certain degree of universality. Brightness models, on the other hand, displayed overall lower predictability than Activity models. Nevertheless, cross-validations within each stimulus set still yielded highly significant correlations, suggesting a certain degree of universality in the perception of this dimension.

A potential limitation of the study is that the listening experiments were conducted in English, while most of the participants were non-native English speakers. This can introduce “semantic noise” due to the variance in subjective understanding of the connotative meanings of words from person to person. It is noteworthy that similar

patterns in semantic associations were nevertheless found across all participant groups. Therefore, it can be assumed that semantic noise did not greatly affect the results.

A natural extension to this work would be to include Indian and Western musicians as participants in order to investigate the relative importance of enculturation and musical expertise on polyphonic timbre perception. In addition, investigating polyphonic timbre perception in the neural domain would allow us to gain a more holistic comprehension of the neural underpinnings of this phenomenon and the effect of cultural background thereon.

Author Note

We would like to thank Dr. Michael Hall for his valuable suggestions and comments. We would also like to thank Gayatri Prasad Rao, Dr. Olivier Lartillot, and Dr. Geoff Luck for their help. This research was supported by the Academy of Finland (project number 7118616).

Correspondence concerning this article should be addressed to Vinoo Alluri, Finnish Centre of Excellence in Interdisciplinary Music Research, Department of Music, PL 35(M), 40014 University of Jyväskylä, Jyväskylä, Finland. E-MAIL: vinoo.alluri@jyu.fi

References

- ADACHI, M., TREHUB, S. E., & ABE, J. (2004). Perceiving emotion in children's songs across age and culture. *Japanese Psychological Research*, 46, 322-336.
- ALLURI, V. & TOIVIAINEN, P. (2010). Exploring perceptual and acoustic correlates of polyphonic timbre. *Music Perception*, 27, 223-241.
- AUCOUTURIER, J. J. (2006). *Dix expériences sur la modélisation du timbre polyphonique* [Ten experiments on the modelling of polyphonic timbre] (Unpublished doctoral dissertation). Université Paris 6, France.
- AUCOUTURIER, J. J. & PACHET, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research*, 32, 83-93.
- AYABE-KANAMURA, S., SCHICKER, I., LASKA, M., HUDSON, R., DISTEL, H., KOBAYAKAWA, T., & SAITO, S. (1998). Differences in the perception of everyday odors - A Japanese-German cross-cultural study. *Chemical Senses*, 23, 31-38.
- BALKWILL, L.-L., & THOMPSON, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, 17, 43-64.
- BALKWILL, L.-L., THOMPSON, W. F., & MATSUNAGA, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, 46, 337-349.
- BEAUCHAMP, J. W. (1982). Synthesis by spectral amplitude and “brightness” matching of analyzed musical instrument tones. *The Journal of Audio Engineering Society*, 30, 396-406.
- BIGAND, E., VIEILLARD, S., MADURELL, F., MAROZEAU, J., & DACQUET, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, 19, 1113-1139.
- BOX, G. & COX, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26, 211-252.
- BRATTICO, E., TERVANIEMI, M., & PICTON, T. W. (2003) Effects of brief discrimination-training on the auditory N1 wave. *Neuroreport*, 14, 2489-2492.
- BURNS, B., SHEPP, B. E. (1988). Dimensional interactions and the structure of psychological space: the representation of hue, saturation, and brightness. *Perception and Psychophysics*, 43, 494-507.
- 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.
- CARTERETTE, E. C., & KENDALL, R. A. (1999). Comparative music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (pp. 725-791). San Diego, CA: Academic Press.

- CASTELLANO, M. A., BHARUCHA, J. J., & KRUMHANSL, C. L. (1984). Tonal hierarchies in the music of north India. *Journal of Experimental Psychology: General*, 113, 394-412.
- CURTIS, M. E., & BHARUCHA, J. J. (2009). Memory and musical expectation for tones in cultural context. *Music Perception*, 26, 365-375.
- DEMOREST, S. M., MORRISON, S. J., BEKEN, M. N., & JUNGBLUTH, D. (2008). Lost in translation: An enculturation effect in music memory performance. *Music Perception*, 25, 213-223.
- EEROLA, T., LOUHIVUORI, J., & LEBAKA, E. (2009). Expectancy in North Sami yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musica Scientiae*, 13, 39-70.
- EEROLA, T., & VUOSKOSKI, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39, 18-49.
- FLETCHER, M. L., & WILSON, D. A. (2002). Experience modifies olfactory acuity: Acetylcholine-dependent learning decreases behavioral generalization between similar odorants. *Journal of Neuroscience*, 22, RC201.
- GIBSON, E. J. (1953). Improvement in perceptual judgments as a function of controlled practice or training. *Psychological Bulletin*, 50, 401-431.
- GILBERT, C.D., SIGMAN, M., & CRIST, R.E. (2001). The neural basis of perceptual learning. *Neuron*, 31, 681-697.
- GJERDINGEN, R. O., & PERROTT, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, 37, 93-100.
- GOLDSTONE, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585-612.
- GREEN, S. B. (1991). How many subjects does it take to do regression analysis? *Multivariate Behavioral Research*, 26, 449-510.
- GREGORY, A. H. & VARNEY, N. (1996). Cross-cultural comparisons in the affective response to music. *Psychology of Music*, 24, 47-52.
- GREY, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61, 1270-1277.
- HALL, M. D., & BEAUCHAMP, J. W. (2009). Clarifying spectral and temporal dimensions of musical instrument timbre. *Canadian Acoustics*, 37, 3-22.
- HANNON, E. E., & TRAINOR, L. J., (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11, 466-472.
- HARWOOD, D. L. (1976). Universals in music: A perspective from cognitive psychology. *Ethnomusicology*, 22, 521-533.
- HELMHOLTZ, H. L. F. VON (1885/1954). *On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans. of the 1877 edition). New York: Dover.
- KAISER, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- KRUMHANSL, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzenand & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43-53). Amsterdam: Elsevier.
- KRUMHANSL, C. L. (2010). Plink: "Thin slices" of music. *Music Perception*, 27, 337-354.
- KRUMHANSL, C. L., TOIVAINEN, P., EEROLA, T., TOIVAINEN, P., JÄRVINEN, T., & LOUHIVUORI, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, 76, 13-58.
- LARTILLOT, O., & TOIVAINEN, P. (2007). A Matlab toolbox for musical feature extraction from audio. *Proceedings of the 10th International Conference on Digital Audio Effects*, Bordeaux, France, 237-244.
- LI, W., LUXENBERG, E., PARRISH, T., & GOTTFRIED, A. (2006). Learning to smell the roses: Experience-dependent neural plasticity in human piriform and orbitofrontal cortices. *Neuron*, 52, 1097-1108.
- MCADAMS, S., WINSBERG, S., DE SOETE, G., & KRIMPHOFF, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58, 177-192.
- MERRIAM, A. (1964). *The anthropology of music*. Bloomington, IN: Northwestern University Press.
- MEYER, R. K., PALMER, C., & MAZO, M. (1998). Affective and coherence responses to Russian laments. *Music Perception*, 16, 135-50.
- MISDARIHS, N., SMITH, B., PRESSNITZER, D., SUSINI, P., & MCADAMS, S. (1998). Validation and multidimensional distance model for perceptual dissimilarities among musical timbres. *Proceedings of the Joint Meeting of the 16th Congress on ICA and 135th Meeting of ASA*, Seattle, WA.
- MORRISON S. J., DEMOREST, S. M., AYLWARD E. H., CRAMER S. C., & MARAVILLAB, K.R. (2003). fMRI investigation of cross-cultural music comprehension, *NeuroImage*, 20, 378-384.
- NAN, Y., KNÖSCHE, T. R., & FRIEDERICI, A. D. (2006). The perception of musical phrase structure: A cross-cultural ERP study. *Brain Research*, 1094, 179-191.
- NAN, Y., KNOSCHE, T. R., ZYSSET, S., & FRIEDERICI, A. D. (2008). Cross-cultural music phrase processing: An fMRI study, *Human Brain Mapping*, 29, 312-328.
- PERETZ, I., GAGNON, L., & BOUCHARD, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, 68, 111-141.
- SMITH, L. D., & WILLIAMS, R. N. (1999). Children's artistic responses to musical intervals. *American Journal of Psychology*, 112, 383-410.
- TOIVAINEN, P., & EEROLA, T. (2003). Where is the beat? Comparison of Finnish and South African listeners. In R. Kopiez, A. C. Lehmann, I. Wolther, & C. Wolf (Eds.), *Proceedings of the 5th Triennial Escom Conference* (pp. 501-504). Hanover, Germany: Hanover University of Music and Drama.

- TREHUB, S. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 427-448). Cambridge, MA: MIT Press.
- TZANETAKIS, G., & COOK, P. (2002). Music genre classification of audio signals. *Institute of Electrical and Electronics Engineers. Transactions on Speech and Audio Processing*, 10, 293-302.
- VAN SELST, M., & JOLICŒUR, P. (1994). A solution to the effects of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, 47A, 631-650.
- ZACHAROPOULOU, K., & KYRIAKIDOU, A. (2009). A cross-cultural comparative study of the role of musical structural features in the perception of emotion in Greek traditional music. *Journal of Interdisciplinary Music Studies*, 3, 1-15.