ASSESSMENT OF PERCEPTUAL MUSIC SIMILARITY

Alberto Novello
Philips Research Laboratories, Eindhoven
The Netherlands
alberto.novello@philips.com

Martin McKinney
Philips Research Laboratories, Eindhoven
The Netherlands
martin.mckinney@philips.com

ABSTRACT

This paper extends a study on music similarity perception presented at ISMIR last year, in which subjects ranked the similarity of excerpt-pairs presented in triads [1]. The larger number of subjects and stimuli in the current study required a modification of the methodological strategy. We use here two nested incomplete block designs in order to cover the full set of song-excerpts comparisons (triads) while limiting the experimental time per subject. In addition to the two variable factors of the previous experiment, tempo and genre, we examine here the effect of prevalent instrument timbre. We found that 69 of 78 subjects where significantly consistent in their judgments of repeated triads. Furthermore, we found significant across-subject consistency on all 10 repeated triads. A significant difference was found in the distributions of inter- and intra-genre excerpt distances. The stress values in the Shepard's plot shows evidence of increased complexity in the present study compared to the previous smaller study.

1 INTRODUCTION

Recently, there has been an increasing interest in music similarity, both in the applicative [2] and research [5][6] fields. Various theoretical [3][4] and experimental works [5][6] have concentrated on which dimensions underlie listeners' perception of similarity. These studies were run on a small number of stimuli or on a limited number of genres, making it difficult to extend conclusions to the large corpus of Western music.

One of the most challenging problems in conducting an experiment on music similarity perception is dealing with the trade off between experimental time and the number of stimuli required for a complete representation of the complexity of the musical world.

Our recent study showed that a method combining triadic comparisons and Balanced Incomplete Block Design (BIBD) limited the reasonable experiment duration per subject to a reasonable length ($<$ 1 hour) while examining 18 excerpts. Here, we show how it is possible to further optimize the experimental design using two nested BIBDs to increase the number of stimuli, and thus to examine a broader range of musical styles.

2 METHOD

We employed a method using triadic comparisons of song-excerpts, because it is a straightforward procedure for subjects and it alleviates problems associated with scale interpretation. We used two nested BIBD to achieve triad reduction: one to create an incomplete but overlapping set of genres for each subject, the second to create a set of triads within each genre-set.

The first BIBD was calculated to determine the musical genres for each subject: we used quadratic comparisons (4 genres) per subject. The BIBD formula shows in this case, the number of genre-sets, b:

$$b = \frac{\lambda n(n-1)}{k(k-1)}. \quad (1)$$

With n=13 genres, k=4 (quadratic) and λ=3 (each genre pair appeared in three subject designs), we obtain b=39 genre-set, one for each subject.

For each genre-set, a BIBD on excerpts was generated. With 6 excerpts per genre, the number of excerpts per genre-set is 24. We used k=3 (triadic) and λ=2 (each excerpt-pair appears twice) reaching 184 triads per genre-set (subject). We added ten repeated triads for each subject for evaluation of within and across subject consistency.

2.1 The Web Experiment

We performed a web experiment which involved 78 subjects, running the total genre BIBD design twice. We had 59 males and 19 females and 50 musicians and 28 non-musicians. The average subject age was 28 years. After listening to a triad of excerpts, the subject had to choose the most similar and dissimilar pair among the three possibilities. The stimuli were 15-second excerpts of Western popular music covering a range of 13 musical styles. Genre classification was based on the "allmusic" website [7]. The song-excerpts belonged to one of two tempo categories: *fast*, for excerpts whose tempo was faster than 120 BPM for a quarter note; and *slow*, for excerpts whose tempo was slower than 100 BPM for a quarter note. The excerpts were also selected to fall into one of three timbre categories depending on their dominant musical instrument. The timbre categories were vocal, guitar and piano allowing us to find excerpts containing these instruments in all selected genres.

2.2 Analysis

The data analysis consisted of three main stages: within-subject consistency, across-subject consistency, and Multidimensional Scaling (MDS). We used the Kendal Coefficient of Concordance (KCC) [8] to evaluate consistency for both within and across subjects.

To model the multidimensional perceptual space, we first built a dissimilarity matrix of all subjects' rankings, assigning the value '2' for the least similar pair, '1' to the middle pair and '0' for the most similar. The ALSCAL multidimensional scaling algorithm [9] was used to estimate the coordinates of the excerpt positions best fitting the original data using a range of dimensionality orders. The Shepard's plot shows stress as a function of the number of dimensions. The optimal number of dimensions necessary to achieve an acceptable fit is typically the smallest number given the stress value is less than 0.2.

3 RESULTS

We calculated the within-subject consistency using the ten repeated triads of each subject. 69 subjects showed significant consistency at the 0.5 significance level in their repeated rankings, among these 4 subjects where very close to significance and only one subject performed very low in consistency. We also calculated the across-subject consistency on the repeated triads rankings. Figure 1 shows that the across subjects consistency is significant on all 10 testing triads, which is in line with our previous result of significant across-subject consistency in 97 over 102 triads.

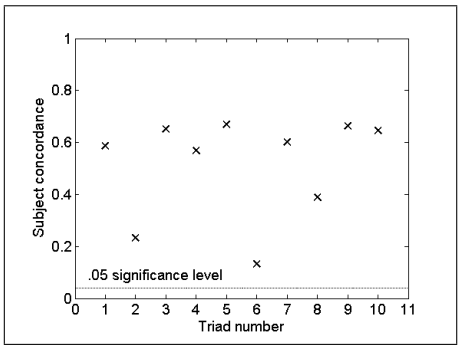

Figure 1. Across-subject consistency per triad

We calculated the Shepard's plot, shown in Figure 2 from our MDS analysis. The best compromise here, between dimension-order and stress value is six dimensions, giving a stress value of 0.175, while the previous study showed three dimensions to be the optimal choice.

From the coordinates of the excerpt positions in the six-dimensional space, we calculated inter- and intra-genre distances. The two distributions show significant difference, in agreement with our past results, confirming the importance of genre a factor in subjects' ranking.

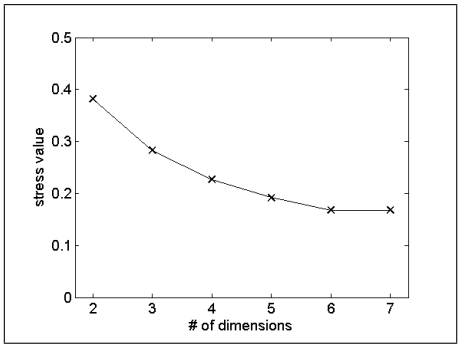

Figure 2. Stress value vs number of dimensions

4 DISCUSSION

The results confirm the findings of the previous study [1]: there is significant across-subject consistency on excerpt similarity ranking; most subjects also show significant consistency in repeated triads. However the stress values in the Shepard's plot shows evidence of increased complexity in the present study, most likely due to the lager set of stimuli. Thus the perception of music similarity in the current study is not easily mapped to three dimensions as in the previous study.

5 REFERENCES

[1] Novello A., McKinney .M, Kohlrausch .A, "Perceptual Representation of Music Similarity", 246-249, ISMIR 2006 Proceedings.

[2] Pampalk E., "Computational Models of Music Similarity and their Application in Music Information Retrieval", Ph.D. dissertation 2006.

[3] Deliege I.,"Introduction, Similarity Perception - Categorization - Cue Abstraction", *Music Percept.*, 18(3), 233-43, (2001).

[4] Ockelford A., "On similarity, derivation and the cognition of musical structure", *Psychology of Music*, 32(1), 23-74, (2004).

[5] Chupchik G. C., "Similarity and preference judgements of musical stimuli", *Scand. J. Psychol.* , 23, 273-282, (1982).

[6] Lamont A., Dibben, N., "Motivic Structure and the Perception of Similarity", *Music Percept.*, 18(3), 245-74, (2001).

[7] http://www.allmusic.org/.

[8] Kendall M., "Rank Correlation Methods", Charles Griffin, London (1975).

[9] Young F. W., Lewyckyj R., "ALSCAL User's Guide (5^{th} Ed.)", L.L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, NC (1996).