# Identification of perceptual qualities in textural sounds using the repertory grid method

Thomas Grill Austrian Research Institute for Artificial Intelligence (OFAI) Freyung 6/6 Vienna, Austria thomas.grill@ofai.at Arthur Flexer Austrian Research Institute for Artificial Intelligence (OFAI) Freyung 6/6 Vienna, Austria arthur.flexer@ofai.at Stuart Cunningham Centre for Applied Internet Research (CAIR) Glyndŵr University Mold Road, Wrexham, UK s.cunningham@glyndwr.ac.uk

# ABSTRACT

This paper is about exploring which perceptual qualities are relevant to people listening to textural sounds. Knowledge about those *personal constructs* shall eventually lead to more intuitive interfaces for browsing large sound libraries. By conducting mixed qualitative-quantitative interviews within the repertory grid framework ten bi-polar qualities are identified. A subsequent web-based study yields measures for inter-rater agreement and mutual similarity of the perceptual qualities based on a selection of 100 textural sounds. Additionally, some initial experiments are conducted to test standard audio descriptors for their correlation with the perceptual qualities.

## **Categories and Subject Descriptors**

H.5.5 [Sound and Music Computing]: Modeling

## **General Terms**

Experimentation, Human Factors

## **Keywords**

textural audio, auditory perception, verbal description, personal constructs, repertory grid, machine listening

# 1. INTRODUCTION

Today's electronic musician has a universe of sounds – synthetically generated or recorded – at his/her disposal, piled up in folder structures on one's computer hard-disk or stuffed into huge databases in the web. As a consequence of this huge amassment, a common problem is to find a specific sound for an application in composition, sound design or live performance in the vast forest of sounds available. This calls for efficient ways to organize sounds – strategies including manual semantic tagging [1] or using the various techniques from the field of Music Information Retrieval (MIR) to automatically *classify* and *cluster* sounds according to their digital audio content. Since manual tagging of large databases is very time-consuming, computer-based organization of data based on audio analysis almost suggests itself for many fields of application.

One approach to such automatic organization is to use some computational model for the sounds e.g. MFCCs, fluctuation patterns or a combination of standard audio descriptors (see e.g. [2, 3]) and apply techniques of unsupervised learning to cluster the modeled data in some low-dimensional (typically two-dimensional) space. This has e.g. been demonstrated with Pampalk's Islands of Music [4]. A drawback of this method is that the projection axes of the low-dimensional space are not pre-defined and/or not interpretable. When used as a visualization as part of a user interface, the systematics of organization consequently have to be learned by the user by means of exploration. These systematics can change when new data is introduced to the collection. Another approach is to find audio descriptors that are both technically feasible and meaningful to humans, and organize the sounds along the dimensions of those descriptors. Examples of such descriptors would be *pitch*, *noisiness*, *flux* etc. In this context a fundamental question arises: What are the most significant qualities of sounds that users of a musical interface (i.e. musician) would find intuitive and would want as organization criteria? The body of existing systematic research in this direction seems exclusively focused on timbre [5, 6, 7, 8, 9, 10] in relation to instrumental sounds, not elaborating on a more general scope of the sonic universe.

This research project is directed towards the discovery of common terminology to describe sounds. At present, we focus on textural sounds, that is, sounds that appear stationary – as opposed to evolving over time. This restriction seems useful in order to reduce the number of influence factors for a first approach. Textural sounds are interesting for musicians also because of their neutrality in respect to gestures – functioning as *sound material*. Temporal evolutions can always be additionally modulated onto the textural sound using volume, filter or pitch-shifting variations. At a future stage of research, qualities of temporal evolution shall be explored in a similar manner.

The paper is organized as follows: In section 2 we describe our use of the repertory grid method as a means to identify notions for describing qualities in sound, including some theoretical background, as well as the conduction and evaluation of listening experiments. In section 3 further evaluation on a larger scale is described, accompanied by quantitative analyses and an outlook to a potential application for novel musical interfaces. Section 4 concludes with a discussion of the findings and possible improvements.

# 2. REPERTORY GRID

The experience of sound and music is obviously a highly subjective one, leaving potential for ambiguity to occur. So, can we expect people to use a common language (resp. common systematics) when describing sounds? As a first step, we are concerned with finding out which could be important qualities for the description of sounds. Searching the web, there's already a substantial choice of 'adjectives to describe sounds', e.g. yielding notions like "loud, soft, silent, vociferous, screaming, shouting, thunderous, blaring, quiet, noisy, talkative, rowdy, deafening, faint, muffled, mute, speechless, whispered, hushed"<sup>1</sup> etc. However, there's potential for a researcher to influence the outcome of an experiment by the use of existing descriptions. Therefore, it seems advantageous to use a technique where the listener subjects may come up with their own descriptions of the elements under investigation. The repertory grid method (see [11, 12]) is just such a technique.

## 2.1 Background

The repertory grid technique is a tool that is based in Personal Construct Psychology (PCP) and implements Personal Construct Theory (PCT), best defined by George Kelly in 1955 [11]. As the name suggests, constructs are grounded in the psychological concept of constructivism, where human beings draw their understanding, descriptions, and theories about the world around them based upon their own interactions and personal experiences. More recently, the same principal techniques have been applied in a wide range of disciplines, especially where research crosses interdisciplinary bounds as is often the case in investigations into the roles of music and sound, such as in Picking's analysis of the computer technologies in music education [13] and Berg & Rumsey's study of spatial attributes of sound [14], both of which employ repertory grid analysis to elicit and understand people's experiences of their respective research domains.

Repertory grid is a recognized research method that employs a structured approach to elicit personal constructs from individuals. Typically this takes the form of a one-on-one interview between the researcher and subject. To begin a repertory grid investigation, a domain or topic of interest must be first defined. In our case the perceptual qualities of textural sounds were chosen as the domain of interest. Once the domain has been confirmed, particular instances, representative of that domain, must be defined. These instances are known in as *elements*. In the case of our study these are audio files containing textural sounds chosen by the first author of this paper.

The grid itself is built-up by the subject during the course of the interview which is used to elicit constructs from the subject. Constructs are determined by a reflective and comparative evaluation of the elements presented to the subjects.

<sup>1</sup>http://www.enchantedlearning.com/grammar/ partsofspeech/adjectives/, retrieved May 4th 2011 Constructs must be bi-polar, so the subject is required to define two terms that can be used at either end of a rating scale, which commonly has five or seven points. For example, suitable constructs for the rating of images might be: *light-dark*. The subjects are asked to describe the difference between two elements and then form a construct by articulating the opposite of that term. This method is known as straight differentiation. Another technique is known as triads which works by randomly selecting three elements from the set and asking the subject to group together the two elements they perceive as being most similar. By then questioning the reasons behind the grouping, and the difference between the two and remaining one element, bi-polar constructs can be elicited. After a bi-polar construct has been defined the subject is invited to rate all of the elements on that particular scale. This further removes ambiguity when a subject provides a rating, since the interrelation between the opposing ends of the scale have been specified by the subject themselves [11]. This process generates a matrix of ratings which represents the repertory grid for that subject. If the investigation employs multiple subjects the grids for each subject are concatenated at the end of the interviews to form a single, large repertory grid which represents the group's perceptions of the elements and therefore the domain of investigation.

The primary modes of investigation once the grid is complete can work on various levels depending upon the research question or hypothesis behind the investigation. Generally, these would take the form of analyzing constructs to identify descriptions of the elements and domain, where clustering of elements occurs and where clustering of constructs occurs. These can be identified using the numerical ratings. By using constructs and repertory grid analysis insight is gained into the constructs of the individual and the shared constructs of a group [11, 13]. Importantly, as Tan & Hunter [15] explain, shared constructs "... can yield information about group norms", which is of particular relevance to the clustering analysis and common constructs that are of interest for the purposes of this paper.

It is usual to interview a number of subjects, generally between 8 and 15 [13, 14, 15, 16]. This leads to a common criticism of the repertory grid technique, which is that it does not engage with enough subjects to make the data quantitatively meaningful. Whilst the main focus of the exercise is on elicitation of constructs, which are qualitative, repertory grids also provide quantitative robustness as the granularity of information comes from the number of constructs elicited and the numerical ratings provided for each. In essence, this makes it a mixed-methods technique, adaptable to a range of investigation scenarios in sciences, psychology and sociology.

# 2.2 Study methodology

When using the repertory grid method to elicit personal constructs, those constructs are strongly related to the specific items presented to the subjects. Hence, as for the project at hand we are interested in finding out general descriptions for textural sounds, our choice of sounds has to be general too, that is, it has to cover a variety of sonic characteristics as broad as possible. Since one subject should not be involved with the interview process for longer than about an hour, and preliminary tests showed that about 10–12 constructs can be expected, we limited the number of sounds to 20 in total. These sounds have been taken from a large collection of mostly abstract sounds used for electro-acoustic music performance of the first author. At a first step 100 sounds were selected from the library, fulfilling the criteria of being textural and not strongly exhibiting their provenience. The rationale for the latter is that a highly evident origin of sound production (e.g. recognizable materials, cultural or natural contexts etc.) could distract listeners from qualities related to the sound matter. This is strongly related to an *acousmatic* 'reduced listening' mode as formulated by Pierre Schaffer [17]. From the selection of 100 sounds (between 5 and 10 seconds of length), 20 sounds as diverse as possible were chosen as items for the repertory grid method. The sounds have also been normalized in regard to perceived loudness and mixed down to mono in order to minimize influences connected to the listening situation<sup>2</sup>. During the interview process, the sounds were played from a laptop computer using one monophonic speaker with reasonable quality placed at a distance of about one meter from the listener.

The subject listeners for the study have been recruited from the research and artistic contexts of the authors. 16 persons responded to our search request, aged between 23 and 45, among them 6 female and 10 male. By and large, those persons are all used to talking about sound and music and able to formulate their listening experiences, being either professional artists or researchers in sound. Prior to the elicitation, subjects were instructed to focus rather on the quality of the sounds than on potential origins (sound sources). The labels of the sounds have been anonymized in order to obscure potential semantic tagging or, again, potential sources<sup>3</sup>.

For each participant the process consisted of two parts. The first part was conducted as an interactive interview situation where the bi-polar personal constructs were elicited. In most cases *straight differentiation* was used, by presenting two different sounds and ask the person in which ways those sounds differ. In cases where this method proved unsuccessful, a third sound was introduced to form a *triad*. After all constructs have been found, the person was asked to rate all the available 20 sounds on the dimensions of one's personal construct, and 5 the other). The persons were asked to rate a construct for all sounds before proceeding to the next construct, thus staying in the same reference frame. The two parts of the process took no longer than 30 minutes each.

#### 2.3 Filtering constructs

From the 16 participants a total of 202 bi-polar constructs for 20 textural sounds have been elicited, yielding a repertory grid of  $202 \times 20$  elements. As shown in [12] several standard analyses can be performed on the data, including Principal Component Analysis (PCA) and hierarchical clustering (therein called FOCUS Cluster analysis).

Hierarchical clustering and its visualization in the form of a dendrogram can help to identify groups of constructs that have been evaluated on the items in a similar way, that is, have been used synonymously. The method is based on a distance measure (in our case the Euclidean metric) computed on vectors of item evaluations of the individual constructs in the repertory grid. We used the *hcluster* [18] module for Numerical Python in its 'complete-linkage' mode.

Using the clustering shown in Figure 1 it was possible to identify four clusters of constructs that exhibit a close linkage. We decided to use alternative wordings to outline the semantic range within each cluster – see Table 1, upper half, for the German constructs and their English equivalents. It is obvious that also outside those relatively well-defined clusters many constructs appear in semantically identical and varied forms. Hence, we additionally collected those constructs that appear repeatedly in the grid and tried to find distinct terms to represent the somewhat blurry clouds of semantical variations. This leads to the additional constructs, listed in the lower half of Table 1.

Table 1: Major constructs filtered from the 202 personal constructs elicited by use of the repertory grid method. The upper half contains those constructs that have been retrieved using a hierarchical clustering method, the lower half lists constructs that have been identified by semantic grouping. Constructs in the same cell are used synonymously.

Constructs	Constructs	
(German original)	(translated to English)	
hoch-tief	high–low	
hell–dunkel	bright-dull	
regelmäßig-unregelmäßig	ordered-chaotic	
geordnet-chaotisch	coherent-erratic	
glatt-rau	smooth-coarse	
weich-rau	soft-raspy	
natürlich-künstlich	natural-artificial	
analog-digital	analog-digital	
statisch-dynamisch	static-dynamic	
starr-bewegt	rigid-eventful	
nahe–fern	near–far	
klar-verschwommen	clear-blurred	
kantig-rund	edgy-flowing	
zerrissen-kompakt	disjointed-continuous	
dicht-spärlich	dense-sparse	
flächig-punktuell	expansive-selective	
homogen-heterogen	homogeneous-heterogeneous	
gleichförmig-differenziert	uniform-differentiated	
tonhaltig-geräuschhaft	tonal–noisy	

Since most constructs have been elicited using the German language (by 13 out of 16 subjects) the English equivalences have been worked out with the help of a native bilingual translator who is specialized in artistic projects. Although we triggered the subjects to concentrate on the quality of sound rather than on potential sources, the construct *natürlich-künstlich* (*natural-artifical*) seems to have been unavoidable and appears prominently in the repertory grid and therefore also in the resulting filtered constructs.

<sup>&</sup>lt;sup>2</sup>This seems to have been successful since no constructs involving loudness or volume, respectively directionality of sound have been retrieved.

<sup>&</sup>lt;sup>3</sup>The sounds are available at http://grrrr.org/test/ classify/grid.html.php



Figure 1: Dendrogram resulting from hierarchical clustering of the elicited personal constructs evaluated on 20 textural sounds. The dissimilarities between the individual constructs have been calculated using Euclidean distances for a rating scale from 1 to 5. The l.h.s. part of the plot is continued on the r.h.s.

# 3. EVALUATION OF CONSTRUCTS

Given 10 bi-polar constructs uncovered using the repertory grid method, we carried out further evaluation targeted towards three main purposes: First, to test the constructs on a larger, more general body of textural sounds than just the 20 examples used in the elicitation process. Second, to find out which are the more important, respectively, most unambiguous constructs. Third and final, because some of the filtered constructs were compiled using the researcher's common sense and imagination, there is the need to test reliability of that filtering process on the agreement among a larger number of listeners.

# 3.1 Study methodology

In order to collect a (for statistical reasons) desirable, substantial amount of data we decided to design a web-based survey that we could send out to interested individuals and forums. The web form programmed in HTML/PHP with an SQL database backend (see Figure 2)<sup>4</sup> allows to first choose



#### Figure 2: Online survey for the rating of constructs on a random sound (out of a selection of 100).

English or German as the preferred language and enter a unique ID (e.g. an email address) to allow the user to pause the process and come back at a later time. We deliberately refrained from asking for more user-centered data, like age, sex, profession, mother-tongue etc. since we wanted to keep the entry barrier (e.g. by having to enter private data) as low as possible, and such data from an online form would not be very trustworthy anyway.

After that initial step, one random anonymous sound (out of 100 available) is presented to the listener who is asked to

grade it on the dimensions of the 10 filtered constructs. If a construct seems inappropriate or ambiguous for a sound, the user can leave it out. As in the course of the repertory grid ratings subjects have often chosen to use finer grading than just integer steps on the scale from 1 to 5, the online form features 9 steps from one pole of the construct to the other one. On pressing the 'submit' button the rating is submitted to an SQL server, along with information on the chosen language, user ID, time/date and eventual additional remarks given by the user.

It has been observed from the data that different users tend to exhibit considerably differing rating habits: some orient themselves along the middle of the rating scale, while others like to exploit the extremes of the scale. In order to account for that, the grades have been normalized to unit variance, individually for each user. Up to the current stage of the project no advanced data filtering (e.g. concerning outliers) is being performed. The only sanity check (also for reasons of per-user variance normalization) is that users have to rate at least 10 sounds to be included into the evaluation, to eliminate potentially meaningless ratings that result from pure curiosity or fun submissions.

We announced the web survey on several specialized mailing lists, including 'music-ir'<sup>5</sup>, 'auditory'<sup>6</sup>, 'Pd-ot'<sup>7</sup>, as well as mailing lists for local research and artist forums, with a potential total audience of several thousand persons. Although the distribution is much broader than with the one-on-one interviews in section 2, subjects completing the survey would still be mostly in research or artistic contexts with a specific interest for sound and music. It goes without saying that the listening situation for each subject is beyond our control – however, as we are targeting a certain generality, we accept this variance to be part of the various influence factors.

# 3.2 Results

Up to now<sup>8</sup> 104 subjects have contributed data to our study, with 59 rating at least 10 sounds. From those, 35 subjects chose the German version of the form, and 24 the English version. By those 59 filtered subjects, a total of 16808 constructs has been rated on 1796 playbacks of the available sounds. This means that on average 9.4 constructs have been rated per sound – the possibility of omitting inappropriate constructs for sounds has not been frequently used. On average, each construct has been rated more than 16 times on each sound, with a minimum of 10 and a maximum of 33 rates. Seven subjects rated all 100 available sounds.

Figure 3 shows all the 100 exemplary sounds ordered according to mean grades for the construct *ordered-chaotic*. The standard deviations of the grades are plotted as bars centered at the mean values. As can be observed also for most of the other constructs, the range of grades has been applied quite evenly which indicates an even distribution of characteristics over the sound selection. The only construct

<sup>&</sup>lt;sup>4</sup>Available online at http://grrrr.org/test/classify

<sup>&</sup>lt;sup>5</sup>http://listes.ircam.fr/wws/info/music-ir, retrieved May 12th 2011

<sup>&</sup>lt;sup>6</sup>http://www.auditory.org, retrieved May 12th 2011

<sup>&</sup>lt;sup>7</sup>http://lists.puredata.info/listinfo/pd-ot, retrieved May 12th 2011

<sup>&</sup>lt;sup>8</sup>July 1st 2011, the web survey being still online.

that sticks out a bit in this respect (not shown as a figure) is *tonal–noisy* where the distribution is slightly warped to the noisy side of the scale. For all of the constructs standard deviations tend to be lower in the extremes – indicating unambiguity – and higher close to the center of the scale.



Figure 3: Normalized (per subject) ratings of all 100 sounds for the construct *ordered-chaotic*. Means (dots) and standard deviations (bars) are shown.

An important evaluation concerns inter-rater agreement, that is, how well individual subjects correspond in their ratings. We used Krippendorff's alpha [19] for interval data to evaluate the agreement of the ratings<sup>9</sup>. A value of 1 would indicate perfect agreement, while 0 signifies statistical unrelatedness. The results are listed in Table 2 for two different groups of raters: Firstly, for a core group of nine subjects who already took part in the repertory grid interviews. And secondly, for all 59 subjects who rated at least 10 sounds using the web-survey. It is immediately obvious that for every construct the measure of mutual agreement for raters that Table 2: Inter-rater agreement (Krippendorff's alpha) among raters, calculated for each construct, respectively. The left results column shows the agreements among the core group of nine subjects who already took part in repertory grid interviews, the right one the agreements among all 59 subjects who rated at least 10 sounds. Constructs are ordered by decreasing agreement – top ones have been rated more consistently than those at the bottom.

Construct	Agreement	Agreement
	$\alpha$ (core group)	$\alpha(\text{all } n \ge 10)$
high-low	0.588	0.519
ordered-chaotic	0.556	0.447
natural-artificial	0.551	0.492
smooth-coarse	0.527	0.420
tonal-noisy	0.523	0.435
homogeneous-heterogeneous	0.519	0.416
dense-sparse	0.492	0.342
edgy-flowing	0.465	0.376
static-dynamic	0.403	0.383
near-far	0.252	0.249

took part in the elicitation process is consistently higher than for a more general audience. This is not surprising, also because of the fact that currently no substantial filtering is being performed on the online data and the smaller group of people seems more reliable because of personal contact during the repertory grid phase. Nevertheless the ranking of the constructs in respect to inter-rater agreement is consistent, indicating the constructs high-low, orderedchaotic, natural-artificial, smooth-coarse, tonal-noisy and homogeneous-heterogeneous to be shared among subjects of both groups with sufficient agreement. Note that the computed values here are generally lower than the threshold value of 0.667 recommended by Krippendorff [19]. This can be explained by the complexity and subjectivity of the underlying problem, whereas in the measure's original domain of 'content analysis' predominantly true rater errors account for disagreement.

Since each subject may choose either German or English as the language for the web survey, checking the correlation of ratings between both languages seems worthwhile. Table 3 shows inter-rater agreement of average ratings for each sound carried out using the English or German terms for the constructs, respectively. It is worth pointing out that for the constructs showing high inter-rater agreement in Table 2, agreement in the usage of the terms in both languages is very high.

Another interesting evaluation can be performed on correlations concerning the subjects' usage of the individual constructs. Figure 4 shows a self-similarity matrix depicting the pairwise Pearson correlations of the constructs. For each pair of constructs between 1546 and 1648 ratings are taken into account, resulting in very low thresholds for statistical significance. Apart from the diagonal some considerable correlations between constructs can be found which indicate certain amounts of ambiguity: -0.60 for smooth-coarse vs. edgy-flowing, 0.58 for static-dynamic vs. homogeneous-heterogeneous, 0.56 for

<sup>&</sup>lt;sup>9</sup>normalized to unit variance for each subject

Table 3: Agreement (Krippendorff's alpha) of average ratings for each sound between raters using the English and German versions of the web survey, calculated for each construct, respectively.

Construct	Agreement $\alpha$
natural-artificial	0.871
high–low	0.859
smooth-coarse	0.852
tonal-noisy	0.843
ordered-chaotic	0.831
homogeneous-heterogeneous	0.831
near-far	0.735
dense-sparse	0.706
static-dynamic	0.608
edgy-flowing	0.593

homogeneous-heterogeneous vs. ordered-chaotic, 0.54 for static-dynamic vs. ordered-chaotic, and 0.51 for smooth-coarse vs. tonal-noisy. On the other hand, some constructs don't correlate substantially with any other, specifically natural-artificial (maximum correlation is 0.16), near-far (0.26), and high-low (0.26).



Figure 4: Pearson correlation between constructs, evaluated on the data of all subjects having rated at least 10 sounds. The smallest significant correlation value (at  $\alpha = 0.05$ , two-tailed) is  $\pm 0.049$ .

As a final step the retrieved constructs have been tested against some standard audio descriptors commonly used in Music Information Retrieval (MIR). This evaluation does not claim to be exhaustive in any way, it is presented merely to show the potential of future application. A future goal would be to find computable audio descriptors that correlate well with each (and exactly one) of the constructs, thereby allowing automatic classification of sounds according to perceptually oriented constructs. For the time being we simply tried out all of the scalar audio descriptors available in the YAAFE (yet another audio feature extractor) library<sup>10</sup>. It contains many of the standard features like *spectral slope*, *spectral flux*, zero crossing rate, spectral centroid etc. We calculated the means of the descriptors over the duration of the individual sound files, using a frame length of 1024 samples and a hop size of 512 samples (at 44100 Hz sample frequency).

Figure 5 shows measures of correlation for each of the constructs versus the various audio descriptors in the YAAFE library<sup>11</sup>. Again we used Pearson correlation, in this case



Figure 5: Pearson correlation between perceptual constructs (rows) and computed audio features (columns). The smallest significant correlation value (at  $\alpha = 0.05$ , two-tailed) is  $\pm 0.19$ .

on the mean ratings of the perceptual constructs, and on the mean frame-based YAAFE descriptors. Both data sets have been whitened in regard to overall mean and standard deviation. The highest correlations can be found for the construct *high-low* which has a high negative correlation of -0.76 with *perceptual sharpness*, and also high correlations of -0.69 with the spectral centroid (0th moment of the SpectralShapeStatistics descriptor) and -0.69 with spectral slope. This makes perfect sense since all those descriptors measure the shape of a spectral distribution which corresponds to the semantic meaning of *high-low*. Other constructs also exhibit considerable correlations to individual descriptors, notably smooth-coarse (-0.63 to spectral skewness which is the 2nd moment of the *SpectralShapeStatistics* descriptor), *near*-far (-0.62 to spectral centroid and spectral slope), and edgy-flowing (-0.61 to spectral flatness). The fact that spectral centroid and spectral slope correlate well with a number of constructs motivates the construction of specifically sensitive descriptors (or combinations thereof) in order to be

<sup>&</sup>lt;sup>10</sup>http://yaafe.sourceforge.net, retrieved May 12th 2011 <sup>11</sup>Labels with numerical appendices indicate moments of a statistical distribution, being *centroid*, *spread* and *skewness*, respectively, for indices 0 to 2.

able to discriminate those perceptual qualities. Some constructs like *natural-artificial*, *homogeneous-heterogeneous*, *static-dynamic*, *ordered-chaotic* and *dense-sparse* cannot be reasonably modeled by any of the audio descriptors. The reason seems to be that none of those constructs refer to timbral, but predominantly rather to temporal qualities, which are not covered by the selection of descriptors. As mentioned earlier, the construct *natural-artificial* seems to be connected more to contextual information (regarding the source of sound production) than to sonic qualities, hence we would expect finding corresponding audio descriptors to be virtually impossible.

## 4. CONCLUSION

By conducting systematic listening tests with the repertory grid method, we were able to elicit ten major bi-polar constructs describing qualities of a varied selection of textural sounds. Subsequent data collection employing a larger scale online survey with a selection of 100 sounds and over 100 participants resulted in measures for the inter-rater agreement of each of those constructs, and also yielded an order of relevance and measures for the similarity of the constructs. As a first outlook into further research envisioned using those findings we looked into the correlation between the resulting constructs and some standard MIR-style audio descriptors. The fact that mainly timbre-related qualities are covered by those audio descriptors leads the way to future exploration for suitable audio descriptors capable of modeling also temporal, dynamic, etc. qualities.

It should be noted that currently no substantial filtering is being performed on the data retrieved from the online survey. We cannot expect all inputs to be reliable, therefore some more sanity checks should be done, e.g. to identify outliers resulting from joke submissions or functionality tests. The rare use of the possibility to omit grading inappropriate qualities for individual sounds (obviously often resulting in grading close to the center of the scale) also calls for some respective filtering strategy. It seems also worthwhile to look into more advanced methods for data normalization, e.g. using histogram equalization per user and per construct.

## 5. ACKNOWLEDGMENTS

This work has been supported by the Vienna Science and Technology Fund WWTF (Audiominer, project number MA09-024). Many thanks to the reviewers for their valuable comments!

#### References

- D. Turnbull, L. Barrington, and G. Lanckriet, "Five approaches to collecting tags for music," in *Proceedings of the 9th International Conference on Music Information Retrieval*, (Philadelphia, USA), pp. 225–230, 2008.
- [2] P. Herrera, X. Serra, and G. Peeters, "A proposal for the description of audio in the context of MPEG-7," 1999. MPEG.
- [3] P. Herrera, X. Serra, and G. Peeters, "Audio descriptors and descriptor schemes in the context of MPEG-7," in *Proceedings of ICMC*, 1999.
- [4] E. Pampalk, "Islands of music: Analysis, organization, and visualization of music archives," Master's thesis,

Vienna University of Technology, Vienna, Austria, December 2001.

- [5] K. Ueda, "A hierarchical structure for adjectives describing timbre," Journal of the Acoustical Society of America (JASA), vol. 100, p. 275, October 1996.
- [6] G. Darke, "Assessment of timbre using verbal attributes," in *Proceedings of the Conference on Interdisciplinary Musicology (CIM)*, (Montréal, Canada), March 2005.
- [7] C. Nicol, Development and Exploration of a Timbre Space Representation of Audio. PhD thesis, Department of Computing Science, University of Glasgow, 2005.
- [8] C. G. Johnson and A. Gounaropoulos, "Timbre interfaces using adjectives and adverbs," in *Proceedings of* the 2006 conference on New interfaces for musical expression, NIME '06, (Paris, France, France), pp. 101– 102, IRCAM — Centre Pompidou, 2006.
- [9] J. Štěpánek, "Musical Sound Timbre: Verbal Description and Dimensions," Proceedings of the 9th International Conference on Digital Audio Effects (DAFx06), Montreal, Canada, September 18–20, 2006, September 2006.
- [10] M. Sarkar, B. Vercoe, and Y. Yang, "Words that describe timbre: a study of auditory perception through language," in *Language and Music as Cognitive Systems Conference (LMCS-2007)*, (Cambridge, UK), May 2007.
- [11] G. A. Kelly, The Psychology of Personal Constructs. New York, NY, USA: Norton, 1955.
- [12] S. Cunningham, "Applying personal construct psychology in sound design using a repertory grid," in *Proceed*ings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound, AM '10, (New York, NY, USA), pp. 8:1–8:6, ACM, 2010.
- [13] R. Picking, Supporting Students' Construction of Hypermedia. PhD thesis, Loughborough University, February 1996.
- [14] J. Berg and F. Rumsey, "Identification of quality attributes of spatial audio by repertory grid technique," *J. Audio Eng. Soc*, vol. 54, no. 5, pp. 365–379, 2006.
- [15] F. B. Tan and M. G. Hunter, "The repertory grid technique: A method for the study of cognition in information systems," *MIS Quarterly*, vol. 26, pp. 39–57, March 2002.
- [16] S. Cunningham, V. Grout, and R. Picking, Game Sound Technology and Player Interaction: Concepts and Developments, ch. Emotion, Content & Context in Sound and Music, pp. 235–263. IGI Global Publishing, 2010.
- [17] P. Schaeffer, *Traité des objets musicaux*. Editions du Seuil, Paris, 1966.
- [18] D. Eads, "hcluster: Hierarchical clustering for scipy." http://scipy-cluster.googlecode.com/, 2008.
- [19] K. Krippendorff, Content analysis: an introduction to its methodology. Sage, 2004.