

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 excerpts 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 comparisons (triads) and limit 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. The results partially confirm the findings of the previous research, although there is greater variability in the preliminary data from this study: subjects show decreased consistency and a higher dimensionality is required by the multidimensional data fitting.

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 typically run on a small number of stimuli or on a limited number of genres, making it difficult to extend the 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 chose triadic comparisons of songs, 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 $\lambda=3$ (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 songs was generated. With 6 songs per genre, the number of stimuli per genre-set is 24. We used $k=3$ (triadic) and $\lambda=2$ (each song-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. Each subject had to first fill in a questionnaire with her/his general information and was then presented with the experiment which consisted of song triads. After listening to the 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. The songs belonged to one of two tempo categories (fast and slow) and to one of three timbres categories depending on dominant musical instrument: piano, vocal, and guitar.

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) [7] 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 [8] was used to estimate the coordinates of the excerpt positions best fitting the original data using a range of dimensionalities. 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 when the stress value is less than 0.2.

3 RESULTS

For each of the 19 subjects who already finished the experiment, we calculated the within-subject consistency from the ten repeated triads. 10 of 19 subjects showed significant consistency in their repeated rankings. We also calculated the across-subject consistency on the repeated triads rankings. Figure 1 shows that the across subjects consistency is significant on 9 over 10 triads, which is in line with our previous result of significant across-subject consistency in 97 over 102 triads.

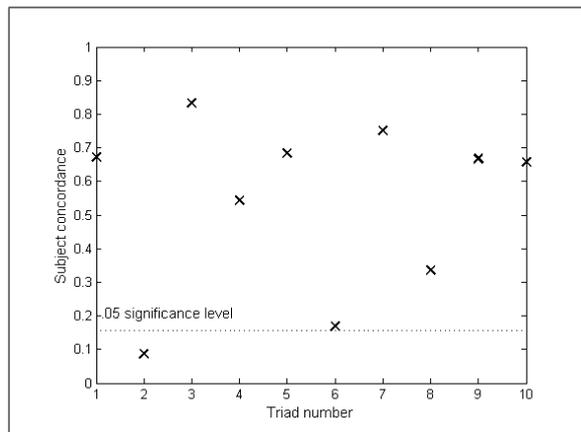


Figure 1. Across-subject consistency per triad

We calculated the Shepard's plot, shown in Figure 2 from our MDS analysis. The best compromise here, between dimensions 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 song 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.

4 DISCUSSION

Although the experiment is not yet completed, the results confirm some of the findings of the previous study [1]: in particular, the significant across subject consistency on 9/10 triads and the difference between inter- and intra-genre excerpt distances. However, the within-subject con-

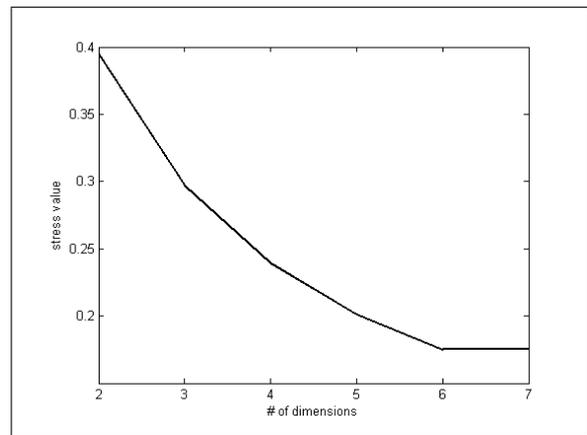


Figure 2. Stress value vs number of dimensions

sistency and the stress values in the Shepard's plot show evidence of increased complexity in the present study. The complete data gathered from the experiment will be presented in the poster paper.

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] Kendall M., "Rank Correlation Methods", Charles Griffin, London (1975).
- [8] Young F. W., Lewyckj R., "ALSCAL User's Guide (5th Ed.)", L.L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, NC (1996).