TOWARDS THE COMPARABILITY AND GENERALITY OF TIMBRE SPACE STUDIES

Saleh Siddiq[1], Christoph Reuter[1], Isabella Czedik-Eysenberg[1], Denis Knauf[2]

[1]Musicological Department, University of Vienna, Austria
[2]Vienna University of Technology, Austria
saleh.siddiq@univie.ac.at

ABSTRACT

This article describes an empirical meta study that was carried out to assess the comparability of timbre spaces [1]. A recent comparison of three popular timbre spaces revealed a lack of consistency among those studies [2]. It is most likely caused by the stimuli-sets that were vastly different from study to study. Thus far, instruments were reduced to a single tone, compared at the same pitch, and only (re-)synthesized sounds were used.

These findings raise the question whether an empirical meta timbre space would rather comply with the results of the original timbre spaces or confirm the inconsistency. Based on the original stimuli of the compared timbre spaces [3][4] [5], and additional natural instrument sounds out of the *Vienna Symphonic Library*, a hearing experiment was performed. By the means of multidimensional scaling, the obtained dissimilarity matrix was graphed into a new meta timbre space and eventually structured through a hierarchical clustering.

The inconsistency is confirmed. The meta timbre space yields a clear clustering of stimuli-sets. Apparently, there is a greater timbral resemblance among the different instrument sounds from the same stimuli-set than among the sounds of the same instrument across the different stimuli-sets. Hence, the timbral differences between the stimuli-sets prevail as primary discrimination cue and thus, impair the comparability and generality of timbre space studies.

1. INTRODUCTION

Although most people intuitively “know” what timbre, or tone quality, is, it in fact remains a very elusive phenomenon when it comes down to the hard facts. Timbre can be associated with many facets of music like instrumentation, pitch range, articulation, and musical dynamics. Hence, it is impossible to describe timbre with a distinctive sound feature—as compared to other characteristics such as pitch (~periodicity) or loudness (~intensity). Timbre research can be basically divided into two main branches: the investigation of sound production (i.e. musical acoustics) and the investigation of sound perception (i.e. music psychology) [6]. A popular focal point of musical acoustics is the acoustics (which in this case means timbre) of musical instruments in terms of identification and discrimination. So far, several acoustic parameters have been identified as contributors to musical instrument timbre (see [6] for a brief summary), and it’s a complex interaction of these features that makes up the instrument sound. This knowledge about timbre from an acoustical perspective is reflected in a definition provided by Stumpf as early as 1890 [7]. Stumpf’s definition was, in fact, an adaptation of a definition previously published by Helmholtz [7][8][9]. Helmholtz then considered the harmonic spectrum as the only physical correlate of timbre. Stumpf accepted it as the main feature, labeled it “Klangfarbe im engeren Sinn” (roughly translated: timbre in a narrow sense), and further packed all the other (temporal) features, such as noise, transients, fluctuations, musical phrases etc. together and labeled them “Klangfarbe im weiteren Sinn” (timbre in a wider sense).

While the acoustic components of timbre are thus well explored, there’s still not much known about their psychological correlates that are actually used by the ear to perceive an impression of timbre. Since the publications of Helmholtz’ “On the Sensations of Tone” (1863) [8]—especially the English translation by (1875) Ellis [9]—and Stumpf’s “Tonpsychologie II” (1890, unfortunately never translated), timbre has gained attention by empirical scientists. Since then, several approaches have emerged in order to describe the perception of timbre. Thereof, especially the so-called *timbre spaces* (TS) have generally been accepted. TS are (most often Euclidean) virtual spaces that translate timbral dissimilarities into spatial distances. That means, the closer two sounds are located in such a space, the stronger their timbres resemble each other. Although most of those studies were somehow productive, there are some common noticeable drawbacks: (1) musical instrument sounds were generally (re-)synthesized instead of being actually recorded. (2) Instruments were reduced to a single tone, (3) they were, in each case, compared on the same pitch that (4) inevitably had to be out of range for many instruments (imagine trying to find a common pitch for double bass and flute or even piccolo). Musical instruments obviously can’t be properly represented through a single tone. Such methodical weaknesses considerably reduce the data basis, thus minimizing the chance of significant data overlap between two studies, and, as a consequence, have a negative impact on the comparability and therefore the validity of the studies.

1.1 Comparison of Timbre Spaces

If it is assumed that TS studies yield significant results about timbre similarities of musical instruments, then the same instruments should appear in roughly the same spatial regions across the different studies and thus, the spatial relations between all the instruments should be consistent across the studies as well.

In an earlier meta study, the results of three of the most popular TS studies by Grey 1975 [3], Krumhansl 1989 [4], and McAdams et al. 1995 [5] were compared [2]. These TS were chosen not only because of their popularity and significance but also because they contain a decent number of the same instruments (comparing flute A with flute B obviously makes more sense than comparing flute A with double bass B) and the coordinates of all three Euclidean dimensions were, somehow, accessible.

The 3D-coordinates of all instruments were extracted from every TS, uniformly scaled (the lowest value becoming 0, the highest 100) and aligned, and finally graphed into a new 3D scatter plot. The result was a meta TS (MTS, see Figure 1) that revealed a notable inconsistency among the compared TS. That means that the same instruments, across all studies, were located in widely different regions of their respective spaces.

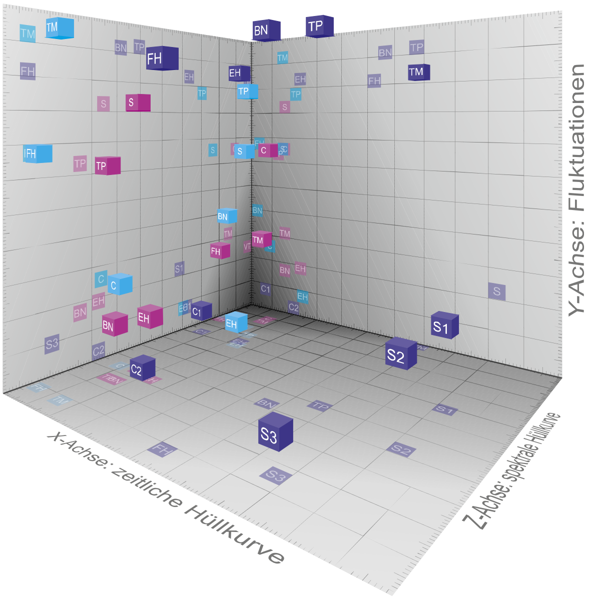

Figure 1: Meta timbre space (MTS). Colors: blue = Grey (GRY), cyan = Krumhansl (KRH), magenta = McAdams (MCA); dimensions: x = temporal envelope, y = fluctuations, z = spectral envelope; abbrev.: BN = bassoon, C = clarinet, EH = cor anglais, FH = french horn, S = strings, TM = trombone, TP = trumpet.

2. QUESTION

These results, and the fact that same instruments are obviously often represented by very different stimuli, raise a new question: Will an empirical TS, based on the original stimuli from the compared studies, support the inconsistency or instead reconcile with the original TS? Or to put it another way: How much of an influence do the utilized stimuli have on the comparability and generality of TS studies?

3. METHODS

To investigate the question, the same TS as in the earlier comparison [3][4][5] were now compared in an empirical meta study [1]. The methods basically match those of the compared studies. By means of a hearing experiment (pairwise comparison) and a multidimensional scaling (MDS), a meta TS was ascertained [3][4][5][10]. For the first time, the new empirical meta TS (EMTS), allows to compare the stimuli of different TS as well as actually recorded instrument sounds of the *Vienna Symphonic Library* (VSL) in the same context. On top of that, a hierarchical clustering was performed in order to closely examine the (spatial) arrangement and relations of the sounds.

3.1 Stimuli

This study includes every instrument that is represented in each of the compared TS. Hence, the following seven instruments were tested: bassoon, clarinet, English horn, French horn, strings (i.e. celli), trombone, and trumpet. The utilized 24 stimuli were exactly the same stimuli used in the original studies by Grey (1975) [3] (including three celli, two clarinets; thus a total of ten stimuli) and Krumhansl (1989) [4] (seven stimuli; McAdams et al. (1995) [5] utilized the same set of stimuli) and moreover actually recorded instrument sounds (seven stimuli) from the *Vienna Symphonic Library* (VSL). According to the stimuli from the original studies, the pitch was Eb4 (roughly 313 Hz).

3.2 Subjects

A total of 35 subjects, including 15 females and 20 males, participated in the experiment. Their ages ranged from 19 to 72 (Ø=30.9, SD=13.3). The subjects had to assess their amount of musical experience by completing a short questionnaire before the experiment. 24 subjects were musicians (including playing instruments, singing, and conducting), eight were formerly active and three were non-musicians (Ø=19.6 years of experience, SD=14.2).

3.3 Procedure

The experimental session lasted roughly 45 to 60 minutes and consisted of four phases: an instruction and familiarization phase, a training phase, the actual experimental phase, and the questionnaire. The instructions were presented orally as well as in written form on the screen. Subjects were allowed to ask any question to avoid possible misconceptions. After that, all 24 stimuli were presented in a randomly ordered sequence in order to familiarize the subjects with the range of timbre variation among the test sounds. Unlimited repetitions were allowed but the order of presentation was random every time.

The timbral dissimilarity of the stimuli was rated subjectively in a pairwise comparison. Subjects did their ratings on a one-dimensional ten-point-scale (0 = most similar, i.e. identical; 9 = least similar). The succession of both the sounds within a pair and the pairs themselves was fully randomized for each subject. The randomized order of the sounds within a pair enabled us to cut the session duration in half by dropping all complementary ordered pairs (i.e. B–A instead of A–B) while still controlling the confounding influence of the order of presentation. Identical pairs (A–A) were excluded as well. Hence, the experimental phase consisted of ((24·23):2=) 276 pairs. Each trial could be repeated as often as needed. Subjects were allowed to take unlimited breaks at any time during the session, provided they were not in the middle of a trial.

The preceding training phase, consisting of 20 trials (randomly drawn out of the available 276 pairings), was identical to the actual test, so subjects could get accustomed to the procedure and hopefully develop some kind of a consistent rating strategy.

The experiment was performed on a specially developed browser-based software. The stimuli were presented through external sound cards (Roland Quad-Capture UA55) and electrostatic headphones (Koss ESP 950 with amplifier E 90).

3.4 Evaluation

The perceptual ratings of each subject were stored in a separate symmetric dissimilarity matrix. In fact, the matrices were half-matrices (i.e. only the upper triangle) because leaving out the complementary pairs led to an empty lower half-matrix. Since systematic asymmetries have never been found [10], leaving out the lower half, the complementary pairs respectively, is not a constraint. The individual half-matrices were averaged into an overall half-matrix. Using the median instead of mean value considerably lowered the stress value of the subsequently calculated spatial configuration. The calculation was carried out by means of a non-metric multidimensional scaling (MDS). Based on the eigenvalues, a four-dimensional configuration proofed to be an appropriate fit (Kruskal's stress = 0,0362). Thereon, eventually, a hierarchical clustering was performed in order to in depth study the spatial arrangement of the sounds.

4. RESULTS

The ascertained new empirical meta TS (EMTS, see Figure 2) strongly confirms the previously found inconsistency. It is visible to the naked eye, that, in the EMTS, sounds of the same instruments do not reside in the same spatial regions. So the sounds of the same instruments (e.g. flute A and flute B) do not possess significant timbral similarities, which hence causes the expected instrument-clusters not to evolve. That would be enough to state that there's a lack of comparability among and thus a lack of generality to (the compared) TS-studies.

Furthermore, another striking phenomenon was found: while the VSL-sounds tend spread over the whole space, the Grey-stimuli (GRY) and Krumhansl-stimuli (KRH) primarily group on opposing sides of the space. In other words: the sounds group into some kind of "stimuli-set-clusters" instead of the aforementioned expected instrument-clusters. The hierarchical clustering (see Figure 3) reveals a strict division into three main clusters. Interestingly, (1) the clarinets (VSL and KRH) set themselves apart as a separate cluster. The rest splits into (2) an almost exclusive GRY-cluster (only disturbed by the VSL bassoon) and (3) a KRH-cluster. The latter one is spread wider since it also contains most of the VSL-sounds.

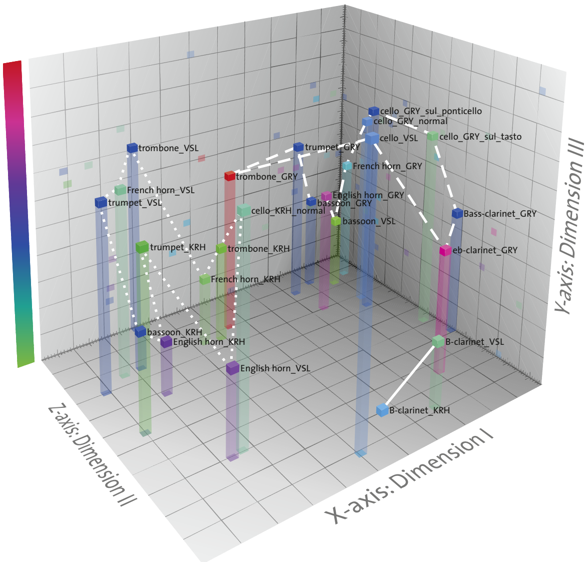

Figure 2: Empirical meta timbre space (EMTS). Color scale = dimension IV, red = high value, green = low value. No physical correlates of timbre were attached to any of the dimensions. The three main clusters are highlighted by the white lines: solid = b-clarinets (VSL & KRH), dotted = KRH-cluster, dashed = GRY-cluster.

5. DISCUSSION AND OUTLOOK

The results of this empirical meta study do not comply with the data of the compared studies, but rather confirm the inconsistency among those studies that was found in an earlier comparison. Possible reasons for this inconsistency could be, (1), subjective data diverging from study to study —although these are unlikely to fully account for the vast differences—and (2) more probably, different distance models and MDS algorithms that were used to process the data (see [10] for an in-depth discussion). If we furthermore take the conspicuous grouping in stimuli-set-clusters into account, it seems likely that utilizing natural sounds might have the biggest impact on the validity of TS studies. In fact, the obvious systematic timbral toning (or—in this case—rather "ton*e*ing") of the stimuli-sets seems to stand out so much that it simply overrules the initially sought-after instrumental characteristics and become the primary cues of timbre discrimination. The timbral toning most probably stems from the process of (re-)synthesizing the stimuli, which further supports the notion that natural sounds from actually recorded instruments are better suited for studying musical timbre. The notion that the sound system might be the crucial factor can be dismissed, since the systematic timbral differences among the stimuli-sets were apparent even though the stimuli were presented through the same speakers (in this case headphones). Nonetheless, it is clear that the frequency response of loudspeakers always, more or less, influences timbre perception. Thus, diverging results from different studies may—not fully but to some extent–be also put down to different sound systems.

However, it needs to be stated that stimuli duration could have a biasing influence. The GRY-stimuli averaged 350 ms [3] while the KRH-stimuli averaged 673 ms [5]. The different durations, taken by themselves, are not a problem because (1.) they only contribute to the studied differences between the stimuli-sets and (2.) by a stimulus duration of max. 300 ms, the ear has reached its maximal performance. That means, even by the end of the shorter GRY-stimuli the impression of timbre through the ear is well established and any further increase of duration won't increase the quality of perception by the ear [11][12][13] [14]. But the problem became evident when the duration of the VSL-sounds had to be adjusted. Since manipulating the stimuli of the compared studies was not an option, the only question was then towards which stimuli-set the VSL-sounds would be tilted. The perceptual durations of the VSL-sounds were eventually matched with those of the KRH-stimuli. At first glance, this complies with the basal clustering. A further examination, however, shows that same instruments of both sets still do not possess considerable timbral similarities. If we examine the leafs of the tree model (i.e. the single items/instruments), the same instruments of different sets never share the same nodes (see Figure 3). So the durations apparently do not interfere with timbre perception but might be a factor on a more general level of over-all similarity. While different durations become confounding factors in meta studies that are bound to have some kind of impact, it might be fair to assume that they do not generally impair the validity of TS studies per se.

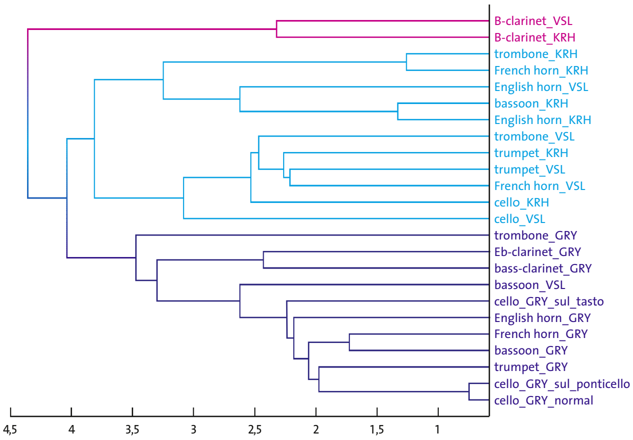

Figure 3: Dendrogram of the hierarchical clustering. The main clusters are seperated by color: magenta = clarinet-cluster, blue = GRY-cluster, cyan = KRH-cluster.

Other than that, the results indicate that natural sounds are better suited to yield reliable data with regards to timbral similarities of musical instruments. In other words: the usage of actually recorded sounds presumably would significantly enhance the external validity, reliability and thus generality of TS studies.

The next steps planned include further empirical studies, exclusively using real instrument sounds and taking musical dynamics and pitch as influencing variable of timbre. Therefor, musical instruments will be tested over a vast range of their respective ranges of pitch and dynamics. This considerable broadening of the data basis for each instrument will certainly lead to results that are (1) reproducible and hence reliable, (2) closely related to the actual circumstances in music, and thus will (3) yield more realistic and universal information about the perceptual similarities of the timbre of musical instruments.

6. ACKNOWLEDGEMENTS

Supported by funds of the Österreichische Nationalbank (Österreichische Nationalbank, Anniversary Fund, project number: 16473). Special thanks to the *Vienna Symphonic Library* for supporting our work with two free “Vienna Super Package” licenses.

7. REFERENCES

[1] Siddiq, S. et al., “Kein Raum für Klangfarben – Timbre Spaces im Vergleich”, Fortschritte der Akustik – 40. DAGA 2014, Oldenburg, Germany, 2014

[2] Siddiq, S., “Timbre Space revisited – Was ist Fakt, was ist Mythos?” Proc. FAMA (of the DEGA), Detmold, Germany, pp. 51–52, 2014

[3] Grey, J. M., “An exploration of musical timbre using computer-based techniques for analysis, synthesis and perceptual scaling,” Stanford University, Report No. STAN-M-2, Stanford, 1975

[4] Krumhansl, C., “Why is musical timbre so hard to understand?” Nielzen, S. and Olsson, O. (eds.), “Structure and perception of electroacoustic sound and music,” Amsterdam, Netherlands, pp. 43–53, 1989

[5] McAdams, S. et al., “Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes,” Psychological Research, 58(3): 177–192, 1995

[6] Auhagen, W., “Zu Fragen der Klangfarbenwahrnehmung und der Klanggestaltung durch Musiker,” Eberl, K. and Ruf, W. (eds.), “Musikkonzepte – Konzepte der Musikwissenschaft. Bericht über den Internationalen Kongress der Gesellschaft für Musikforschung, Halle (Saale) 1998,” Vol. 1, Kassel, Germany, pp. 86–100, 2000

[7] Stumpf, C., “Tonpsychologie,” Vol. 2, Stuttgart, Germany, 1890

[8] Helmholtz, H. v., “Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik,” Braunschweig, Germany, 1st ed. 1863

[9] Helmholtz, H. v. and Ellis, A. J., “On the sensations of tone as a psychological basis for the theory of music,” London, United Kingdom, 1st english edition, 1875

[10] McAdams, S., “Perspectives on the Contribution of Timbre to Musical Structure,” Computer Music Journal, 83(3): 85–102, 1999

[11] Mertens, P.-H., “Die Schumannschen Klangfarbengesetze und ihre Bedeutung für die Übertragung von Sprache und Musik,” Frankfurt/M, Germany, 1975

[12] Reuter, C., “Die auditive Diskrimination von Orchesterinstrumenten. Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel,” Frankfurt/M, Germany, 1996

[13] Gefand, S. A., “Hearing. An introduction to psychological and physiological acoustics,” Yew York, NY, 2004

[14] Yost, W. A., “Fundamentals of hearing. An introduction,” San Diego, CA, 2007