Guimarães - Portugal
acústica 2004
paper ID: 077 /p.1

Relations between Perceptual Space and Verbal Description in Violin Timbre

J. Stepanek

Sound Studio of the Faculty of Music, Academy of Performing Arts Prague, Malostranské nám. 13, 11800 Praha 1, Czech Republic, e-mail: stepanek@hamu.cz

ABSTRACT: Perceptual spaces acquired from timbre dissimilarity listening tests are compared with verbs describing timbre of the same violin tones. Five sets of violin tones (pitch B3, F#4, C5, G5, D6) were used in the study. The results of dissimilarity listening tests with seventeen sounds for each pitch were processed using CLASCAL method and perceptual space for each pitch was obtained. Spontaneous verbal description was carried out on the representative subsets of original violin tones. Various approaches to qualification of relations between individual perceptual dimension or the whole perceptual space and verbal description represented by verbal attribute frequency of occurrence in spontaneous verbal description are used (perceptual dimensions qualification, verbal attributes immersion and projection into perceptual space) and compared. Perceptual spaces of different pitches are compared in virtue of their interpretation based on verbal description.

1. INTRODUCTION

An efficient and relatively uncomplicated method for the timbre study of a set of sounds is dissimilarity listening test. Listener's task may be just to quantify dissimilarity in timbre in each pair of sound stimuli. The result of dissimilarity listening test (dissimilarity matrix) can be evaluated using one of various types of Multidimensional Scaling method (MDS) which offers geometric representation of sound stimuli in a space (perceptual space). Now the problem arises how to interpret this perceptual space. In the study of one instrument type sounds (violins in our case) there is no *apriori* information about sounds (like it can be instrument type in classic test of sounds of miscellaneous musical instruments [1-3]. There are at least two possible resources for the external interpretation of perceptual space of sound stimuli: spectral (spectro-temporal) characteristics of sound signals or verbal descriptions of sounds. The interpretation of perceptual space using spectral characteristics is possible without any additional listening test but it represents difficult and usually not enough resultful task, namely in the case of very complex violin spectral envelopes [4].
Verbal description is dependent on the language used by judges (with well-known problem of authentic and appropriate translation into any other language) but it can offer very detailed and precise description of all main and most important features utilised by judges for dissimilarity assessment. It is possible to obtain verbal description immediately during the dissimilarity test or in additional listening test, ask the judge for spontaneous verbal description or let him to choose from in advance prepared list of verbal attributes.
As concerns methods how to use verbal description (or external data in common) for the interpretation of MDS solution (perceptual space) two basic approaches are usually used [5]:

1) Use of scales (perceptual dimensions as internal scales qualified by means of verbal description or optimally fitted external scales defined by verbal attributes).
2) Regional interpretation of perceptual space.
Application of both methods was used for the interpretation of perceptual spaces of violin timbre and is described later in this article.

2. METHOD

2.1 Previous experiments

Violin timbre was studied on five different pitches: **B3** (fundamental frequency 247 Hz, played on G string), **F#4** (370 Hz, D), **C5** (523 Hz, A), **G5** (784 Hz, E), and **D6** (1175 Hz, E). The same monophonic recordings of tones played on 24 violins of various qualities in *détaché*, *naturale,* non-vibrato, and *mezzoforte,* and recorded in an anechoic room were used in later described experiments. The same loudness, pitch and tone duration was maintained during the recording or later equalized. Attack and decay transients were subsequently manipulated to weaken their influence on perception. Stimuli (sounds) in tests were listened to using headphones. Two different experiments were carried out.
Experiment I was **dissimilarity listening test**. Seventeen of 24 most appropriate sounds for each pitch were used in the test. Twenty experienced listeners – students and professors of violin from Prague Academy assessed dissimilarities in timbre in pairs of violin tones. Results were evaluated using Latent Class MDS approach: weighted Euclidean model (CLASCAL) and extended CLASCAL [6]. Data most appropriate (best fits) models are summarized in Table 1, more detailed description of resulted **perceptual spaces** see in [7].

Table 1 – *Most appropriate latent class models (CiDjSk, i ... number of listener classes, j ... number of perceptual dimensions, k ... stimuli specificities: 0=No, 1=Yes).*

Pitch	B3	F#4	C5	G5	D6
Model	C2D3S0	C2D3S0	C2D2S1	C2D2S1	C2D2S1

Experiment II was **Spontaneous Verbal Description** of sounds (**SVD**), a modified free identification (verbal labeling) technique. Eleven of seventeen sounds best representing each perceptual space from the first experiment were used in tests. Listeners described perceived differences in timbre in all pairs of sounds spontaneously in words (verbal attributes). A separate test for each pitch was carried out with ten listeners (subgroup of listeners from the first experiment). The results of SVD of timbre were the **frequencies of occurrence** of each used verbal attribute on each sound (summed over the whole group of listeners). These verbal attributes were merged into five lists; only verbal attributes with number of occurrence of at least ten were evaluated. This represented 65 verbal attributes for the B3 pitch, 64 for F#4, 58 for C5, 64 for G5 and 65 for D6 pitch. The most frequently used verbal attributes are in Table 2 (see also [8]).
Frequencies of occurrence of verbal attributes acquired in the second experiment were used for external interpretation of perceptual spaces from the first experiment.

Table 2 - *Rank and number of occurrence of most frequently used verbal attributes in individual pitches, high ranks (1 – 9) are bold. The English translation of attributes is in the left column; the original Czech attributes are in parentheses.*

	B3	F#4	C5	G5	D6
sharp (ostrý)	**2** *(176)*	**1** *(185)*	**1** *(208)*	**1** *(174)*	**1** *(249)*
narrow (úzký)	**6** *(105)*	**3** *(107)*	**4** *(140)*	**2** *(151)*	**3** *(124)*
bright (světlý)	**4** *(130)*	**5** *(95)*	**3** *(143)*	**4** *(126)*	**5** *(114)*
dark (tmavý)	**1** *(208)*	**2** *(157)*	**2** *(152)*	10 *(68)*	11 *(60)*
soft (měkký)	**8** *(89)*	**4** *(96)*	**6** *(121)*	**8** *(75)*	**2** *(134)*
round (kulatý)	**5** *(119)*	**8** *(73)*	**5** *(123)*	**7** *(76)*	**7** *(92)*
gloomy (temný)	**3** *(144)*	**6** *(78)*	10 *(75)*	11 *(66)*	**8** *(71)*
clear (jasný)	15 *(48)*	10 *(71)*	**8** *(83)*	**3** *(146)*	**4** *(120)*
metallic (kovový)	**7** *(99)*	**9** *(72)*	**9** *(76)*	**9** *(72)*	**9** *(62)*
delicate (jemný)	18 *(45)*	**7** *(75)*	**7** *(108)*	15 *(50)*	10 *(61)*
voiced (znělý)	16 *(46)*	11 *(67)*	20 *(35)*	**5** *(86)*	13 *(58)*
rustle (šustivý)	14 *(48)*	12 *(51)*	30 *(24)*	18 *(44)*	**6** *(101)*
damped (přidušený)	26 *(33)*	23 *(38)*	13 *(58)*	**6** *(77)*	18 *(40)*
bleat (mečivý)	**9** *(87)*	24 *(31)*	27 *(27)*	56 *(11)*	61 *(10)*

2.2 Perceptual space interpretation using SCALES

Individual **perceptual dimension qualification** was acquired by using Pearson correlation coefficient between dimension coordinate of sounds and frequency of occurrence of verbal attribute on the same set of sounds (internal scale qualification [5]). The attributes with most significant positive or negative correlations were used for the interpretation of the dimension.

Verbal attribute immersion into N dimensional perceptual space was defined as direction in perceptual space, based on the optimal fitting of external scale defined by frequency of occurrence of verbal attribute in individual sounds [5]. External scale fitting is calculated using multiple regression formula (1) with verbal attribute as dependent variable y and dimension coordinates as independent variables $x_1, ..., x_N$:

$$y = b_0 + \sum_{n=1}^{N} b_n x_n \qquad (1)$$

Immersion is determined by direction cosines which are calculated using regression weights. Success of immersion is given by multiple correlation coefficient, its value is equal to the Pearson correlation between frequency of occurrence of verbal attribute and coordinates of the projection of sounds in perceptual space onto attribute immersion. So the most successful immersion reproduces well frequencies of occurrence, reproduction is expressed by coordinates of the projection.

2.3 Perceptual space interpretation using REGIONS

Regional interpretation of perceptual space is based on search of areas in perceptual space containing a group of sounds sharing some common property, in our case for example described with high frequency of occurrence of some specific verbal attribute. Due to huge

number of attributes used by listeners for the description of sounds in each pitch it was necessary to simplify the search for attributes of local meaning in perceptual space.
Verbal attribute projection into N dimensional perceptual space with M objects is defined as centre of gravity C of objects (sounds) in perceptual space with verbal attribute frequency of occurrence as masses or weights w. The calculation of individual coordinates of centre C is given by formula (2), where d denotes coordinate of object m in perceptual dimension n.

$$C_n = \frac{\sum_{m=1}^{M} w_m d_{mn}}{\sum_{m=1}^{M} w_m} \qquad n = 1, ..., N \tag{2}$$

Stability S of the projection is defined as a scattering of objects around centre of gravity C in perceptual space, weighted by frequency of occurrence, see formula (3), where l is the distance of the object m from the centre C.

$$S = \frac{\sum_{m=1}^{M} w_m l_m}{\sum_{m=1}^{M} w_m} \tag{3}$$

Vicinity of projections of verbal attributes was further investigated using hierarchical clustering. More stable projections of attributes offer the candidates of local meaning and thus helped for the search of interpretable regions in perceptual space.

3. RESULTS

More detailed description and graphing of results of perceptual dimension qualification, verbal attribute immersion and projection will be given for highest tested pitch D6.
The most appropriate MDS model for pitch D6 is two dimensional perceptual space. **Perceptual dimensions qualification** based on Pearson correlation coefficient is in Table 3.
Verbal attribute immersion is illustrated in Table 4, verbal attributes are ordered according to immersion success (given by correlation coefficient r). Optimal fitted coordinates for attributes *soft* and *high* are illustrated in Figure 1.
Verbal attribute projections for most frequently used verbal attributes (number of occurrence at least 20) are in Figure 2 (left part). The dendrogram of their hierarchical clustering (Figure 2 right) is the base for the division of perceptual space into regions (Figure 2). Most stable projections have attributes *flute like*, *velvety*, *strident*, *sandy*, *scrubs*, *smooth*, *delicate*, *round*, *wide* and *rustle*, which are the candidates for attributes with local meaning.
Examples of the distribution of occurrences of best immersed and local meaning attributes are in Figure 3.

Table 3 – *Pearson correlation coefficient r (significant at least on 1 % level) between dimension coordinates and frequency of occurrence of verbal attributes in pitch D6.*

Dimension D1, r>0		Dimension D1, r<0	
rustle (šustivý)	+0.91	**soprano-like** (sopránový)	-0.88
sharp (ostrý)	+0.88	**soft** (měkký)	-0.87
sandy (pískový)	+0.82	**round** (kulatý)	-0.84
dusty (zaprášený)	+0.80	**damped** (přidušený)	-0.83
scrubs (drhne)	+0.79	**delicate** (jemný)	-0.83
dirty (nečistý)	+0.74	**smooth** (hladký)	-0.82
		weak (slabý)	-0.81
		velvety (sametový)	-0.74
Dimension D2, r>0		**Dimension D2, r<0**	
dark (tmavý)	+0.90	**high** (vysoký)	-0.82
		tinny (plechový)	-0.79
		metallic (kovový)	-0.74
		whistle (pískavý)	-0.73

Table 4 – *Verbal attributes with most successful immersion scales into perceptual space in pitch D6. Multiple correlation coefficient r is significant at least on 0.1 % level. Angles in counter-clockwise direction starts from direction of positive first dimension axis.*

Attribute	r	Angle [°]	Attribute	r	Angle [°]
soft (měkký)	0.98	148	**gloomy** (temný)	0.88	133
sharp (ostrý)	0.95	334	**soprano-like** (sopránový)	0.88	177
round (kulatý)	0.94	150	**damped** (přidušený)	0.87	162
rustle (šustivý)	0.92	351	**delicate** (jemný)	0.87	162
smooth (hladký)	0.92	149	**sandy** (pískový)	0.87	338
dark (tmavý)	0.90	95	**tinny** (plechový)	0.86	295
penetrating (pronikavý)	0.90	315	**high** (vysoký)	0.86	254
hissy (syčí)	0.89	320			

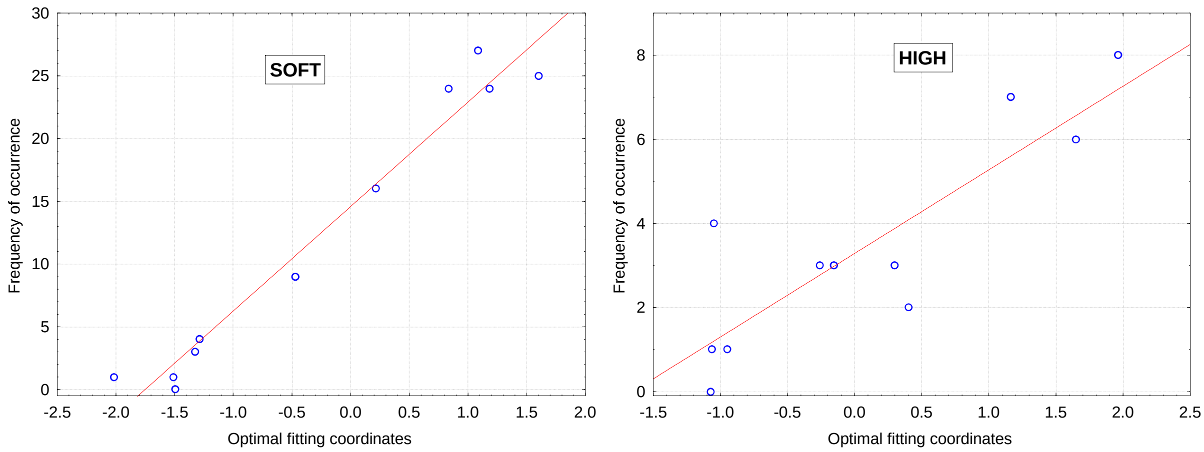

Figure 1 – *Scatterplots of immersion success for* ***soft*** *(r=0.98) and* ***high*** *(r=0.86) in pitch D6.*

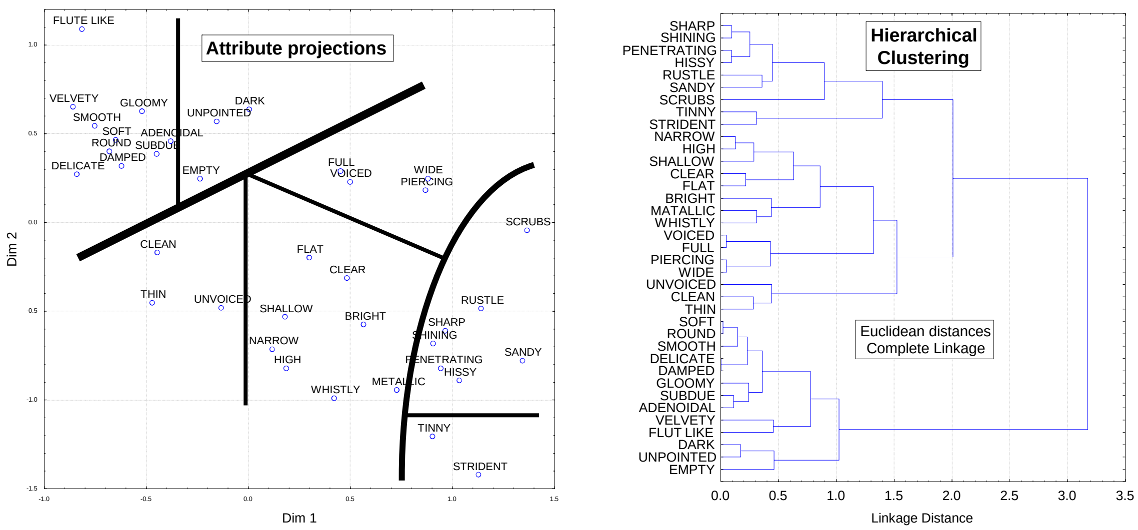

Figure 2 – *Projections of verbal attributes into perceptual space of pitch D6 (left) with regions based on dendrogram of hierarchical clustering of projection distances (right).*

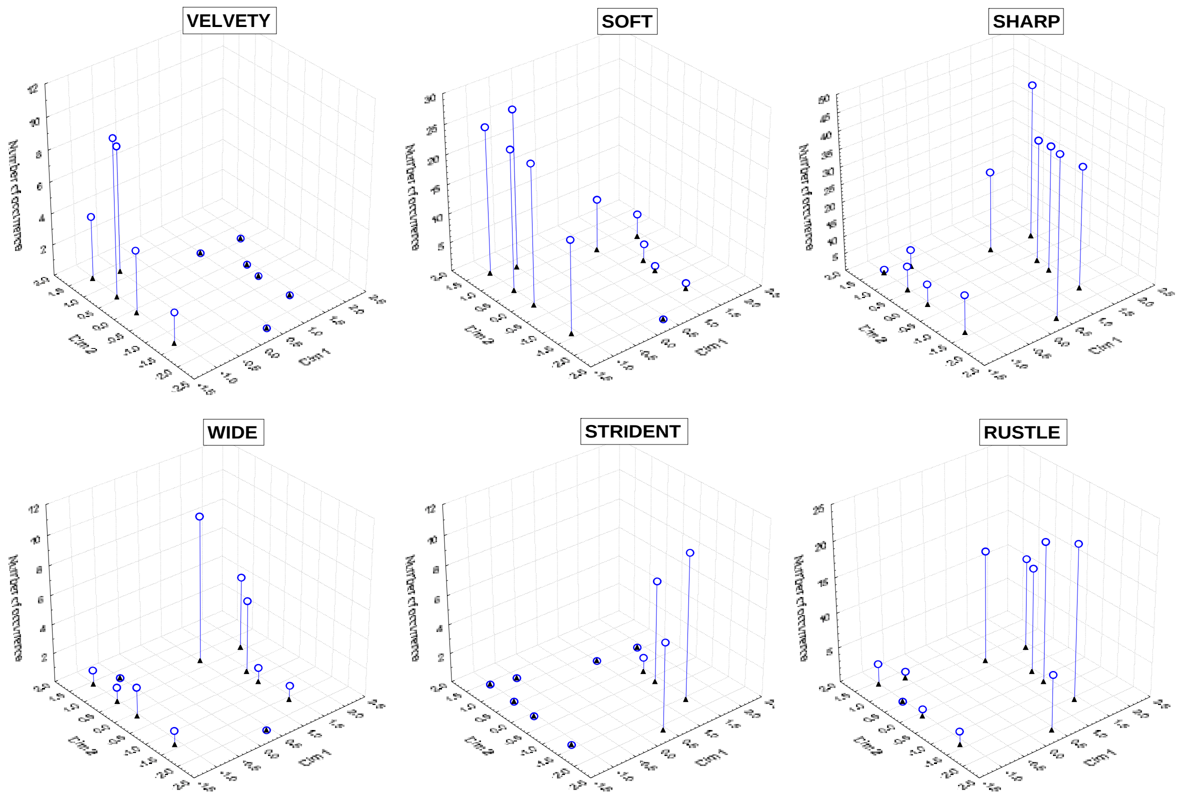

Figure 3 – *Distribution of occurrences of best immersed (**soft**, **sharp**) and local meaning (**velvety**, **wide**, **strident**, **rustle**) verbal attributes of pitch D6. Triangles represent configuration of sounds in 2-dimensional perceptual space, lines frequency of occurrence.*

4. DISCUSSION

4.1 Immersion and projection of verbal attributes into perceptual space (pitch D6)

Verbal attribute **immersion** into perceptual space can improve the statement of attribute about the space in two viewpoints according to perceptual dimension qualification:

1) To improve the embedding (expressed by correlation coefficient), for example in *soft* (from 0.87 to 0.98), or *penetrating* (from 0.71 to 0.90) – see also Table 3 and 4.
2) To unfold optimal direction of the change of attribute amount in perceptual space (*rustle* immersion has 9^0 deviation from the first dimension, *dark* 5^0 from the second dimension, *penetrating* 45^0 from both dimensions) – see Table 4.

Two other features can help for the perceptual space understanding:

1) Review of 'directional' filling of perceptual space according to all directions of successfully immersed attributes and establishing 'main' directions in perceptual space (around 150^0 or $150^0+180^0=330^0$) – see Table 4.
2) Search for similarity, polarity or ortogonality in directions of attributes immersion (*soft* and *round* contain the angle of 2^0, *soft* and *sharp* 174^0, *delicate* and *high* 92^0) – see Table 4. These findings can be base for the specification of more natural dimensions describing perceptual space.

Verbal attribute **projection** into perceptual space enabled to divide perceptual space into regions. Correlations between frequencies of occurrence of attributes from the same region provided significant values in most cases. Projection together with projection stability can help to discover attributes with more or less local presence in perceptual space (Figure 3).
Most stable projections emerged on the fringe of projections (Figure 2) and near to 'main' immersion direction.

4.2 Results of other studied pitches

Projections of verbal attributes into two dimensional perceptual spaces (pitch C5, G5) yielded to the most stable projections near 'main' immersion direction(s) and border of perceptual space, similarly to pitch D6. In three dimensional solutions (pitch B3, F#4) this phenomenon was not so strictly pronounced.
Only verbal attribute *soft* was immersed on 0.1 % significance level in all five pitches but the group of significantly (at least on 5% level) immersed attributes was relatively stable and similar. Pitch G5 emerged distinct from other studied pitches having only four verbal attributes with immersion on 0.1% level.

5. CONCLUSION

Immersion and projection of verbal attributes into perceptual space revealed to be beneficial tools for the perceptual space interpretation. The perceived variability in presence of successfully immersed or local attributes on listened sound context is virtually main source of judged dissimilarity. More detailed study of immersion angles can lead to the search for attribute categories and then to the finding verbally described 'psychological' dimensions better explaining perceptual space. Comparison of results for different sound contexts may even separate attributes with only apparent proximity.

ACKNOWLEDGEMENT

The research was supported by the Ministry of Education, Youth and Sports of the Czech Republic (project MSM 511100001).

REFERENCES

[1] PLOMP, R. (1976): *Aspects of tone sensation*. London, Academic Press.

[2] WESSEL, D. L. (1973): Psychoacoustic and music: A report from Michigan State University. *PACE: Bulletin of the Computer Arts Society*, 30, 1-2.

[3] GREY, J. M. (1977): Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61 (5), 1270-1277.

[4] STEPANEK, J. (2004): Spectral sources of basic perceptual dimensions of violin timbre. Proceedings of 7. CFA / 30. DAGA, Strasbourg, in print.

[5] BORG, I., GROENEN, P. (1997): *Modern Multidimensional Scaling*. Theory and Applications, Springer-Verlag, New York.

[6] MCADAMS, S., WINSBERG, S., DONNADIEU, 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.

[7] STEPANEK, J., OTCENASEK, Z. (2003): Listener common and group perceptual dimensions in violin timbre. Proceedings of the Stockholm Music Acoustics Conference (SMAC 03), Stockholm, 663-666.

[8] STEPANEK, J. (2002): The Study of Violin Timbre Using Spontaneous Verbal Description and Verbal Attribute Rating. Forum Acusticum Sevilla 2002 (ISBN 84-87985-06-8), MUS-06-008.