MUSICAL SOUND TIMBRE: VERBAL DESCRIPTION AND DIMENSIONS

Jan Štěpánek

MARC — Musical Acoustics Research Centre Prague
Music Faculty, Academy of Performing Arts in Prague, Czech Republic
jan.stepanek@hamu.cz   http://www.hamu.cz/sound

ABSTRACT

Two approaches to the study of musical sound timbre are described and documented by psychoacoustic experiment examples.

The classical bottom-up approach is demonstrated on the study of contexts of violin sounds and pipe organ sounds. Verbal attributes collected during listening tests were used for the interpretation and comparison of resulted perceptual spaces of sounds.

The proposed top-down approach is based on the collection of musical experts experiences and opinions going from very common to more specific ones. Here the common perceptual space (perceptual space of verbal attributes) was constructed from non-listening test of dissimilarity of verbal attributes describing timbre (verbal or soundfree context of stimuli).

The verbal interpretation of perceptual spaces of sound contexts and perceptual space of verbal attributes are compared and the hypothesis of the four basic dimensions of timbre is formulated: 1. gloomy — clear, 2. harsh — delicate, 3. full — narrow, 4. noisy — ?.

1. INTRODUCTION

In the past the timbre research was focused on different aspects of this phenomena (overview see for example in [1]): timbre of steady state part of the sound, transient influence, different contexts — many kinds of natural sounds of musical instruments, synthetic sounds, respondent contexts (not frequently pointed out for example in relation to [2]), isolated tones and tone sequences, etc. Most frequently the goal was to explain used scales or to found dimensions using time and/or spectral properties of stimuli (sound signals) qualitatively or even quantitatively. Sometimes the effort to generalize the results in timbre research led to the changes of the sound context (used stimuli in [3] and [4]), or to the study more sound contexts in parallel [5].

Plomp in [6] recognizes two types of processing in hearing: bottom-up (audition) and top-down (cognition). Analogically, we will speak about bottom-up and top-down approach to the general study of timbre (unlike using the same terms for the description or classification of individual experiment, see [7]).

Bottom-up approach is based on the study of more contexts (of basic level) and then the selection of representative stimuli to create contexts of the second level, etc. Bottom-up approach is demonstrated here on the study of steady state timbre of five contexts of violin sounds (tones B3, F♯4, C5, G5 and D6) [5, 8, 9, 10] and five contexts of pipe organ sounds (Principal 8' contexts C2, C3, C4, C5 and C6 tones) [13, 14] (thus only basic level of bottom-up approach was provided up to now but on many sound contexts). Dissimilarity in timbre in pairs of sounds was acquired in listening tests of each sound context. The test results (individual dissimilarity matrices) were evaluated using latent class approach of multi-dimensional scaling (MDS) method (CLASCAL) [15, 16]. Perceptual space of appropriate dimensionality was obtained as an optimal MDS model for each sound context. The objects of each perceptual space represent test stimuli (sounds), dissimilarities between pairs of sounds were transformed into Euclidean distances. Verbal attributes (collected as spontaneous verbal description of timbre differences during listening tests) were used for the interpretation of resulted perceptual spaces (attributes in Czech for violin and in German for organ) and their comparison.

Top-down approach is based on the collection of musical experts experiences and opinions going from very common to more specific ones. Experts judged (without any listening) dissimilarity measure between pairs of verbal attributes describing timbre (attributes in the Czech language comes from questionnaire survey). Common perceptual space of verbal attributes was constructed using CLASCAL [17].

The results of above-mentioned studies were presented in several " conferences ([5, 8, 9, 10, 11, 12, 13, 14, 16, 17] and others) but never discussed together. The goal of the comparison of results of our studies presented here is the description of timbre via "common verbal" dimensions. They can be used not only for the interpretation of perceptual spaces, constructed from the results of listening test, but also for better communication among experts by the description of qualities of any kind of musical sounds and audio effects as well.

2. LISTENING EXPERIMENTS

Each stimuli context consisted of sounds of (nearly) the same pitch, level and duration.

2.1. Method

Listening experiments were focused on the description of timbre perception of quasistationary parts of sounds. Sound stimuli for tests were manipulated to decrease the influence of transients on the perception (uniform fade in and fade out).

The pairwise comparison listening tests were used for the collection of dissimilarities in timbre. Verbal attributes describing timbre dissimilarity in judged sound pairs were also collected (Spontaneous Verbal Description, SVD, [9]), in organs during each test or in violins for pairs of selected representative sounds.

External interpretation of perceptual spaces was made using collected verbal attributes. The number of use of verbal attribute in (optimal fitting of the external scale) [11, 18] was applied. Only successfully embedded verbal attributes (represented by significant reproduction of external scale values by embedding) were selected for further analysis. The angles between the embedding vectors were calculated [18] and the relations between the embedded verbal attributes were observed and discussed. Additional experiment was provided with violin sounds — Verbal Attribute Ranking and Rating (VARR) [8]. VARR test consisted in two parts supported by PC program: 1. ranking stimuli according verbal attribute, 2. rating the magnitude of verbal attribute on ranked stimuli.

2.2. Violin sound

2.2.1. *Contexts, stimuli and respondents*

Violin sounds of tones B3, F♯4, C5, G5 and D6 played the same technique (*détaché*, *naturale*, non-vibrato and *mezzoforte*) in anechoic room were recorded and prepared for tests as stimuli. The same loudness, pitch and tone duration was maintained during monophonic recording. Seventeen sounds for each tone were used in dissimilarity test (with twenty judges), after the selection based on their positions in constructed perceptual space eleven of them were used in verbal attributes collection SVD tests (ten judges). The test respondents were professors and students of violin play from Prague Academy. Verbal attributes were expressed in Czech language, but their approximate translation into English is also used here.

In VARR tests the same eleven sounds were used as stimuli, different eleven judges take part in tests. The selection of attributes was based on the results of dissimilarity tests (comparison of perceptual spaces and verbal attributes relations [5]). Verbal attributes *sharp*, *dark*, *clear*, *narrow* and also perceived sound quality were selected for the test and judged.

2.2.2. *Results*

Results related to verbal attributes in selected tones are summarized in Tables 1, 2.

Tone		B3	F♯4	C5	G5	D6
No. of dimensions		3	3	2	2	2
No. of attributes	$f_{abs} \geq 10$	65	64	58	64	65
	$\alpha \leq 5\%$	45	41	32	31	49
	$\alpha \leq 1\%$	27	28	22	12	31
	$\alpha \leq 0.1\%$	13	13	12	4	15

Table 1: *Number of perceptual space dimensions of optimal MDS solution and number of verbal attributes which have number of use at least* 10 *and embedding correlation significance at least* α.

The magnitudes of verbal attributes *sharp*, *dark*, *clear*, *narrow* and perceived sound quality received in VARR test were averaged over all judges' and evaluated using Factor Analysis (FA) in each tone. Two-dimensional factor space of attributes was optimal in all five tones [8]. The positions of attributes in factor space are shown in Figure 3. Successful prediction of perceived sound quality from these four attributes in each tone was also verified [8].

2.2.3. *Discussion*

The significant embedding of verbal attribute (number of it use on individual sounds — external scale) into perceptual space of sounds can be interpreted as follows: the feature represented by

Tone B3

Verbal attribute				
Czech	English	no. of use	rank	R_{embed}
temný	gloomy	144	3	0.97
hladký	smooth	45	17.5	0.95
úzký	narrow	105	6	0.94
mečivý	bleaty	87	9	0.93
tmavý	dark	208	1	0.92
kovový	metallic	99	7	0.90
ostrý	sharp	176	2	0.87
jemný	delicate	45	17.5	0.87

Tone C5

Verbal attribute				
Czech	English	no. of use	rank	R_{embed}
měkký	soft	121	6	0.98
jemný	delicate	108	7	0.97
kovový	metallic	76	9	0.96
ostrý	sharp	208	1	0.96
světlý	bright	143	3	0.94
kulatý	round	123	5	0.93
vysoký	high	45	16.5	0.92
tmavý	dark	152	2	0.92
pronikavý	penetrating	35	21	0.91

Tone D6

Verbal attribute				
Czech	English	no. of use	rank	R_{embed}
měkký	soft	134	2	0.98
ostrý	sharp	249	1	0.95
kulatý	round	92	7	0.94
šustí	rustle	95	6	0.92
hladký	smooth	59	12	0.92
tmavý	dark	60	11	0.90
pronikavý	penetrating	57	14	0.90
temný	gloomy	71	8	0.88
přidušený	damped	40	18	0.87
jemný	delicate	62	10	0.87
pískový	sandy	53	15	0.87
vysoký	high	38	19	0.86

Table 2: *Selected verbal attributes with number of use at least* 30 *and embedding significance* $\alpha \leq 0.1\%$*, sorted according to embedding correlation* R_{embed}.

the attribute has significant influence on perception and gives important contribution to the timbre dissimilarities on judged sound context. Moreover the significant embedding defines the direction along which this feature grows in perceptual space. Thus the features with embeddings containing small angles (near 0°) have similar influence on perception (their use on sounds is similar or proportional, they have coincident occurrence on sounds), large angles (near 180°) imply perceptually opposite and nearly orthogonal embeddings (near 90°) perceptually independent features.

The embedding of selected verbal attributes into perceptual spaces of optimal dimensionality are in Figure 1. The presented attributes were selected using following criteria: a) high number of spontaneous use in listening test, b) perceptual space filling, c) embedding success (significance level typically 0.1%). The attributes creating orthogonal system in sound-context free test (see section 3) were preferred. Since the sharpness was found as substantial feature of sound timbre [2], the presented comparison of perceptual features in individual violin tones will be related to embeddings of attribute sharp. Summary of angles contained between *sharp* and some other attributes is in Table 3. *Sharp* and *soft* are nearly opposite features in all five tones like *sharp* and *gloomy*, but *dark* changes the relative position according to *sharp* in tones G5 and D6. *Narrow* has similar influence with *sharp* in B3, F♯4 and C5, but is nearly independent in G5 and D6.

Verbal attribute	B3	F♯4	C5	G5	D6
soft	164	171	163	160	174
gloomy	160	163	167	171	159
dark	166	159	168	94	121
narrow	28	19	15	75	86

Table 3: *Angles [°] of embeddings of selected attributes contained with embedding of sharp.*

Attributes *sharp* and *dark* in factor space of VARR test results (Figure 2) have opposite positions in all five tones, the change of position *narrow* is more pronounced then in SVD embedding.

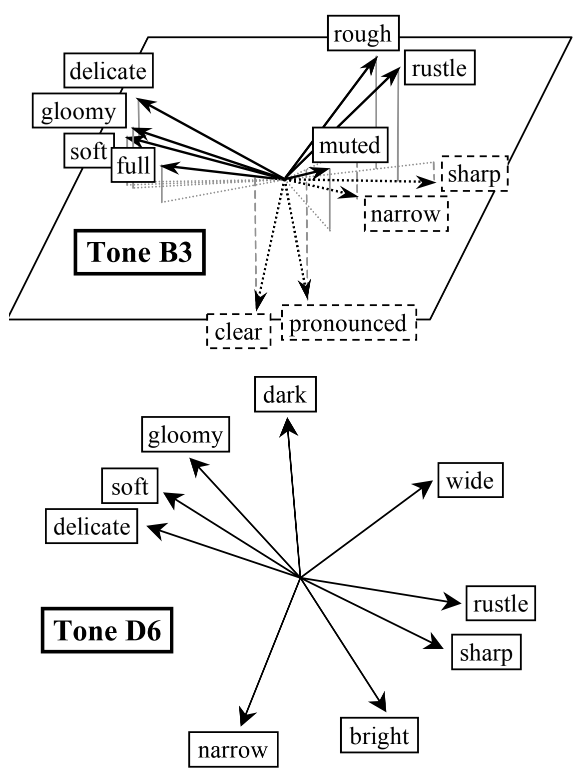

Figure 1: *Directions of embeddings of selected verbal attributes into perceptual spaces (tones B3, D6).*

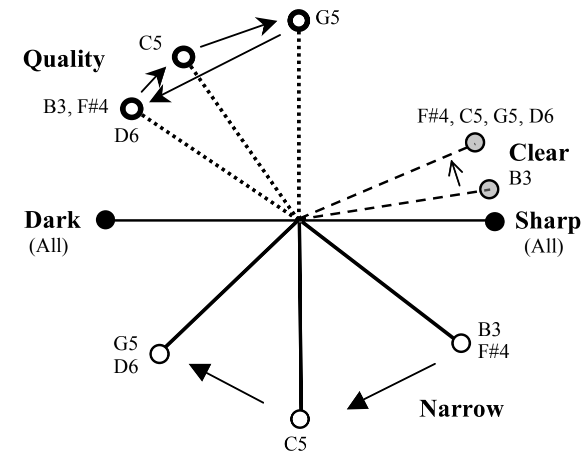

Figure 2: *Violin timbre: schematic factor space of verbal attributes and perceived sound quality.*

2.3. Pipe organ sound

2.3.1. Contexts, stimuli and respondents

Principal 8' sounds of tones C2, C3, C4, C5 and C6 of twelve organs from four European countries recorded *in situ* during the EU CRAFT project [13] were prepared for dissimilarity tests as stimuli. The tests were performed with twelve respondents — organ players and organ builders. Verbal attributes were expressed in German language, but their approximate translation into English is also used here.

2.3.2. Results

Results related to verbal attributes are summarized in Table 4. The three-dimensional perceptual space was identified as the optimal model for dissimilarity data for all studied C tones. Verbal attributes embeddings were calculated and representative attributes were selected based on following criteria: a) embedding significance at least 5presence in many C tones. For better comparison with other experiment results the attributes specific only for organ sound description were omitted: *streichend* (stringed in English), *obertönig* (rich on high harmonics), *flötig* (flute like), *prinzipalig* (principal like), etc.

2.3.3. Discussion

Angles contained between embedding of selected verbal attributes in three-dimensional space do not fill up the whole space, their relative positions can be simplified with satisfactory precision in the plane, see Figure 3. The perceptual spaces were rotated to fix the position of *round* for further view simplification. The perceptual features represented by *round* and *soft* have similar influence in all C tones. The relative positions of *round* and *noisy* is stable. The attributes *sharp* and *narrow* have similar positions and position changes relative to round.

3. MUSICIANS OPINIONS

The collection of musicians opinions was based on nonlistening experiments: a) collection of verbal attributes describing timbre

Verbal attribute	German	eng	rauschig	rund	scharf	weich
	English	narrow	noisy	round	sharp	soft
C2	no. of use	15	6	10	6	15
	rank	1.5	12.5	5	12.5	1.5
	R_{embed}	**0.86**	—	**0.85**	0.61	0.78
C3	no. of use	9	4	6	4	9
	rank	3.5	14	6.5	14	3.5
	R_{embed}	0.62	0.63	0.76	0.67	**0.94**
C4	no. of use	18	12	10	5	8
	rank	2	4.5	6.5	14	8.5
	R_{embed}	**0.95**	0.64	0.81	0.66	0.70
C5	no. of use	19	6	7	19	19
	rank	3.5	12.5	10	3.5	3.5
	R_{embed}	0.65	0.66	0.58	**0.86**	0.72
C6	no. of use	9	1	4	38	12
	rank	5.5	58.5	14	1	4
	R_{embed}	0.81	—	0.71	0.98	0.58

Table 4: *Selected verbal attributes describing Principal 8' sound timbre (embedding significance $\alpha \neq 5\%$, values of R_{embed} with $\alpha \leq 0.1\%$ are bold).*

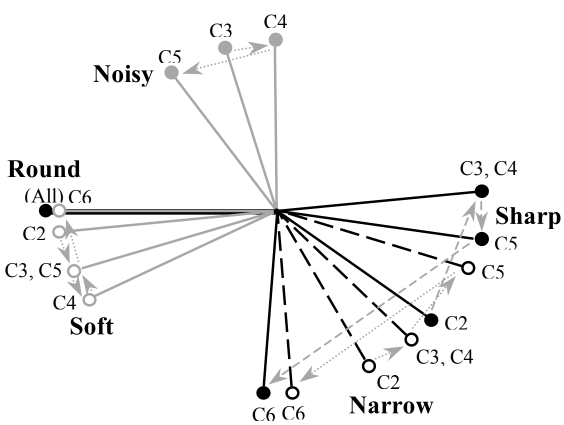

Figure 3: *Organ Principal 8' timbre: schematic perceptual space of selected verbal attributes.*

-questionnaire survey, b) test with selected verbal attributes in dissimilarity according to timbre understanding — attributes relations; then construction of common perceptual space of timbre using CLASCAL. The more detailed description is in [17].

3.1. Questionnaire survey

The questionnaire consisted of two parts: a) respondents personal profile (age, education, music profession and experience time, played instrument, etc.), b) words and expressions the respondent uses for the description of musical sound timbre. One hundred twenty respondents wrote 1964 different verbal attributes in total. The list of 25 most frequently used attributes is in Table 5.

3.2. Sound-context free test

3.2.1. Stimuli, respondents and methods

The 25 most frequently used verbal attributes from questionnaire (Table 5) were used as stimuli, 43 music professionals took part in the test (five professional groups: players of string, wind and key instruments, composers & conductors and sound designers).

Verbal attribute		Frequency	
Czech	English	f_{abs}	f_{rel}
ostrý	sharp	94	78.3
temný	gloomy	79	65.8
měkký	soft	78	65.0
jasný	clear	75	62.5
sametový	velvety	61	50.8
jemný	delicate	58	48.3
kulatý	round	58	48.3
tupý	unpointed	55	45.8
drsný	harsh	54	45.0
světlý	bright	54	45.0
tvrdý	hard	54	45.0
sladký	sweet	53	44.2
plný	full	51	42.5
hrubý	rough	46	38.3
tmavý	dark	46	38.3
teplý	warm	43	35.8
zářivý	radiant	42	35.0
čistý	pure	40	33.3
vřelý	hearty	40	33.3
barevný	colored	38	31.7
zvonivý	ringing	38	31.7
chladný	cool	36	30.0
průzračný	lucid	36	30.0
široký	wide	36	30.0
úzký	narrow	36	30.0

Table 5: *List of 25 most frequently used verbal attributes with their absolute(f_{abs}) and relative(f_{rel}) frequencies.*

Dissimilarity pair test (dissimilarities from 0 to 5) was provided, the respondent quantified his/her opinion on the dissimilarity in each pair of verbal attributes going from his/her internal image of timbre. The test results were evaluated using CLASCAL [15] and common perceptual space of verbal attributes was constructed.

The angles between attribute positions in the coordinate system of common perceptual space were calculated and the relations between them were qualified as follows [18]:

a) small angles ($\alpha \leq 20°$) $\Rightarrow$ similar,

b) nearly orthogonal positions ($70° \leq \alpha \leq 110°$) $\Rightarrow$ independent,

c) nearly opposite positions ($\alpha \geq 160°$) $\Rightarrow$ opposite meaning according to timbre.

The aim of the interpretation was to search for a system of verbal attributes, which completely describes the perceptual space and constitutes nearly orthogonal system.

3.2.2. Results

The three-dimensional common perceptual space as the optimal model for dissimilarity data and following nearly orthogonal system of verbal attributes were identified (see also Table 6 and Figure 4):

1. **temný** / tmavý — **jasný** / světlý	**gloomy** / dark — **clear /** bright
2. **drsný** / hrubý — **jemný**	**harsh /** rough — **delicate**
3. **plný** / široký — **úzký**	**full /** wide — **narrow**

Verbal attributes creating nearly orthogonal system in common perceptual space can be regard as an approximation of the new coordinate system of this space with dimensions defined by these attributes.

Angle [°]	gloomy	clear	harsh	delicate	full	narrow
gloomy	—		84	91	71	109
clear	176	—	97	89	106	75
harsh			—		92	77
delicate			159	—	106	85
full					—	
narrow					169	—

Table 6: *Angles between selected verbal attributes representing dimensions of common perceptual space.*

3.2.3. Discussion

The position of the most frequently used attribute *sharp* in perceptual spaces of all respondents and of individual professional groups [17] is similar and approaches to the plain of the first two dimensions in quadrant defined by attributes harsh and clear.

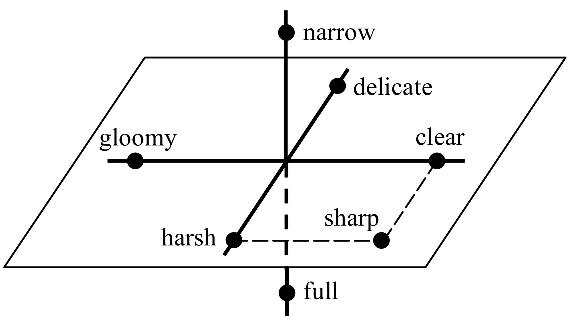

Figure 4: *Nearly orthogonal system of verbal attributes in three-dimensional common perceptual space of timbre.*

4. GENERAL DISCUSSION

The nearly orthogonal systems of verbal attributes fully explaining perceptual spaces constructed from results of listening tests of violin and organ sounds were not found. Let us try to look on these perceptual spaces (Figure 1–3) by monitoring positions of attributes of nearly orthogonal system describing common perceptual space of verbal attributes (Figure 4).

Attribute positions (when significant embeddings were found) in violin sound contexts are rarely opposite (1st dimension in violin tones F♯4, C5 and D6, 3rd dimension only in B3) or orthogonal (only some attributes in 2nd and 3rd dimensions in F♯4 and D6). The pair of attributes sharp — soft seems to be more prominent 'dimension' (see angles in Table 3), dimensional attributes of common perceptual space (Figure 4) surround this pair of attributes (Figure 1) in agreement with their positions according to sharp in common perceptual space: *clear*, *bright*, *harsh*, *rough* to *sharp*, and *gloomy*, *dark*, *delicate* to *soft*, but with small (acute) angles. The changing position of *narrow* is moreover supported in VARR test results (Figure 2). Attribute *rustle* with independent spectral source out of harmonic part of the sound in high violin tones [10, 12] is another candidate for orthogonal (independent) timbre dimension.

In organ sound contexts the features of *sharp* and *narrow* attributes have similar influence and *round* with *soft* and also *noisy* have strong influence on the perception of organ sound.

Many spontaneously pronounced verbal attributes in described experiments of timbre assessment are the words used in the description of other senses perception and are "borrowed" by hearing. Only *noisy* and *rustle* are originated in hearing but were not a part of timbre imagination in nonlistening test, probably due to musicians effort to avoid undesired disturbing of noise from the ideal concept of timbre.

This provokes to the formulation of hypothesis of the four basic dimensions of timbre: 1. *gloomy — clear* (vision, sense of sight), 2. *harsh — delicate* (touch), 3. *full — narrow* (volume, filling), 4. *noisy* / *rustle* — ? (hearing).

5. CONCLUSIONS

Musicians internal imagination of timbre supports orthogonal dimensions but their saliency or relationship in real sounds is sound context dependent (for example depends on pitch or type of the instrument — violin, organ, etc.). The imagination of the timbral features orthogonality probably comes just from the large span of their relations (contained angles) in real situations.

The independence of musical sound timbre features defined by discovered verbal attributes must be proved by the future search for their sound-invariant spectral resources (like it was started in [10, 12, 13]) and followed by their quantitative description (like it was done for sharpness and for some other features of other sound contexts). The knowledge of these resources can be used for the intended control of musical instrument sound and audio effects as well.

6. ACKNOWLEDGEMENTS

Research was supported partly by the Ministry of Education and Youth of the Czech Republic, project No. 1M6138498401 and by the EU CRAFT project G1ST-CT-2002-50267–DEMORGPIPE.

7. REFERENCES

[1] J. M. Hajda, R. A. Kendall, E. C. Carterette, and M. L. Harshberger, "Methodological issues in timbre research," in *Perception and Cognition of Music*, I. Deliège and J. E. Sloboda, Eds. Psychology Press, 1997, pp. 253–286.

[2] G. von Bismarck, "Timbre of steady sounds: A factorial investigation of its verbal attributes," *Acustica*, vol. 30, pp. 146–159, 1974.

[3] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Am.*, vol. 61, no. 5, pp. 1270–1277, 1977.

[4] J. M. Grey and J. Gordon, "Perceptual effects of spectral modification on musical timbres," *J. Acoust. Soc. Am.*, vol. 63, no. 5, pp. 1493–1500, 1978.

[5] J. Štěpánek, Z. Otčenášek, and A. Melka, "Comparison of five perceptual timbre spaces of violin tones of different pitches," in *CD-ROM of Joint Meeting 137th ASA, 2nd EAA Forum Acusticum 1999, 25th DAGA Berlin*, 1999, 5aMUb5.

[6] R. Plomp, *The Intelligent Ear*, 2nd ed. London: Lawrence Erlbaum Associates, 2002.

[7] M. Garnier, D. Dubois, J. Poitevineau, N. Henrich, and M. Castellengo, “Perception et description verbale de la qualité vocale dans le chant lyrique : une approche cognitive,” in *Journées d’Étude sur la Parole (JEP04)*, Fes, Maroc, 2004.

[8] J. Štěpánek, “Evaluation of timbre of violin tones according to selected verbal attributes,” in *32nd Int. Acoustical Conf., European Acoustics Association (EAA) Symp. "Acoustics Banská Štiavnica"*, 2002, pp. 129–132.

[9] ——, “The study of violin timbre using spontaneous verbal description and verbal attribute rating,” in *Forum Acusticum*, Sevilla, Spain, 2002.

[10] J. Štěpánek and Z. Otčenášek, “Interpretation of violin spectrum using psychoacoustic experiments,” in *CD of Proc. Int. Symposium on Musical Acoustics (ISMA2004)*, vol. 4-S2-1, Nara, Japan, 2004.

[11] J. Štěpánek, “Relations between perceptual space and verbal description in violin timbre,” in *acústica 2004*, 2004, CD ROM, AFP 077-S.

[12] J. Štěpánek and Z. Otčenášek, “Acoustical correlates of the main features of violin timbre perception,” in *Proc. Colloq. Interdisciplinary Musicology (CIM’05)*, Montréal, Québec, Canada, Electronic proceedings, 2005, [Online] http://www.oicm.umontreal.ca/cim05/cim05_articles/STEPANEK_J_CIM05_01.pdf.

[13] J. Štěpánek, Z. Otčenášek, V. Syrový, C. Täsch, and J. Angster, “Spectral features influencing perception of pipe organ sounds,” in *Forum Acusticum 2005*, Budapest, 2005, pp. 465–469.

[14] J. Štěpánek, V. Syrový, Z. Otčenášek, C. Täsch, and J. Angster, “Verbal description of organ principal 8’ sound,” in *32. DAGA, Braunschweig, in print*, 2006.

[15] S. Winsberg and G. De Soete, “A latent class approach to fitting the weighted euclidean model, CLASCAL,” in *Psychometrika*, vol. 58, 1993, pp. 315–330.

[16] J. Štěpánek and Z. Otčenášek, “Listener common and group perceptual dimensions in violin timbre,” in *Proc. Stockholm Music Acoustics Conf. (SMAC’03)* Stockholm, Sweden, 2003, pp. 663–666.

[17] J. Štěpánek and O. Moravec, “Verbal description of musical sound timbre in Czech language and its relation to musicians profession and performance quality,” in *Proc. Colloq. Interdisciplinary Musicology (CIM’05)*, Montréal, Québec, Canada, Electronic proceedings, 2005, [Online] http://www.oicm.umontreal.ca/cim05/cim05_articles/STEPANEK_J_CIM05_02.pdf.

[18] J. Štěpánek, “Interpretation and comparison of perceptual spaces,” in *Inter-Noise 2005*, Rio de Janeiro, 2005, CD-ROM, file: /papers/doc_1627.pdf.