

# The Tonnetz at First Sight: Cognitive Issues of Human–Computer Interaction with Pitch Spaces

José L. Besada<sup>1,2</sup> , Erica Bisesi<sup>3</sup>, Corentin Guichaoua<sup>4,5</sup>   
and Moreno Andreatta<sup>4,5</sup>

## Abstract

Pitch spaces allow pitch relations to be expressed through geometrical representations for many different purposes. The *Tonnetz* is a well-known pitch space in the field of music theory; equivalent representations have been described in the field of cognitive science, especially Krumhansl's model of perceived triadic distance. Despite her empirical approach, we know very little about the way people interact, cognitively speaking, with *Tonnetz*-based computational platforms involving multimodal stimuli. Our study has approached this issue by means of empirical experimentation for the first time. A total of 88 participants, with varying backgrounds in music and mathematics, were asked to interact with a *Tonnetz* interface; they did not have prior knowledge of this pitch space. Results of our experiment confirmed our main hypotheses. On the one hand, strong skills in music theory are needed to partially grasp the overall structure of the *Tonnetz* at first sight; this aspect is mainly related to the quality recognition of triads and the detection of shared pitch classes in harmonic motions. On the other hand, the particular geometry of the *Tonnetz* may bias this understanding when non-functional harmonic sequences are displayed on it.

## Keywords

Harmony, human–computer interaction, multimodal interaction, pitch spaces, *Tonnetz*

Submission date: 8 October 2023; Acceptance date: 26 March 2024

The gradual development of notational elements, especially concerning pitch, has marked—although not exclusively—the history of Western music (Grier, 2021). At first, neumatic notation facilitated oral transmission of religious monodies; subsequent Western notations crystalized a sophisticated symbolic abstraction that today facilitates the encoding of a broad number of musical repertoires on staves. Beyond the standard staff notation, a number of people involved in musical practices have developed many spatial configurations to annotate pitch-related information with diverse purposes such as composition, music theory and analysis, the psychology of music, or music information retrieval—to name a few. Depending on this variety of purposes, different spatial representations may overlap or strongly differ.

This article aims to inspect a spatial representation of pitch coming from the field of music theory through the lense of cognitive science. Although the psychology of

music and systematic musicology have explicitly approached neo-Riemannian theories a few times (e.g.,

<sup>1</sup> Departamento de Musicología, Universidad Complutense de Madrid, Madrid, Spain

<sup>2</sup> Instituto Complutense de Ciencias Musicales, Madrid, Spain

<sup>3</sup> Faculté de Musique, Université de Montréal, Montréal, Canada

<sup>4</sup> IRMA/CREAA, Université de Strasbourg and CNRS, Strasbourg, France

<sup>5</sup> STMS Lab (UMR9912), CNRS, Sorbonne Université and IRCAM, Paris, France

## Corresponding author:

José L. Besada, Department of Musicology (Faculty of Geography and History) of the Universidad Complutense de Madrid and Instituto Complutense de Ciencias Musicales. Calle del Profesor Aranguren s/n, 28040 Madrid, Spain.

Email: jlbésada@ucm.es

Data Availability Statement included at the end of the article



Brower, 2008; Krumhansl, 1998; Moss, 2014), this study stands—to our knowledge—as the first psychological investigation on the *Tonnetz* with quantitative methods from the perspective of human–computer interaction. Prior to the presentation of the study, we contextualize our approach within the underlying theoretical scenario and the existing literature.

## Approaching the *Tonnetz* Empirically

### Tonal Pitch Spaces

Fred Lerdahl is a scholar at the crossroads of several fields mentioned above: composition, music theory, and the psychology of music. In the late 1980s, he implicitly defined pitch spaces as “topological models” expressing pitch relations; particularly, tonal pitch spaces are, from his perspective, “intended to capture the sense of proximity and distance among pitch configurations that listeners bring to bear when hearing tonal pieces” (Lerdahl, 1988, p. 315).

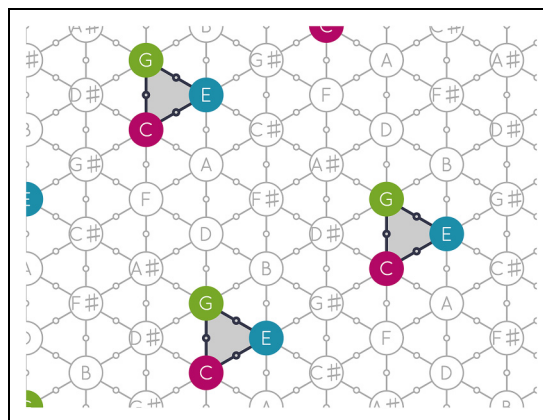
The first pitch space reproduced by Lerdahl was the double helical model developed by Roger Shepard (1982a): it vertically unfolds the dimension of pitch height while orthogonally projecting a circle corresponding to pitch chroma. For Western musicians, the notion of pitch height is trivial as the staff notation strongly manifests this dimension.<sup>1</sup> Traces of the conceptual metaphor *PITCH RELATIONSHIPS ARE RELATIONSHIPS IN VERTICAL SPACE* are pervasive in Western musical culture (Zbikowski, 2002). Although this fact seems to be rooted in a cross-modal correspondence between pitch height and visual elevation (Parise et al., 2014; Parkinson et al., 2012; Walker et al., 2010), recent research points to broader conceptual primitives for pitch beyond spatial verticality (Antović et al., 2020). Curved pitch spaces are not as pervasive as straight ones, but music theorists often deal with circular representations of pitch denoting octave equivalence (Tymoczko, 2011). Pitch spaces combining linear and circular features are also found—more rarely, though—beyond the literature of the psychology of music—we can mention, for instance, Archimedean pitch spirals by Iannis Xenakis in his early compositional sketches (Besada, 2022).

Lerdahl (1988) raised a criticism against the helical model due to its homogeneity, as it gathers “two strands of whole-tone scales” which do not account for the “less symmetrical” distribution of the diatonic system (p. 318).<sup>2</sup> His criticism also applied to another highly “symmetric” tonal pitch space, proposed by Christopher Longuet-Higgins (1962) and Gerald Balzano (1982), wherein pitch height unfolds in two independent directions: horizontal—perfect fifths—and vertical—major thirds. Lerdahl equally acknowledged that, long before these pitch spaces were introduced in music psychology, analogous rectangular lattices and schemes were adopted by music theorists to depict proximities of tonal keys (Schoenberg 1954/1969; Weber 1824/1842).

## The *Tonnetz*: An Overview

In spite of Lerdahl’s accurate mention of Gottfried Weber’s and Arnold Schoenberg’s representations, his overview does not mention similar earlier schemes, such as the one dating from the 18th century by mathematician Leonhard Euler (1739).<sup>3</sup> On this path, further pitch spaces by German theorists in the 19th century, such as those presented by Arthur von Oettingen (1866) and by Hugo Riemann (1873/1992)—for a historical review of these schemes and similar ones, see Cohn (1998; 2011)<sup>4</sup>—paved the way at the end of the 20th century for the modern 12-TET *Tonnetz*. This pitch space is the most widespread schematic representation among the neo-Riemannian theories—or Riemann systems (Lewin, 1982)—a particular branch of transformational music theory.

The modern *Tonnetz* is a geometrical lattice made of equilateral triangles (Figure 1),<sup>5</sup> where the vertices—here marked with circles—stand for pitch classes instead of pitches, that is, regardless of octave registers. The triangular tiling defines three axes: The vertical axis unfolds perfect fifths—for instance C-G-D-A and so on—while the two diagonal ones respectively unfold major thirds—C-E-G $\sharp$ -C and so on—and minor thirds—C-D $\sharp$ -F $\sharp$ -A and so on. Notice the equivalence of the axes related to perfect fifths and major thirds with the pitch space by Longuet-Higgins (1962); the third axis, in contrast, depends on the other two. Additionally, relevant properties of this pitch space—like the morphology of its elementary triangles and their relationships (explained below)—are invariant under axial rotation and symmetries. This axial configuration induces the interval content of the *Tonnetz* elementary triangles: There are only two layouts—both invariant under translation—respectively corresponding to major and minor triads. Due to the particular cyclicity of the perfect-fifth axis, which generates the aggregate, the *Tonnetz* gathers all the 24 possible major and minor triads within the 12-TET system. Such a pitch space periodically replicates itself; thus, its planar structure can be wrapped—which gives rise to a torus.



**Figure 1.** Partial screenshot of the *Tonnetz* webpage in which the C major chord is highlighted on the *Tonnetz*.

Any triangle of the lattice is surrounded by three others with opposite layout, each sharing two vertices with the former. Musically speaking, this means that any major (minor) triad is related to three minor (major) triads by sharing two pitch classes with each of them. These relationships depend on three possible transformations that non-trivially maximize pitch-class intersections (Figure 2): the Parallel operation—henceforth P—preserves the root of both triads and the shared perfect fifth while swapping the order of the harmonic minor and major thirds; the Relative operation—henceforth R—relates a pair of triads with a common major third while their roots are separated by a minor third; the Leading-Tone operation—*Leitonwechsel*, henceforth L—relates a pair of triads with a common minor third while their roots are separated by a major third. As the modern *Tonnetz* is structured in terms of pitch classes (Cohn, 1997), none of these operations entails any particular motion. However, as shown in Figure 2, pitches can be arranged to produce parsimonious voice leadings with these operations: One voice smoothly moves up or down—a chromatic semitone for P, a diatonic semitone for L, and a tone for R—while the other two voices remain sustained.

The PRL family of elementary operations outlined above generates the entire space of major and minor triads—although R and L alone suffice as generators because P can be derived from them. Any pair of perfect triads can be therefore connected through a limited chain of elementary operators—in fact not more than five. These neo-Riemannian tools have proven to be useful for analyzing many tonal passages that would be hard to explain by means of only functional tonal models, since tonal functionality mainly accounts for proximity through the cycle of fifths instead of parameters like pitch-class intersection or voice-leading smoothness. The model and further neo-Riemannian extensions fit particularly well with the requirements for the analysis of the chromatic music of the late-Romantic period (e.g., Childs, 1998; Cohn, 1996; 2012) and is also appropriate for approaching varied musical repertoires in the extended common practice, such as jazz and bossa nova (e.g., Briginshaw, 2012; Capuzzo, 2006; de Lemos Almada, 2020), pop and rock (e.g., Bigo & Andreatta, 2014; Capuzzo, 2004), or minimalist music (e.g., Cohn, 2019).

Further research in music psychology has also dealt with pitch spaces similar to those proposed by Longuet-Higgins

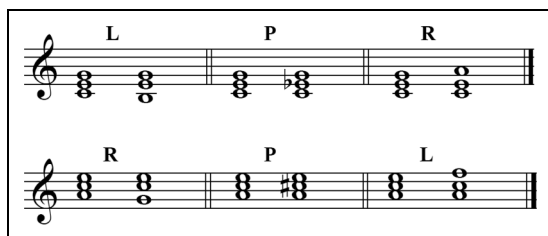
and Balzano—for an account, see Shepard (1982b). A significant contribution was provided by Krumhansl and Kessler (1982), who introduced a pitch space derived from an empirical study on perceived triadic proximity. Years later, Carol Krumhansl (1998) explicitly discussed the resemblances of her model with the dual representation of the *Tonnetz*—mathematically speaking, the so-called chicken-wire torus (Douthett & Steinbach, 1998)—by trying to find a compromise between tonal functionality and pitch-class intersections, which led later to some neuroscientific speculation about mental key maps (Zatorre & Krumhansl, 2002). More recent empirical research on the *Tonnetz* and further neo-Riemannian concepts are mainly based on computational inspection, encompassing perspectives from psychoacoustic models (e.g., Bernardes et al., 2016; Milne & Holland, 2016) to music information retrieval (e.g., Aminian et al., 2020; Chuan & Herremans, 2018; Lieck et al., 2020).

### Considering Transformational Music Theory Through Human–Computer Interaction

Although the previous literature review attests to the importance of pitch spaces in both music theory and cognition, some spatial implications of these geometric representations have been largely disregarded. They are, however, important from the point of view of embodied cognition and multimodal interaction. For instance, some authors have started exploring bodily and gestural responses to *Tonnetz*-based environments (Cannas, 2018; Hedges & McPherson, 2013; Holland et al., 2009; Mandanici et al., 2016), but without a systematic discussion of their cognitive implications. This issue is also relevant for the so-called isomorphic musical instruments, as several of them depend on *Tonnetz*-like real or virtual keyboards (Graf & Barthet, 2023; Maupin et al., 2011; Milne et al., 2007; Park & Gerhard, 2013).

From the perspective of embodied experience, a comprehensive discussion on cognitive paradoxes in the presentation of several pitch spaces has been proposed by Candace Brower (2008). In particular, she suggested a modification of the modern *Tonnetz*, retaking actual pitches—instead of pitch-classes—and slightly distorting the axes to make them, visually speaking, more consistent—and therefore more comprehensible—with an underlying vertical pitch schema. Although Brower's criticism is purely theoretical—as she does not rest on any empirical research—her formulations opened the door for the main hypotheses of our study:

- Despite the apparent geometrical simplicity of the *Tonnetz*, its deep structural implications are very complex. Only people with high skills in music theory are able to partially grasp its overall structure at first sight. This partial grasp should not be understood as a threshold of accuracy in the recognition of



**Figure 2.** Possible parsimonious transformations for major and minor triads.

the structure but as a direct or implicit understanding of fragmentary properties of the *Tonnetz*.

- The geometry of the *Tonnetz* is not consistent with the widely spread vertical schema for pitches as it unfolds differently in its three axial directions; nor does it provide self-evident patterned representations of tonal functionality—even the implicit circle of fifths, unfolding in one of its axes, may go unnoticed at first sight. Both facts may bias the comprehension of the *Tonnetz*, particularly when non-functional harmonic sequences are represented on it.

The formulation of these hypotheses also comes from our scholarly experience with human–computer interaction in the field of music theory. The SMIR project<sup>6</sup> led us to develop a robust audiovisual web platform for the *Tonnetz* (Guichaoua et al., 2021); our observation of varied people interacting with it raised some intuitions around the mentioned hypotheses that we are formalizing in this study, through a controlled multimodal experiment by means of an adaptation of our web platform. Among the plethora of digital musical-learning resources (Mandanici et al., 2023), a growing number of software applications has been devoted to multimodal interaction with the *Tonnetz* and *Tonnetz*-related pitch spaces (e.g., Bergstrom et al., 2007; Bigo et al., 2015; Holland, 1992). The pitch relations enabled by software approaches visually suggest the metaphor of motion that stands at the core of transformational music theory (Attas, 2009).

## Method

### Participants

A total of 88 participants (44.32% female, 55.68% male) took part in the study. In order to have 4 different groups of 22 people each, participants were recruited according to the following criteria: Music professionals—henceforth MusPro participants—held a conservatory diploma and developed their careers as singers, instrumental performers, composers, conductors, music teachers, and/or musical radio broadcasters (age:  $M = 38.32$ ,  $SD = 8.07$ ); science professionals—henceforth SciPro participants—held a master’s degree or a PhD in a scientific or technological field and developed their careers as mathematicians, physicists, computer scientists, engineers, or architects (age:  $M = 43.68$ ,  $SD = 12.10$ ); music students—henceforth MusStu participants—were bachelor students in musicology with formal music training not yet professionalized (age:  $M = 22.82$ ,  $SD = 2.89$ ); science students—henceforth SciStu participants—were bachelor’s or master’s students in mathematics, physics, or engineering (age:  $M = 20.31$ ,  $SD = 1.77$ ). For a clearer differentiation of the groups, MusPro and MusStu participants—henceforth, when pooled, Mus participants—were chosen from people with no mathematical education at university; conversely, SciPro and SciStu participants—henceforth, when pooled, Sci participants—were chosen

from people with 0–4 years’ musical training—whether formal or informal. Although all Mus participants reported significant knowledge in music theory—in some cases they were relatively proficient—none of them were familiar with neo-Riemannian theories prior to the participation in our study. Color blindness was also a criterion for exclusion.

The choice of the groups was guided by the hypotheses mentioned above. On the one hand, Sci participants are more than a mere control group in terms of musical training; these participants have received or were receiving scholarly education in which geometrical abstraction was a pivotal feature. On the other hand, the difference between groups MusPro and MusStu is aimed at measuring the impact of the proficiency in terms of music theory skills.

### Audiovisual Stimuli

The aural stimuli consisted of 12 harmonic sequences of 7 major and minor block triads. Each chord lasted 1 s, except for the last one of each sequence, which was sustained for 8 s. Sequences were presented with synthesized organ-like sounds—with a piano-like decay for the last sustained chord—and were dynamically balanced to avoid any chord or particular voice becoming more salient in terms of volume.

The 6 first chords of every sequence unfolded progressions by using only 2 of the 3 possible parsimonious relationships: 4 sequences were built on LR- or RL-progressions, 4 on LP- or PL-progressions, and 4 on RP- or PR-progressions. Besides, half of the sequences opened with a major triad and the other half with a minor triad. Each sequence started with a different pitch class as the root of its first chord, in order to minimize potential biases of persistent openings. The last chord of each sequence was chosen according to six possible options (see Table 1). Therefore, there were 2 representatives of each option in the total pool of 12 sequences: Typologies 1 and 2 keep parsimony, respectively prolonging or not the harmonic direction of the sequence; Typologies 3 and 4 are tonally stable as they close the sequence with a perfect-fourth or perfect-fifth motion in the bass layer; Typologies 5 and 6 produce the most unattended cadential endings, tonally speaking, of the harmonic sequences. The choice of six such options homogenizes the set of stimuli, so facilitating statistical data analysis, in particular the estimation of expected values—as shown in the Results section.

**Table 1.** Description of the typologies for the last chords of the sequences in function of their relationship with the preceding triad.

Last chord typology	1	2	3	4	5	6
Continuing progression	Yes	No	No	No	No	No
Shared quality	No	No	Yes	No	Yes	Yes
Shared pitch classes	2	2	1	1	0	0
Forté’s interval class between roots	0, 3, or 4	0, 3, or 4	5	5	1	6

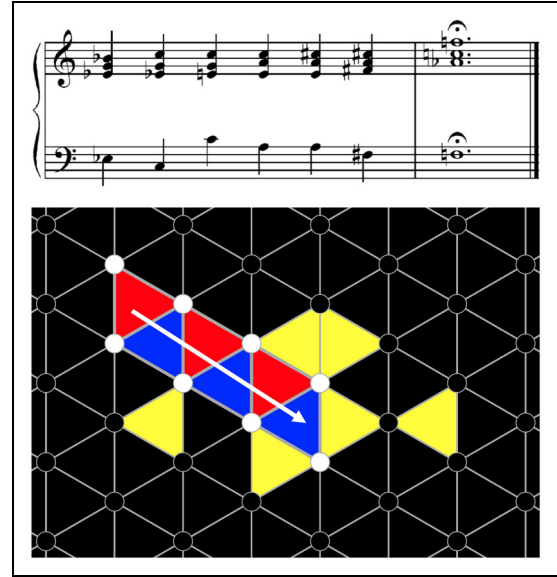
All the triads were displayed in 4 voices—the lowest one always presenting the root, whereas the upper ones always contained the 3 pitch classes of each triad in a closed position. The choice of root positions—fixed by the lowest voice—was aimed at avoiding chord inversions, which could be perceived as more unstable from a tonal viewpoint, due to less spectral harmonicity (McDermott et al., 2010). Layers moved within the average voice range in a SATB choir, ranging from  $E_3$  to  $G_6$ , to avoid register biases (Biasutti, 1997). The voice leading was smooth, that is, the bass layer never unfolded a melodic motion of intervals in the same direction that exceeded a perfect octave; upper voices always moved parsimoniously until the last chord, which was reached by the smallest interval steps and avoided some forbidden voice-leading during common-practice period—mainly parallel and direct octaves or fifths. A comprehensive list of pitches for the 12 sequences in MIDI values is provided in Appendix A.

Every sequence was aligned with its visual counterpart on the *Tonnetz*. The lattice was presented with gray lines on a black background, and the triangle vertices were emphasized with small circles. When the first chord in a sequence started sounding, a triangle of the lattice was lit up—in red if the triad was major; in blue if it was minor. At the same time, its surrounding small circles—standing for pitch classes—were lit up in white. This was repeated for every chord of each sequence, excluding the last one and keeping the previous lit items on screen. Because of the parsimonious harmonic organization of the sequences, the aural chord progression was synchronized with a visual unfolding of aligned triangles with alternating colors. The visual counterpart of the last chord was different: 6 yellow chords appeared at the same time, four of which following the abovementioned Typologies 1–4 and two of which followed Typology 5 or Typologies 5–6. No white circle was simultaneously illuminated accompanying the yellow layouts. One and only one of the 6 yellow triangles matched with the visual representation of the actual sounding triad. For the choice of each yellow triangle on screen, we opted for the closest representative, geometrically speaking, of the triangle corresponding to the penultimate chord (Figure 3).

From one sequence to the next, the visualization of the lattice did not change; the referential pitch classes were adapted to have the 12 highlighted triangles centered as much as possible. The reason for this choice was twofold: to mitigate the potential match between circles and pitch classes by participants with perfect pitch and to avoid potential biases caused by uncentered geometrical representations.

### Procedure

Participants were provided with a tablet and headphones for visualizing the *Tonnetz* and hearing the chords. For each sequence, participants were asked to choose, among the 6 yellow triangles mentioned above, the one which, in their



**Figure 3.** Sequence Major-RP (see Appendix A) used in the experiment. Top: transcription according to the standard music notation. Bottom: visual interface for the participants—of the proposed 6 yellow triangles, the correct one here is the furthest left. The white overlapping arrow—which was not shown during the experiment—highlights the visual unfolding of the sequence.

opinion, best fitted with the last chord they heard; to input their choice, they were provided with an electronic pen. This choice was captured on the tablet, along with the time taken to make the decision; the latter was measured from the start of the last chord—and the illumination of the yellow triangles—to the instant at which the pen touched the tablet surface. Participants were not given feedback on whether they picked the correct triangle.

For all the 12 sequences—whose order was always randomized—the task was performed twice. Before the first trial, participants did not receive any information about the *Tonnetz* structure. At its end, they watched a short video tutorial<sup>7</sup> in which the following explanations—in this sequence—were provided by means of mathematical terminology and aural examples:

1. Pitch classes are represented by small circles.
2. Each one of the three geometrical directions stands, respectively, for a series of perfect fifths, major thirds, and minor thirds.
3. Red triangles pointing to the right correspond to major triads.
4. Blue triangles pointing to the left correspond minor triads.
5. Any chord among the 24 possible major and minor triads can be represented by a triangle on the plane.
6. Each triad-related triangle is surrounded by three other triangles, each of which share two pitch classes with the former.
7. Two triads with no shared pitch classes are represented by distant triangles.

After the video tutorial, participants repeated the task. Before doing this, they were allowed a quick clarification about the tutorial; they were given less than 30 s to restart the experiment to avoid deep reflection. After the second trial, a questionnaire was provided to explore the strategies adopted by the participants to do the tasks. This was also the context in which they could communicate further introspection.

### Data Analysis

We analyzed the significance of the following results using several robust methods. We generally chose non-parametric tests, as normal distributions of the data could not be assumed a priori. Consequently,  $p$ -values for significance were calculated through the Mann–Whitney test when the samples were independent—when comparing different groups—and through the Wilcoxon signed-rank test otherwise—when comparing two attempts, that is, before and after the tutorial within a same group. In addition, we used the Shapiro–Wilk test to flag potential non-normal distributions.

In this study, groups were defined a priori based on the quality and the level of formal music and/or science training. We assumed that such a classification might significantly impact the responses, depending on different reasoning mechanisms related to individual expertise. To check the reliability of this hypothesis, after a close qualitative inspection of the data, we performed a discriminant analysis based on the test of the evenness of eigenvalues of similarity matrices (Feoli & Ganis, 2019). This non-parametric analysis technique is constructed to determine if at least 2 of the 4 groups differ significantly from each other by considering all the variables—both responses recorded on the tablet and from the questionnaire. The test provides statistical significance of group separation based on the fact that separate groups tend to have independent sets of eigenvalues of their similarity matrices—calculated by means of a similarity index. For example, if a matrix contains 4 completely separated groups—as supposed in our case—one would expect that the maximal entropy of the eigenvalues is  $\ln 4$  and the evenness is 1 (Shannon, 1948). However, this would be a very extreme case. To test differences in the similarities between groups, the total similarity matrix was rearranged into 4 groups, and the test of eigenvalues was calculated by means of a permutation technique estimating how many times the evenness was higher than the starting classification. Evenness is defined as follows,  $k$  being the group number and  $\lambda_i$  the eigenvalues:

$$E = - \frac{\sum_{i=1}^k \frac{\lambda_i}{\sum \lambda_i} \ln \frac{\lambda_i}{\sum \lambda_i}}{\ln k}$$

For a measure of similarity, we applied the Jaccard function (Podani, 2021). This is the weighted ratio between the intersection and the union of 2 sets depending on quantitative

variables. Let  $a$  and  $b$  be the scores of the  $i$ -th variable in the objects—that is, participants— $A$  and  $B$ , where all data are rescaled between 0 and 1. The function is defined as follows:

$$J(A, B) = \frac{\sum_{i=1}^k a_i b_i}{\sum_{i=1}^k a_i^2 + \sum_{i=1}^k b_i^2 - \sum_{i=1}^k a_i b_i}$$

The Jaccard function ranges between 0 and 1—which correspond, respectively, to zero and full similarity. The general test of separation of the a priori-defined groups was followed by application of the same test to compare groups two by two—that is, by considering the matrix originated by only two groups at a time—in analogy with the group comparison technique generally adopted in analysis of variance (ANOVA). The significance of the test was estimated after performing 10,000 permutations (Manly, 2006).

For all the variables, we finally extracted a fuzzy set by averaging similarities, as suggested by Feoli and Zuccarello (1986). Fuzzy sets, according to Zadeh (1978), are defined as sets each element—in our case, each participant—of which is associated with a value indicating the relationship with the set itself—within groups—and a value indicating the relationship with other sets—between groups. Such values are called degrees of belonging: high values indicate strong appurtenance to the set; low values indicate weak appurtenance to the set. For each variable, we calculated the average value within the 4 groups and normalized such values so that the sum was equal to 1. The interpretation of this calculation is that the variable with the high value has the highest degree of belonging to the set in question and can be therefore used as a discriminant between sets.

## Results

### Accuracy of the Tablet Answers

The most straightforward approach for measuring the accuracy of the participants' performance comes from the estimation of the correctness in their answers when interacting with the tablet (Table 2). Due to our experimental design, the expected value for the number of correct answers when randomly picking the yellow triangles is 2. Groups SciPro, SciStu, and MusStu obtained average results very close to this expected value in their first attempt. Only the MusPro participants clearly exceeded this value, being significantly different when compared with groups SciPro ( $p = .012$ ), MusStu ( $p = .005$ ), and SciStu ( $p = .005$ ). All participants slightly improved their average score during the second attempt after the tutorial. As a result, the difference from the MusPro participants remained significant—barely, though—when compared with groups MusStu ( $p = .050$ ) and SciStu ( $p = .032$ ); however, the comparison between first and second attempts per group did not lead to significant differences in any group.



**Table 2.** Mean and standard deviation for correct answers in the four different groups, before (attempt 1) and after (attempt 2) the video tutorial. The total statistics are also split into three equally sized categories of stimuli depending on the number of shared pitch classes (s.pc.) between the two last chords of the sequences.

Group	Attempt 1				Attempt 2			
	Total	0 s.pc.	1 s.pc.	2 s.pc.	Total	0 s.pc.	1 s.pc.	2 s.pc.
MusPro	3.32 ± 1.46	0.64 ± 0.66	0.73 ± 0.88	1.95 ± 1.09	3.50 ± 2.06	0.68 ± 0.84	1.32 ± 1.09	1.50 ± 1.06
SciPro	2.23 ± 1.02	0.32 ± 0.65	0.41 ± 0.73	1.50 ± 1.14	2.55 ± 1.26	0.23 ± 0.43	0.82 ± 0.96	1.50 ± 0.74
MusStu	2.09 ± 1.27	0.36 ± 0.58	0.50 ± 0.60	1.23 ± 1.07	2.32 ± 1.70	0.59 ± 0.80	0.82 ± 0.85	0.91 ± 0.68
SciStu	1.95 ± 1.40	0.41 ± 0.59	0.50 ± 0.80	1.05 ± 1.05	2.27 ± 1.12	0.23 ± 0.53	0.68 ± 0.72	1.36 ± 0.90
Mus	2.70 ± 1.49	0.50 ± 0.63	0.61 ± 0.75	1.59 ± 1.13	2.91 ± 1.96	0.64 ± 0.81	1.07 ± 1.00	1.20 ± 0.93
Sci	2.09 ± 1.22	0.36 ± 0.61	0.45 ± 0.76	1.27 ± 1.11	2.41 ± 1.19	0.23 ± 0.48	0.75 ± 0.84	1.43 ± 0.82

Further patterns concerning correct answers emerge when we take into account the number of shared pitch classes between the two last chords of the sequences. For all the groups, the greater the number of shared pitch classes, the higher the average amount of correct answers in both the first and the second attempts. In particular, correct answers for two pitch classes in common—that is, when the last movement is parsimonious—are greater than the sum of those with fewer notes in common in the first attempt. However, the second attempt only led to better scores for parsimonious closures of the sequence for the SciStu participants; in this case, group SciPro scored a similar mean, and, unexpectedly, both MusPro and MusStu participants got worse average results. As the visual bias discussed below may have had an important impact on these results, we did not incorporate them into the overall discussion.

### Further Geometric Features of the Tablet Answers

The accuracy of the answers is perhaps an overly rigid—and evidently poor—approach, as it only measures the exactness of the participants' responses in terms of true or false. Among the false possible responses for each stimulus in the experiment, there were, however, important differences: some of them were closer to the true one; others were very far. Consequently, the inspection of further aspects of the answers—beyond strict accuracy—may shed better light on the evaluation of our first hypothesis. With this objective, our study design made room for partial or indirect inspections on the answers in terms of their spatial distribution, which may help to reveal patterns or tendencies beyond the previous data analysis.

As major and minor chords match with different layouts of the triangles within the *Tonnetz*, the orientation of the selected triangle can be regarded as an indicator of the recognition of the quality of the correct answer. Besides, among the 6 possible yellow triangles per sequence, the one continuing the direction traced by the blue and red ones was always an option. This fact led us to consider that it could be, visually speaking, the most biased option among the possibilities offered. This does not mean that such a biased answer was automatically wrong, as it

**Table 3.** Mean and standard deviation for quality recognition of the last chord of the sequences and for the responses matching with the induced visual bias of the experiment. Outcomes are presented for the four different groups, before (attempt 1) and after (attempt 2) the video tutorial.

Group	Quality		Visual bias	
	Attempt 1	Attempt 2	Attempt 1	Attempt 2
MusPro	7.05 ± 1.79	8.32 ± 2.44	4.68 ± 2.88	3.77 ± 1.54
SciPro	5.91 ± 1.15	6.68 ± 1.21	6.55 ± 3.63	5.82 ± 3.70
MusStu	6.18 ± 1.50	7.45 ± 2.42	4.00 ± 3.51	4.00 ± 2.09
SciStu	6.41 ± 1.53	6.32 ± 1.25	4.50 ± 3.07	4.00 ± 2.02
Mus	6.61 ± 1.69	7.89 ± 2.44	4.34 ± 3.19	3.89 ± 1.82
Sci	6.16 ± 1.36	6.50 ± 1.23	5.52 ± 3.48	4.91 ± 3.09

represented the accurate answer for 2 of the 12 sequences we provided. However, its stronger visual connection with the previous visually unfolded pattern for all the stimuli should not be neglected.

Table 3 summarizes the analysis of quality recognition and biased answers. Due to our experimental design, the expected value for the number of answers matching the correct quality when randomly picking the yellow triangles is 6. Again, groups SciPro, SciStu, and MusStu obtained average results very close to this expected value in their first attempt before the tutorial; the MusPro participants reached a slightly higher value—although not statistically significant—when compared with the other groups. This feature clearly changed in the second attempt. All groups except for the SciStu participants clearly improved their scores, and the MusPro participants became a significantly differentiated population when compared with groups SciPro ( $p = .013$ ) and SciStu ( $p = .003$ ). In addition, the improved results in the second attempt were statistically significant—barely, though—when compared with the ones obtained in the first attempt for groups MusPro ( $p = .045$ ), SciPro ( $p = .033$ ), MusStu ( $p = .038$ ).

Our analysis confirmed the presence of the privileged patterned answer mentioned above, in terms of visual bias. Again, due to the experimental design, the expected value for the number of spatially biased answers when randomly picking the yellow triangles is 2. All groups doubled such a value in their first attempt, and the SciPro

participants even tripled it; the latter were significantly different when compared with the MusStu participants ( $p = .017$ ). All groups except for the MusStu participants lowered their biased answers—not greatly, though—in the second attempt, but no statistically significant difference was found here. A closer look at the information on induced visual bias shows a singular distribution of the data (Figure 4). There is a number of participants within every group that always opted for choosing the visually biased response—value 12—during their first attempt. Such extreme values—and also close ones—vanished during the second attempt for all groups with the exception of the SciPro participants; this group even featured a higher number of people who always opted for the biased triangles. Indeed, this distribution is the only one that returned a significant value ( $p = .029$ ) when testing its non-normality by means of the Shapiro–Wilk test. In other words, the subgroup of extremely biased SciPro participants behaved beyond the norm, which translated into a bimodal distribution—maximum values for the SciPro distribution occur at both 3 and 12.

A second indirect method for measuring the correctness of the answers considers the notion of distance. Considering previously defined geometric distances on the *Tonnetz* (e.g., Krumhansl, 1998; Milne & Holland, 2016; Tymoczko, 2009), we opted for observing the error in the recognition of the number of shared notes between two chords. Consider the number of shared pitch classes between the penultimate sounding chord of each sequence and the one matching with the picked yellow triangle ( $s_p$ ); consider also the number of actually shared pitch classes between the two last sounding chords of each sequence ( $s_a$ ). We

define the perceived distance error ( $d$ )—which is not a mathematical distance but a measure of error—between the correct triangle and the picked one as follows:

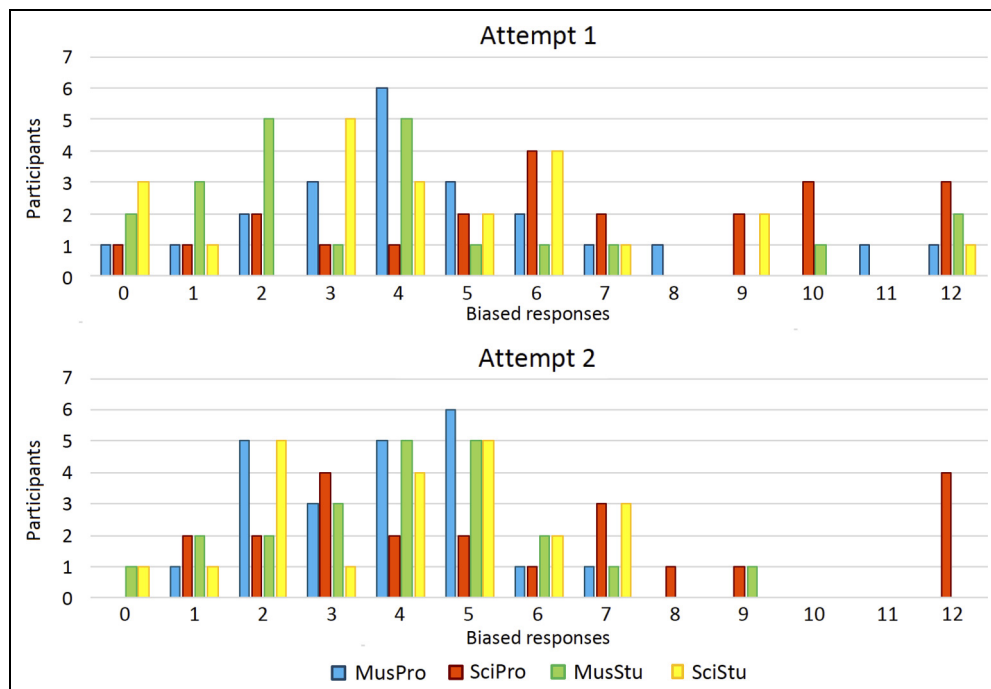
$$d = |s_p - s_a|$$

By simply taking values 0, 1, or 2, this measure estimates the error well—lower is better. Locally speaking, this means measuring the error in the recognition of the number vertices on the *Tonnetz* that are shared with the triangle corresponding to the penultimate chord.

Here, the expected value for the perceived distance error, when randomly picking any yellow one, is 0.89.<sup>8</sup> The mean for all of the groups was slightly lower than this in their first attempt, and only Mus participants—especially MusPro ones—strongly lowered their mean after the video tutorial (Figure 5). In the second attempt, MusPro participants were significantly different when compared with groups SciPro ( $p = .008$ ) and SciStu ( $p = .003$ ); the comparison between the two attempts by MusPro participants was also significant ( $p = .011$ ).

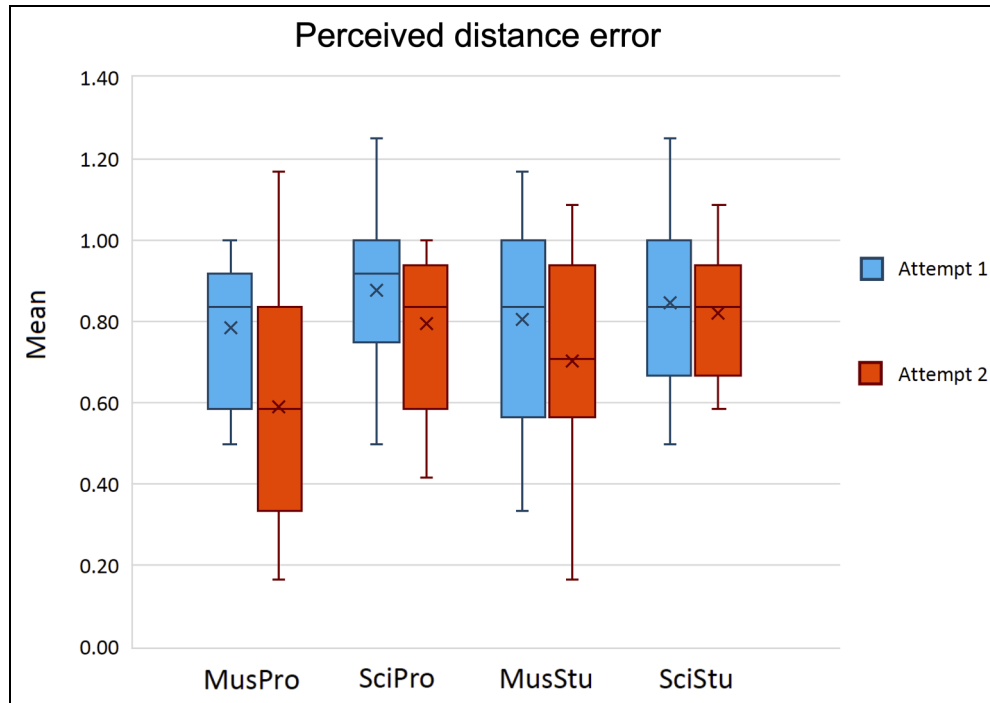
### Response Time

The previous results were based on data that were easy to compare through participants. Dealing with their response time is, conversely, more complex as each person could have had particular needs and adopted individual strategies when choosing their responses; this may have affected their timing in absolute terms. However, as the response time was measured in both attempts, we were able to compare the two responses for a given participant. We define the time ratio as the quotient between the response time for



**Figure 4.** Histograms for the four groups of participants concerning responses that matched with the induced visual bias.





**Figure 5.** Means of the perceived distance error, before (attempt 1) and after (attempt 2) the video tutorial. Crosses in the boxplots stand for the means of the individual means.

any sequence after and before watching the video tutorial. For a clearer analysis, we have adopted a logarithmic scale, so that positive values indicate a longer response time in the second attempt and negative values indicate the opposite.

Once the time ratios—one for each sequence—were averaged, we used the means per participant as values for comparing different groups (Figure 6). The median and the mean of these time-ratio individual means were positive—not higher than 0.50, though—for all the groups, which implies that the response time was, on average, longer for the second attempt independent of the group. Although these medians and means were slightly higher for groups MusPro and MusStu, the differences were not statistically significant when compared with groups SciPro and SciMus. The same analysis was carried out by splitting the data according to the number of shared pitch classes between the penultimate and the last chords of the sequences. Again, the median and mean of the time-ratio individual means were always positive, but in the case of two pitch classes in common, scores were lower independent of the group. Within each group, differences depending on the number of shared pitch classes were not statistically significant except for the MusPro participants; here, the comparison of the time ratio—after removing an outlier—between harmonic sequences ending with 1 and 2 pitch classes in common was significant ( $p = .037$ ). In addition, 28.41% of all the participants provided on average faster responses during the second attempt when the harmonic endings included 2 shared pitch classes, a

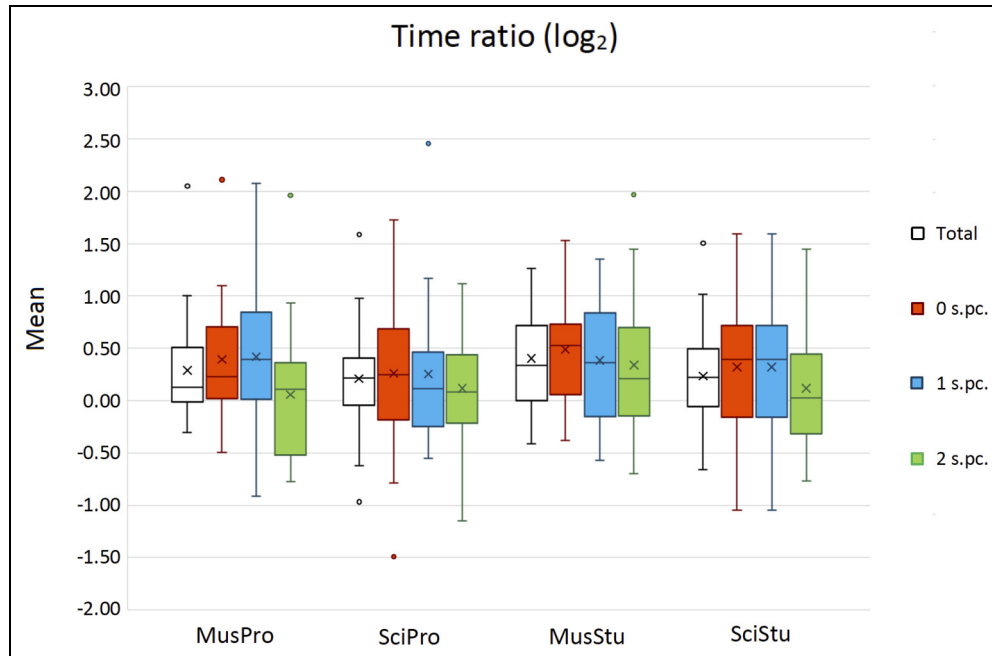
greater percentage compared with harmonic endings when only 1 pitch class was shared (22.73%) or none (17.01%).

### Response Strategies

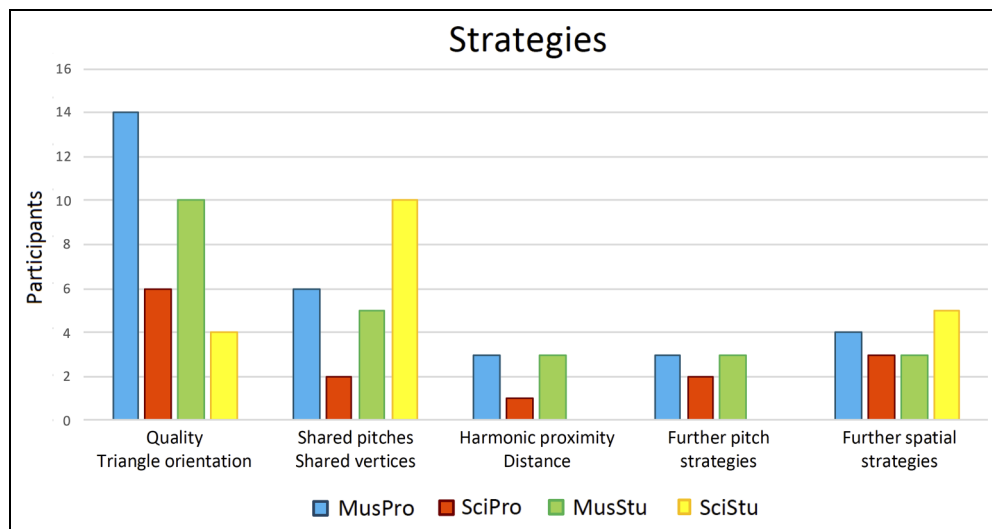
As stated above, participants completed a questionnaire in which they were asked to detail their strategies for choosing the yellow triangles after watching the tutorial. Although we did not provide aprioristic response options, it was possible to detect 5 main tendencies—not necessarily exclusive—in the participants' answers:

1. An explicit reference to the quality of the chords and/or the orientation of the triangles.
2. An explicit reference to the number of shared pitches between chords and/or the number of shared vertices between triangles.
3. A reference to concepts like “proximity” or “distance” in relation to musical harmony and/or visual display.
4. Alternative strategies mentioning pitches.
5. Alternative strategies mentioning spatial features.

Strategies 1 and 2 respectively match with information explicitly provided in the video tutorial; Strategy 3 can be regarded as a particular elaboration of the information relating shared pitches and vertices; Strategy 4 mainly captures misconceptions about the spatial representation of pitches or pitch classes on the *Tonnetz*; Strategy 5 gathers



**Figure 6.** Means of the time ratios, for overall responses and split into three categories of stimuli depending on the number of shared pitch classes (s.pc.). Crosses in the boxplots stand for the means of the individual means.



**Figure 7.** Histograms of the participants' strategies after the video tutorial (attempt 2).

additional approaches that are mostly detached from the aural elements of the experiment.

The distribution of selected strategies (Figure 7) shows a clear preference of Strategies 1 and 2 among Mus participants; unexpectedly, a large number of SciStu participants also mentioned Strategy 2. When Mus participants adopted both Strategies 1 and 2, their report were always written in this order: a check of the quality followed by the analysis of the shared pitches. More than a 40% of the SciPro participants provided unclear information about their strategies or left the question unanswered.

All this information comes from introspections after the task, which limits the ability to directly monitor the impact of the tracking itself. However, we have explored potential correlations between the declared strategies and the outcomes in terms of some of the variables previously analyzed. We opted for estimating correlations between variables indicating strategies and outcomes for both attempts; although the strategies exclusively concern the second attempt, comparison of the degrees of correlation between the two attempts may help to endorse the hypothesis of a potential impact of a given strategy. Notice that

correlations with correctness and quality-recognition means that are close to 1 (anticorrelations close to  $-1$ ) indirectly indicate a favorable (unfavorable) impact of the strategy. Conversely, correlations with perceived-distance-error and visual-bias means that are close to 1 (anticorrelations close to  $-1$ ) indirectly indicate an unfavorable (favorable) impact of the strategy.

A pooled analysis of the correlations for Mus and Sci participants (Table 4) did not provide strong values—whether positive or negative—for any strategy; however, some noticeable changes emerged when comparing the two attempts: we focus on variations of correlations greater than .25 in absolute value. Mus participants show increasing variations—interpretable as a favorable impact—in the correlations of Strategy 2 with correctness and perceived-distance-error means. For this group, a decreasing variation—interpretable as an unfavorable impact—is observed in the correlations of Strategy 5 with all the studied means. Concerning the Sci participants, the clearest variation in the correlations is the one between Strategy 1 and quality and perceived-distance-error means.

Some answers corresponding to Strategies 4 and 5 warrant an additional comment. According to Mus participants' answers, some of them believed that the vertices of the *Tonnetz* stood for actual pitches instead of pitch classes, thus facilitating the representation of voice leading. Finally, one of the SciPro participants provided a very insightful answer: "Although I am aware that I am wrong, the only logical answer for me is to follow the alignment of the triangles";<sup>9</sup> not surprisingly, this person scored 12 for the induced visual bias in both attempts.

### Group Similarity

Differences between groups in terms of the evenness of the eigenvalues are summarized in Table 5; the lower the

indexes, the more dissimilar are the subgroups. MusPro participants do not differ significantly from MusStu ones, but they do differ significantly from both SciPro and SciStu participants; group SciPro differs significantly from both MusPro and MusStu participants but not from group SciStu.

The results of the correlation between the fuzzy sets of the groups—for details, see Appendix B—indicated that MusPro participants are characterized by the majority of the variables involved—both in number and in incidence rate. A more modest contribution was observed for groups SciPro and MusSci, while group SciStu is not characterized except by a few of the variables in play. Finally, Table 6 illustrates a complementary measure for group similarity—the Jaccard similarity ratio between the fuzzy sets. Here, the higher the value, the greater the similarity between groups in terms of contribution of the variables. The two measures are consistent with each other, as the first one returns the similarity between groups in a probabilistic way, while the second provides relative weights of each variable within any group.

### Discussion

The analysis of group similarity confirmed that our criteria for distributing participants into the preconceived groups was worthwhile. In particular, this distribution is helpful for addressing the following discussion, mainly based in the participants' previously acquired skills.

### Influence of Skills in Music Theory

According to our first research hypothesis, the results confirmed a clear influence of previous knowledge in music theory on grasping the *Tonnetz* structure. Although the scores measuring the accurateness of the responses could

**Table 4.** Pearson correlations between the standardized means of previously analyzed data (C: correct answer; Q: recognition of quality; D: perceived distance error; B: induced visual bias) and the strategies declared by the participants after the second trial. Values in italics indicate correlations whose absolute values are greater than .30. Bold numbers indicate variations between attempt 1 and attempt 2 whose absolute values are greater than .25.

Mus group	Attempt 1				Attempt 2			
	Mean C	Mean Q	Mean D	Mean B	Mean C	Mean Q	Mean D	Mean B
(1) Quality	.31	.23	-.13	.19	.36	.32	-.21	.02
(2) Shared pitches	.19	.20	-.08	.02	<b>.45</b>	<b>.44</b>	<b>-.43</b>	-.14
(3) Visual distance	.08	.17	-.06	.15	.02	.10	-.13	-.05
(4) Alt. pitches	-.01	-.15	.19	.02	-.19	-.12	.01	-.20
(5) Alt. spatial	.13	.06	-.23	-.05	<b>-.14</b>	<b>-.31</b>	<b>.25</b>	<b>.23</b>
Sci group	Attempt 1				Attempt 2			
	Mean C	Mean Q	Mean D	Mean B	Mean C	Mean Q	Mean D	Mean B
(1) Quality	.01	.10	-.10	-.02	.23	<b>.36</b>	<b>-.37</b>	-.13
(2) Shared pitches	-.22	-.03	-.21	-.23	-.21	-.17	-.01	-.15
(3) Visual distance	-.14	-.24	.12	-.02	-.05	-.06	<b>-.21</b>	-.10
(4) Alt. pitches	.07	-.27	-.08	.06	.11	<b>.18</b>	-.25	.11
(5) Alt. spatial	.11	-.06	.23	.12	-.11	-.05	.07	-.02

**Table 5.** Similarity between the a priori-defined groups of participants: Jaccard indices. *p*-values in italics correspond to significant cases.

	MusPro	SciPro	MusStu	SciStu
MusPro	1.000	< .001	.141	< .001
SciPro	< .001	1.000	.037	.050
MusStu	.141	.037	1.000	.150
SciStu	< .001	.050	.150	1.000

**Table 6.** Similarity between the a priori-defined groups: Jaccard similarity ratios between the fuzzy sets.

	MusPro	SciPro	MusStu	SciStu
MusPro	1.00	.43	.58	.43
SciPro	.43	1.00	.58	.66
MusStu	.58	.58	1.00	.64
SciStu	.43	.66	.64	1.00

suggest that group SciPro performed the task slightly better than group MusStu, outcomes on quality recognition and perceived distance error point in the opposite direction: Both MusPro and MusStu participants achieved better scores for these variables, compared to those in SciPro and SciStu groups, in both attempts; their improvement after the video tutorial was equally more evident. In particular, these results were mostly significant, statistically speaking, when the comparisons involved MusPro participants—especially when considering both attempts. Consequently, previous knowledge in music theory, in particularly good proficiency, led to better results.

It is, however, important to highlight some differences in the data concerning quality recognition and perceived distance error. As MusPro participants already scored significantly better results for quality recognition before the tutorial, and the influence of Strategy 1 was scarcely different in both attempts for this group, we believe that some of its members may have discovered by themselves, during the first attempt, the relationship between the quality of the triads and the triangular layouts of the *Tonnetz*. This hypothesis is consistent with the particularly high score on accuracy for MusPro participants from their first attempt: When randomly picking a yellow triangle with the restriction of correctly choosing the layout for half of the stimuli, the expected value of accuracy is 3 instead of 2, that is, much closer to the score provided by MusPro participants.

Some statistical data seem to also point to features of quality recognition affecting Sci participants' responses. The statement of Strategy 1 is the most influential one regarding correlations of their results. It is, however, highly unlikely that, just by means of a short video tutorial, Sci participants were able to truly conceptualize the idea of major and minor chords from their qualia, as well as their geometrical correspondences within the *Tonnetz*. This statement apparently goes against the statistically significant

improvement—barely, though—of SciPro participants when performing their second attempt, but we can provide arguments in support of this. Previous research with naive listeners has shown different responses to major and minor chords or tonal keys, like a preference for more harmonicity—i.e., major chords in our context (McDermott et al., 2010)—or different emotional meanings (Bakker & Martin, 2015). Our hypothesis is that, rather than being moved by a factual conceptualization and recognition of the quality of triads, those who improved their scores could have been guided by intuition, based on a sense of consonance or felt emotion while hearing the last chord.

Results from analysis of perceived distance error are more straightforward to interpret. Only the Mus participants clearly improved during the second attempt—with statistical significance in the case of group MusPro—and Strategy 2 appeared to be the most influential when looking at the correlations between different variables. In contrast, neither perceived distance error nor Strategy 2 played a meaningful role for Sci participants. Both aspects underline the need of prior knowledge in music theory, in combination with a short introductory explanation, for grasping—at least partially—the geometrical configuration of the *Tonnetz* beyond the very basic feature of its two elementary triangular layouts.

### *Induced Bias of the Tonnetz Geometry in Non-Functional Musical Contexts*

In considering our second research hypothesis, we did not find the majority of the previous literature on multimodal features of music perception and cognition particularly meaningful. Although the effect of sound—especially music—accompanying moving images has raised some interest in the field of cognitive science—with a particular focus on affect (e.g., Boltz et al., 2009; Cohen, 2015) and multisensory integration (e.g., Lewald & Guski, 2003; Schmiedchen et al., 2012)—knowledge about consistent interactions of music with its visual counterpart representations is sparse.

As suspected, the geometrical directions induced by our stimuli on the *Tonnetz* had a stronger influence on the responses than any potential aprioristic image schema. Previous research confirms the predominance of the vertical schema of pitches in the evaluation of sound coordination with physical motion, whether imagined or visualized—for a review, consult Eitan (2017). Nevertheless, aural stimuli in these studies were often simpler—elementary melodic motions—than those we used—complex four-voice harmonic sequences. The most evident result in our study is the relationship between the visual alignment of the triangles and the participants' skills in music theory. Results also showed the existence of a strong visual bias in the choice of the yellow triangle for all the participants in their first attempt, and also how it was mitigated after the video tutorial—mostly for the Mus participants. In addition,

a subset of group SciPro behaved differently in the second attempt by always choosing the visually biased option. In the absence of formal knowledge of music theory, the use of geometrical logic, regardless of the aural features of the stimuli, was the best—and the only—option for a significant subset of participants proficient in visual abstraction. This seems to even reflect a lack of focus on—or failure to grasp anything from—the harmonic sequence being played. A subtler potential bias could be hidden in the answers by some Mus participants, this time involving, perhaps, the vertical schema for pitches: since several of those who adopted Strategy 5 mentioned pitch classes instead of pitches, and the correlation of this strategy was worse in the second attempt, we believe that such participants probably misunderstood the tutorial and tried to mentally project incongruent pitch schemata onto the *Tonnetz*.

### Further Cognitive Issues around Non-Functional Musical Contexts

The use of sequences based on the PRL-family of neo-Riemannian operations deserves an additional comment. Musical passages based on a RL-progression are found, for instance, in Beethoven's oeuvre, LR-progressions are present in Chopin's work, and PL-progressions appear in Brahms' music (Cohn, 2012). However, these kinds of progressions are proportionally scarce when compared with other passages of the common-practice period, which deal with more evident functional harmonic relationships.

Krumhansl's detection of the overlapping features in her model and in the *Tonnetz* (1998) works well in terms of contextual functional harmony. Relevant studies in the psychology of harmony—some of which were carried out by her—revealed the importance of the musical context for recognition of harmonic hierarchies and creation of expectation, without neglecting tonal modulation (e.g., Bharucha & Krumhansl, 1983; Bigand & Pineau, 1997; Krumhansl et al., 1982; Tillmann et al., 1998). Our stimuli were based on parsimonious harmonic relationships, instead of traditional tonal functions, and almost all of them—except for Major-RL and Minor-LR, see Appendix A—unfolded beyond diatonicism. All these facts, in combination with the visual counterpart of the experiment, demand a deeper consideration of the concepts of perceived proximity and expectation in the context of our study.

The scrutiny of the response time when performing a task may raise a controversial debate regarding the analysis of intuitive and reflective thinking (Isler & Yilmaz, 2023). However, in our study it points to clear evidence: larger response times in the second attempt revealed, on average and regardless of the group, a more thoughtful consideration for picking the yellow triangle. This depended neither on the reported reasoning strategies nor on the success in the choice of the triangle. Even those who barely understood the tutorial were aware of the existence

of a robust logic behind the pitch space, and this could lead them to take more time, although their responses might have been basically intuitive. Other factors leading to this outcome may be related to the experimental conditions: feeling pressure to provide a correct answer, or the fact of being observed. Besides, the smaller time ratio when the last chord was reached parsimoniously—that is, with 2 pitch classes in common with the previous one—again regardless of the group, can be explained as due to contextual proximity. After the tutorial, participants responded faster if there was a strong resemblance between the two last chords in terms of common pitches through the voice leading. This does not mean, however, that their responses were necessarily correct; remember that groups MusPro and MusStu achieved worse scores in terms of accuracy during the second trial and in the case of parsimonious closures.

One might be tempted to say that, on average, the more pitch classes were in common, the lower the time ratio between the two attempts, but this statement fails in the case of group MusPro. Their higher proficiency in music theory probably explains why they required more time to perform the second trial when there was only 1 pitch class in common between the two last chords. This situation matches with Typologies 3 and 4, as explained in the Methods section, in which the roots of the last chords are separated by a perfect fourth or by a perfect fifth. Typologies 3 and 4 also differ in terms of quality, unlike what happens with Typologies 1 and 2, and with Typologies 5 and 6. During the experimental design, we predicted that such a difference would have facilitated the recognition of these cases for group MusPro after the tutorial—and this probably happened, given the clear improvement in the accuracy. We equally predicted—wrongly this time—faster responses in their second attempt. The best hypothesis to explain this divergence from our prediction arises from a collision between non-functional and functional harmony. Typologies 3 and 4 might create a functional cadential sensation, which clashes with the preceding parsimonious non-functional voice leading. Were MusPro participants perhaps trying to discover a rational cadential logic within the *Tonnetz* structure—probably without success, as that was not self-evident from the limited information of the video tutorial?

The following example may illustrate this hypothesis. Consider sequence Major-PL (Figure 8): the expected



**Figure 8.** Transcription, according to standard music notation, of sequence Major-PL (see Appendix A).

seventh triad, within the logic of the unfolded parsimonious progression, would be a B major chord, that is, a return to the first triad by keeping exactly the same pitches in the upper voices—therefore completing a hexatonic cycle (Cohn, 1996), as happens, for instance, in Brahms' *Double Concerto* op. 102. Instead, the seventh triad is a semitone lower than expected, and this creates a strong contrast in the context of parsimonious voice leading. At the same time, the perfect-fourth motion of the root when reaching the last chord induces an accomplished cadential effect: By considering Bb[.18] major as the tonic chord, the last four triads unfold a vi-IV-iv-I progression, that is, the pronounced settling of a minor plagal cadence.

### Limitations of the Study and Future Research

A clear limitation in the design of our experiment comes from the presentation of the questionnaire only after having completed the tasks on the tablet, instead of having requested deep introspection after each attempt. Our choice was aimed at not interrupting the interaction with the audiovisual material, but clearly weakened the analysis of participants' strategies. In addition, the information collected for this purpose has some gaps: The fact that Strategy 1 always preceded Strategy 2 when both were declared by the participants does not automatically entail this order while performing the task after the tutorial. Even if the order reported in the questionnaire reflects what participants truly did, we cannot know if such an order might have been induced by how data were presented in the tutorial—which was, however, more coherent for an appropriate understanding.

Coloring the *Tonnetz* may have also induced an undesired bias. We opted for a warm color matching with major-triad triangles and for a cold color matching with minor-triad triangles, followed by a third different color for the possible answers. Although no participant reported any issue concerning colors in their declared strategies, we cannot measure the potential influence of our choice—that is, colors might perhaps have been pivotal for some participants in the recognition of layouts. We are aware that the use of colors is not an optimal option for settling cross-modal correspondences involving pitch, harmony, or tonal keys, as research on synesthesia reports weak or arbitrary relationships in this sense (e.g., Isbilen & Krumhansl, 2016; Letailleur et al., 2020; Petrović et al., 2012); therefore, other options might have been more appropriate.<sup>10</sup> Moreover, auditory-visual synesthesia may have geometric counterparts (Chiou et al., 2013); this condition could have been taken into account, perhaps, as an additional criterion for participants' exclusion.

We have purposely avoided the use of the term “learning” in the Methods and Discussion sections. Of course, watching the tutorial led many participants to learn something about the *Tonnetz*, and this had an impact on the results during the second attempt—especially for those skilled in music theory. Nevertheless, the whole experiment

was conceived for inspecting a first-sight interaction. Based on our current results, we are able to plan long-term experimental protocols for a more detailed measurement of the learning process; they may include, for instance, methods for eye tracking (Fink et al., 2019) to better grasp the visual features of human–computer interaction with the *Tonnetz*. Improvements could also include a more varied cohort of professional musicians, as our group MusPro involved people who mainly focused on the common practice music period. For instance, octatonic systems are important in the context of jazz, as has recently been empirically tested (Cecchetti et al., 2023); such a different population might respond differently when dealing with *Tonnetz*-based software. It could be also interesting to use more varied stimuli in terms of extraharmonic features (Jimenez et al., 2020; Jimenez, 2023). Finally, the detected clash between functional and non-functional harmonic contexts may inspire further research. This could be particularly useful when considering the whole harmonic progression—instead of primarily focusing on the concluding harmonic motion, as we have done in this study.

### Acknowledgments

We warmly thank África Castillo Morales (Faculty of Physics) and Raquel Díaz Sánchez (Faculty of Mathematics) of the Universidad Complutense de Madrid for helping us to recruit participants. We also thank Xavier Hascher for the English translation of the video tutorial, José O. Martins for his careful reading of the first version of our manuscript, and the reviewers for their valuable criticism.

### Action Editor

Adam Ockelford, University of Roehampton, School of Education

### Peer Review

Fabian Moss, Julius-Maximilians-Universität Würzburg, Institut für Musikforschung

### Contributorship

José L. Besada: overall coordination, experimental design, collection of participants' data, interpretation of statistical data, article writing (corresponding author). Erica Bisesi: experimental design, collection of participants' data, statistical analysis of raw data, article writing. Corentin Guichaoua and Moreno Andreatta: software development, article writing.

### Data Availability Statement

According to the consent of the Research Ethics Committee, raw data will be only available upon direct request to the corresponding author.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Ethics Statement

All participants provided informed consent, and the protocol for collecting and managing data was approved by the Research





Ethics Committee of the Universidad Complutense de Madrid (ref. CE\_20210415-04\_HUM).

## Funding

This research was carried on within the ProAppMaMu project, awarded to Moreno Andreatta, which received financial support from the French CNRS through the MITI interdisciplinary programs. We acknowledge the support by the Interdisciplinary Thematic Institute CREA, as part of the ITI 2021-2028 program of the Université de Strasbourg, the CNRS, and the Inserm, funded by IdEx Unistra (ref. ANR-10-IDEX-0002) and by SFRI-STRAT'US (ref. ANR-20-SFRI-0012) under the French Investments for the Future Program. In addition, José L. Besada is funded by a “Ramón y Cajal” grant (ref. RYC2020-028670-I) by the Spanish MICIU/AEI/10.13039/501100011033 and the European Social Fund.

## ORCID iDs

José L. Besada  <https://orcid.org/0000-0002-8469-2889>

Corentin Guichaoua  <https://orcid.org/0000-0003-2432-2162>

## Notes

1. Notational paradoxes are reported: for instance, interval  $E_4\sharp-F_4b$  is ascending, while its pitch motion is descending.
2. Nowadays, such asymmetry would be discussed through the notions of well-formed scales and maximally even sets (Carey & Clampitt, 1989; Clough & Douthett, 1991; Jedrzejewski, 2008).
3. To be fair, these authors were mentioned later in his book on pitch spaces (Lerdahl, 2004).
4. The expression “Euler–Oettingen–Riemann tone net” (Noll, 2018) better captures the historical evolution of the *Tonnetz*, but we will use the more abridged and extended German term throughout this text.
5. The figure was created with the Tonnetz webpage: <https://imaginary.github.io/web-hexachord/>
6. <http://repmus.ircam.fr/moreno/smir>
7. We provide an English version of the tutorial: <https://youtu.be/oUxVbsDsxvg>. The original video was in Spanish, as it was the participants’ mother language.
8. This happens as the expected value is 1 when the number of actually shared pitch classes is 0 or 2, but 0.67 when the number of actually shared pitch classes is 1.
9. Our translation, original in Spanish.
10. One option might have been the use of visual roughness, based on its cross-modal relationship with harmonicity (Giannos et al., 2021). Another option would have taken into account visual brightness, as some cross-modal correspondence with tonal hierarchies has been reported (Maimon et al., 2020). We chose not to consider these options as we feared they would have generated confusion in the context of our experiment.

## References

- Aminian, M., Kehoe, E., Ma, X., Peterson, A., & Kirby, M. (2020). Exploring musical structure using Tonnetz lattice geometry and LSTMs. In V. Krzhizhanovskaya, et al. (Ed.), *Computational science – ICCS 2020* (pp. 414–424). Springer. [https://doi.org/10.1007/978-3-030-50417-5\\_31](https://doi.org/10.1007/978-3-030-50417-5_31)

- Antović, M., Mitić, J., & Benecasa, N. (2020). Conceptual rather than perceptual: Cross-modal binding of pitch sequencing is based on an underlying schematic structure. *Psychology of Music*, 48(1), 84–104. <https://doi.org/10.1177/0305735618785242>
- Attas, R. (2009). Metaphors in motion: Agents and representation in transformational music analysis. *Music Theory Online*, 15(1). <https://doi.org/https://mtosmt.org/issues/mto.09.15.1/mto.09.15.1.attas.html>
- Bakker, D. R., & Martin, F. H. (2015). Musical chords and emotion: Major and minor triads are processed for emotion. *Cognitive, Affective, & Behavioral Neuroscience*, 15, 15–31. <https://doi.org/10.3758/s13415-014-0309-4>
- Balzano, G. J. (1982). The pitch set as a level of description for studying musical pitch perception. In M. Clynes (Ed.), *Music, mind, and brain* (pp. 321–351). Springer. [https://doi.org/10.1007/978-1-4684-8917-0\\_17](https://doi.org/10.1007/978-1-4684-8917-0_17)
- Bergstrom, T., Karahalios, K., & Hart, J. C. (2007). Isochords: Visualizing structure in music. In *Proceedings of graphics interface 2007* (pp. 297–304). Association for Computing Machinery. <https://doi.org/10.1145/1268517.1268565>
- Bernardes, G., Cocharro, D., Caetano, M., Guedes, C., & Davies, M. E. P. (2016). A multi-level tonal intervalspace for modeling pitch relatedness and musical consonance. *Journal of New Music Research*, 45(4), 281–294. <https://doi.org/10.1080/09298215.2016.1182192>
- Besada, J. L. (2022). Xenakis’ sieve theory: A remnant of serial music? *Music Theory Online*, 28(2). <https://doi.org/https://www.mtosmt.org/issues/mto.22.28.2/mto.22.28.2.besada.html>
- Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13(1), 63–102. [https://doi.org/10.1016/0010-0277\(83\)90003-3](https://doi.org/10.1016/0010-0277(83)90003-3)
- Biasutti, M. (1997). Sharp low- and high-frequency limits on musical chord recognition. *Hearing Research*, 105(1–2), 77–84. [https://doi.org/10.1016/S0378-5955\(96\)00205-5](https://doi.org/10.1016/S0378-5955(96)00205-5)
- Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception & Psychophysics*, 59(7), 1098–1107. <https://doi.org/10.3758/BF03205524>
- Bigo, L., & Andreatta, M. (2014). A geometrical model for the analysis of pop music. *Sonus*, 35(1), 36–48.
- Bigo, L., Ghisi, D., Spicher, A., & Andreatta, M. (2015). Representation of musical structures and processes in simplicial chord spaces. *Computer Music Journal*, 39(3), 9–24. [https://doi.org/10.1162/COMJ\\_a\\_00312](https://doi.org/10.1162/COMJ_a_00312)
- Boltz, M. G., Ebendorf, B., & Field, B. (2009). Audiovisual interactions: The impact of visual information on music perception and memory. *Music Perception*, 27(1), 43–59. <https://doi.org/10.1525/mp.2009.27.1.43>
- Briginshaw, S. B. P. (2012). A neo-Riemannian approach to jazz analysis. *Nota Bene*, 5(1), 57–87. <https://doi.org/10.5206/notabene.v5i1.6581>
- Brower, C. (2008). Paradoxes of pitch space. *Music Analysis*, 27(1), 51–106. <https://doi.org/10.1111/j.1468-2249.2008.00268.x>
- Cannas, S. (2018). Learning geometry and music through computer-aided music analysis and composition: A pedagogical approach. In L. Bigo, et al. (Ed.), *Journées d'Informatique musicale 2018* (pp. 143–146).

- Capuzzo, G. (2004). Neo-Riemannian theory and the analysis of pop-rock music. *Music Theory Spectrum*, 26(2), 177–199. <https://doi.org/10.1525/mts.2004.26.2.177>
- Capuzzo, G. (2006). Pat Martino's *The nature of the guitar*: An intersection of jazz theory and neo-Riemannian theory. *Music Theory Online*, 12(1), <https://doi.org/https://mtosmt.org/issues/mto.06.12.1/mto.06.12.1.capuzzo.pdf>
- Carey, N., & Clampitt, D. (1989). Aspects of well-formed scales. *Music Theory Spectrum*, 11(2), 187–206. <https://doi.org/10.2307/745935>
- Cecchetti, G., Herff, S. A., Finkensiep, C., Harasim, D., & Rohrmeier, M. A. (2023). Hearing functional harmony in jazz: A perceptual study on music-theoretical accounts of extended tonality. *Musicae Scientiae*, 27(3), 672–697. <https://doi.org/10.1177/10298649221122245>
- Childs, A. P. (1998). Moving beyond neo-Riemannian triads: Exploring a transformational model for seventh chords. *Journal of Music Theory*, 42(2), 181–193. <https://doi.org/10.2307/843872>
- Chiou, R., Stelter, M., & Rich, A. N. (2013). Beyond colour perception: Auditory–visual synaesthesia induces experiences of geometric objects in specific locations. *Cortex*, 49(6), 1750–1763. <https://doi.org/10.1016/j.cortex.2012.04.006>
- Chuan, C.-H., & Herremans, D. (2018). Modeling temporal tonal relations in polyphonic music through deep networks with a novel image-based representation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2159–2166. <https://doi.org/10.1609/aaai.v32i1.11880>
- Clough, J., & Douthett, J. (1991). Maximally even sets. *Journal of Music Theory*, 35(1–2), 93–173. <https://doi.org/10.2307/843811>
- Cohen, A. J. (2015). Congruence-association model and experiments in film music: Toward interdisciplinary collaboration. *Music and the Moving Image*, 8(2), 5–24. <https://doi.org/10.5406/musimoviimag.8.2.0005>
- Cohn, R. (1996). Maximally smooth cycles, hexatonic systems, and the analysis of late-romantic triadic progressions. *Music Analysis*, 15(1), 9–40. <https://doi.org/10.2307/854168>
- Cohn, R. (1997). Neo-Riemannian operators, parsimonious tri-chords, and their “tonnetz” representations. *Journal of Music Theory*, 41(1), 1–66. <https://doi.org/10.2307/843761>
- Cohn, R. (1998). Introduction to neo-Riemannian theory: A survey and a historical perspective. *Journal of Music Theory*, 42(2), 167–180. <https://doi.org/10.2307/843871>
- Cohn, R. (2011). Tonal pitch space and the (neo-)riemannian Tonnetz. In E. Gollin & A. Rehding (Eds.), *The Oxford handbook of neo-riemannian music theories* (pp. 322–350). Oxford University Press
- Cohn, R. (2012). *Audacious euphony: Chromaticism and the triad's second nature*. Oxford University Press
- Cohn, R. (2019). Glass graphs. In R. Illiano (Ed.), *Twentieth-century music and mathematics* (pp. 145–170). Brepols
- de Lemos Almada, C. (2020). A theory for parsimonious voice-leading classes. *Musica Theorica*, 5(2), 1–47. <https://doi.org/10.52930/mt.v5i2.139>
- Douthett, J., & Steinbach, P. (1998). Parsimonious graphs: A study in parsimony, contextual transformations, and modes of limited transposition. *Journal of Music Theory*, 42(2), 241–263. <https://doi.org/10.2307/843877>
- Eitan, Z. (2017). Cross-modal experience of musical pitch as space and motion: Current research and future challenges. In C. Wölner (Ed.), *Body, sound and space in music and beyond: Multimodal explorations* (pp. 49–68). Routledge
- Euler, L. (1739). *Tentamen novae theoriae musicae ex certissimis harmoniae principiis dilucide expositae*. Saint Petersburg Academy
- Feoli, E., & Ganis, P. (2019). The use of the evenness of eigenvalues of similarity matrices to test for predictivity of ecosystem classifications. *Mathematics*, 7(3), 1–6. <https://doi.org/10.3390/math7030245>
- Feoli, E., & Zuccarello, V. (1986). Ordination based on classification: Yet another solution?!. In E. Feoli & L. Orloci (Eds.), *Abstracta botanica: Ordination based on classification* (pp. 203–219). Springer Netherlands. [https://doi.org/10.1007/978-94-011-3418-7\\_22](https://doi.org/10.1007/978-94-011-3418-7_22)
- Fink, L. K., Lange, E. B., & Groner, R. (2019). The application of eye-tracking in music research. *Journal of Eye Movement Research*, 11(2). <https://doi.org/10.16910/jemr.11.2.1>
- Giannos, K., Athanasopoulos, G., & Cambouropoulos, E. (2021). Cross-modal associations between harmonic dissonance and visual roughness. *Music & Science*, 4, 1–15. <https://doi.org/10.1177/20592043211055484>
- Graf, M., & Barthet, M. (2023). Reducing sensing errors in a mixed reality musical instrument. In *Proceedings of the 29th symposium on virtual reality software and technology* (pp. 1–2). Association for Computing Machinery. <https://doi.org/10.1145/3611659.3617210>
- Grier, J. (2021). *Musical notation in the West*. Cambridge University Press
- Guichaoua, C., Besada, J. L., Bisesi, E., & Andreatta, M. (2021). The Tonnetz environment: A web platform for computer-aided “mathemusical” learning and research. In B. Csapó & J. Uhoimibhi (Eds.), *Proceedings of the 13th international conference on computer supported education* (pp. 680–689). SciTePress. <https://doi.org/10.5220/0010532606800689>
- Hedges, T. W., & McPherson, A. P. (2013). 3D Gestural interaction with harmonic pitch space. In *10th Sound and music computing conference* (pp. 103–108). <https://doi.org/10.5281/zenodo.850182>
- Holland, S. (1992). Interface design for empowerment: A case study from music. In A. D. Edwards & S. Holland (Eds.), *Multimedia interface design in education* (pp. 177–194). Springer. [https://doi.org/10.1007/978-3-642-58126-7\\_12](https://doi.org/10.1007/978-3-642-58126-7_12)
- Holland, S., Marshall, P., Bird, J., Dalton, S., Morris, R., Pantidi, N., Rogers, Y., & Clark, A. (2009). Running up blueberry hill: Prototyping whole body interaction in harmony space. In *Proceedings of the 3rd international conference on tangible and embedded interaction* (pp. 93–98). Association for Computing Machinery. <https://doi.org/10.1145/1517664.1517690>
- Isbilen, E. S., & Krumhansl, C. L. (2016). The color of music: Emotion-mediated associations to Bach's *Well-tempered Clavier*. *Psychomusicology: Music, Mind, and Brain*, 26(2), 149–161. <https://doi.org/10.1037/pmu0000147>

- Isler, O., & Yilmaz, O. (2023). How to activate intuitive and reflective thinking in behavior research? A comprehensive examination of experimental techniques. *Behavior Research Methods*, 55, 3679–3698. <https://doi.org/10.3758/s13428-022-01984-4>
- Jedrzejewski, F. (2008). Generalized diatonic scales. *Journal of Mathematics and Music*, 2(1), 21–36. <https://doi.org/10.1080/17459730801995863>
- Jimenez, I. (2023). Dos aspectos poco discutidos en la percepción de progresiones armónicas. *Símula: Revista de Teoría y Análisis Musical*, 1(1), 47–68. <https://doi.org/10.59180/29525993.a9378287>
- Jimenez, I., Kuusi, T., & Doll, C. (2020). Common chord progressions and feelings of remembering. *Music & Science*, 3, 1–16. <https://doi.org/10.1177/2059204320916849>
- Krumhansl, C. L. (1998). Perceived triad distance: Evidence supporting the psychological reality of neo-Riemannian transformations. *Journal of Music Theory*, 42(2), 265–281. <https://doi.org/10.2307/843878>
- Krumhansl, C. L., Bharucha, J., & Castellano, M. (1982). Key distance effects on perceived harmonic structure in music. *Perception & Psychophysics*, 32(2), 96–108. <https://doi.org/10.3758/BF03204269>
- Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89(4), 334–368. <https://doi.org/10.1037/0033-295X.89.4.334>
- Lerdahl, F. (1988). Tonal pitch space. *Music Perception*, 5(3), 315–350. <https://doi.org/10.2307/40285402>
- Lerdahl, F. (2004). *Tonal pitch space*. Oxford University Press
- Letailleur, A., Bisesi, E., & Legrain, P. (2020). Strategies used by musicians to identify notes' pitch: Cognitive bricks and mental representations. *Frontiers in Psychology*, 11, <https://doi.org/10.3389/fpsyg.2020.01480>
- Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, 16(3), 468–478. [https://doi.org/10.1016/S0926-6410\(03\)00074-0](https://doi.org/10.1016/S0926-6410(03)00074-0)
- Lewin, D. (1982). A formal theory of generalized tonal functions. *Journal of Music Theory*, 26(1), 23–60. <https://doi.org/10.2307/843354>
- Lieck, R., Moss, F. C., & Rohrmeier, M. (2020). The tonal diffusion model. *Transactions of the International Society for Music Information Retrieval*, 3(1), 153–164. <https://doi.org/10.5334/tismir.46>
- Longuet-Higgins, H. C. (1962). Two letters to a musical friend. *The Music Review*, 23, 244–248 and 271–280
- Maimon, N. B., Lamy, D., & Eitan, Z. (2020). Crossmodal correspondence between tonal hierarchy and visual brightness: Associating syntactic structure and perceptual dimensions across modalities. *Multisensory Research*, 33(8), 805–836. <https://doi.org/10.1163/22134808-bja10006>
- Mandacini, M., Rodà, A., & Canazza, S. (2016). The Harmonic Walk: An interactive physical environment to learn tonal melody accompaniment. *Advances in Multimedia*, 2016, 1–16. <https://doi.org/10.1155/2016/4027164>
- Mandacini, M., Spagnol, S., Ludovico, L. A., Baratè, A., Avanzini, F. (2023). A taxonomy of digital music learning resources. In M. Mandacini et al. (Ed.), *Digital music learning resources* (pp. 53–66). Springer. [https://doi.org/10.1007/978-981-99-4206-0\\_4](https://doi.org/10.1007/978-981-99-4206-0_4)
- Manly, B. F. J. (2006). *Randomization, bootstrap and Monte Carlo methods in biology* (3rd ed). Chapman & Hall. [https://doi.org/10.1111/j.1467-985X.2007.00485\\_5.x](https://doi.org/10.1111/j.1467-985X.2007.00485_5.x)
- Maupin, S., Gerhard, D., & Park, B. (2011). Isomorphic tessellations for musical keyboards. In *8th Sound and music computing conference* (pp. 471–478). <https://doi.org/10.5281/zenodo.849954>
- McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, 20(11), 1035–1041. <https://doi.org/10.1016/j.cub.2010.04.019>
- Milne, A., Sethares, W., & Plamondon, J. (2007). Isomorphic controllers and dynamic tuning: Invariant fingering over a tuning continuum. *Computer Music Journal*, 31(4), 15–32. <https://doi.org/10.1162/comj.2007.31.4.15>
- Milne, A. J., & Holland, S. (2016). Empirically testing Tonnetz, voice-leading, and spectral models of perceived harmonic distance. *Journal of Mathematics and Music*, 10(1), 59–85. <https://doi.org/10.1080/17459737.2016.1152517>
- Moss, F. C. (2014). Tonality and functional equivalence: A multi-level model for the cognition of triadic progressions in 19th century music. In K. Jakubowski, N. Farrugia, G. A. Floridou, & J. Gagen (Eds.), *Proceedings of the 7th International Conference of Students of Systematic Musicology* (pp. 1–8).
- Noll, T. (2018). Dual lattice-path transformations and the dynamics of the major and minor exo-modes. *Journal of Mathematics and Music*, 12(3), 212–232. <https://doi.org/10.1080/17459737.2018.1548035>
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, 111(16), 6104–6108. <https://doi.org/10.1073/pnas.1322705111>
- Park, B., & Gerhard, D. (2013). Rainboard and musix: Building dynamic isomorphic interfaces. In *Proceedings of the international conference in new interfaces for musical expression 2013* (pp. 319–324). <https://doi.org/10.5281/zenodo.1178632>
- Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception*, 41(7), 854–861. <https://doi.org/10.1068/p7225>
- Petrović, M., Antović, M., Milanković, V., & Ačić, G. (2012). Interplay of tone and color: Absolute pitch and synesthesia. In *Proceedings of the 12th ICMPC and the 8th triennial conference of the ESCOM* (pp. 799–806).
- Podani, J. (2021). The wonder of the jaccard coefficient: From alpine floras to bipartite networks. *Flora Mediterranea*, 31, 105–123. <https://doi.org/10.7320/FIMedit31SI.105>
- Riemann, H. (1992). On the imagination of tone. *Journal of Music Theory*, 36(1), 81–117. (Originally published 1873). <https://doi.org/10.2307/843911>
- Schmiedchen, K., Freigang, C., Nitsche, I., & Rübsamen, R. (2012). Crossmodal interactions and multisensory integration in the perception of audio-visual motion – A free-field study. *Brain Research*, 1466, 99–111. <https://doi.org/10.1016/j.brainres.2012.05.015>

- Schoenberg, A. (1969). *Structural functions of harmony*. Norton. (Originally published 1954)
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shepard, R. N. (1982a). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4), 305–333. <https://doi.org/10.1037/0033-295X.89.4.305>
- Shepard, R. N. (1982b). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 343–390). Academic Press. <https://doi.org/10.1016/B978-0-12-213562-0.50015-2>
- Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy. *Music Perception*, 16(1), 99–117. <https://doi.org/10.2307/40285780>
- Tymoczko, D. (2009). Three conceptions of musical distance. In E. Chew, A. Childs, & C. H. Chuan (Eds.), *Mathematics and computation in music. MCM 2009* (pp. 258–272). Springer. [https://doi.org/10.1007/978-3-642-02394-1\\_24](https://doi.org/10.1007/978-3-642-02394-1_24)
- Tymoczko, D. (2011). *A geometry of music: Harmony and counterpoint in the extended common practice*. Oxford University Press
- von Oettingen, A. (1866). *Harmoniesystem in dualer Entwicklung*. Glaser
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synesthetic cross-modality correspondences. *Psychological Science*, 21(1), 21–25. <https://doi.org/10.1177/0956797609354734>
- Weber, G. (1842). *Theory of musical composition*. Wilkins & Carter. (Originally published 1824)
- Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Systems*, 1(1), 3–28. [https://doi.org/10.1016/0165-0114\(78\)90029-5](https://doi.org/10.1016/0165-0114(78)90029-5)
- Zatorre, R. J., & Krumhansl, C. L. (2002). Mental models and musical minds. *Science*, 298, 2138–2139. <https://doi.org/10.1126/science.1080006>
- Zbikowski, L. M. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford University Press

## Appendix A

Sequence	1 <sup>st</sup> triad	2 <sup>nd</sup> triad	3 <sup>rd</sup> triad	4 <sup>th</sup> triad	5 <sup>th</sup> triad	6 <sup>th</sup> triad	7 <sup>th</sup> triad	T 1	T 2	T 3	T 4	T 5	T 5/6
Major RP	El <sup>b</sup> 51, 63, 67, 70	c 48, 63, 67, 72	C 60, 64, 67, 72	a 57, 64, 69, 72	A 57, 64, 69, 73	f <sup>‡</sup> 54, 66, 69, 73	f 53, 68, 72, 77	F <sup>‡</sup> b	D d <sup>‡</sup>	c <sup>‡</sup> E	C <sup>‡</sup> e	f B <sup>b</sup>	g C
Minor RP	d 50, 65, 69, 74	F 53, 65, 69, 72	f 41, 65, 68, 72	A <sup>b</sup> 44, 63, 68, 72	g <sup>‡</sup> 44, 63, 68, 71	B 47, 63, 66, 71	d <sup>‡</sup> 51, 63, 66, 70	b D	G G	f <sup>‡</sup> E	E c	c a <sup>‡</sup>	a <sup>‡</sup>
Major PR	F 53, 69, 72, 77	f 53, 68, 72, 77	A <sup>b</sup> 56, 68, 72, 75	g <sup>‡</sup> 44, 68, 71, 75	B 47, 66, 71, 74	b 47, 66, 71, 74	D 50, 66, 69, 74	D a <sup>‡</sup>	G f	f <sup>‡</sup> F <sup>‡</sup>	E g <sup>‡</sup>	c C	a <sup>‡</sup>
Minor PR	g 43, 67, 70, 74	G 55, 67, 71, 74	e 52, 67, 71, 76	E 52, 68, 71, 76	c <sup>‡</sup> 49, 68, 73, 76	C <sup>‡</sup> 49, 68, 73, 77	g <sup>‡</sup> 49, 68, 73, 77	a <sup>‡</sup> El <sup>b</sup>	f F <sup>‡</sup>	F <sup>‡</sup> a <sup>‡</sup>	g <sup>‡</sup> A <sup>b</sup>	C e	D
Major LR	A 45, 64, 69, 73	c <sup>‡</sup> 49, 64, 68, 73	E 52, 64, 68, 71	g <sup>‡</sup> 44, 63, 68, 71	B 47, 63, 66, 71	d <sup>‡</sup> 51, 63, 66, 70	e 52, 59, 64, 67	El <sup>b</sup> g	F <sup>‡</sup> a <sup>‡</sup>	a <sup>‡</sup> E <sup>b</sup>	A <sup>b</sup> f	e A	a
Minor LR	e 52, 67, 71, 76	C 48, 67, 72, 76	a 45, 69, 72, 76	F 53, 69, 72, 77	d 50, 69, 74, 77	B <sup>b</sup> 46, 70, 74, 77	g 55, 70, 74, 79	g C <sup>‡</sup>	a <sup>‡</sup> F	E <sup>b</sup> a <sup>‡</sup>	f C	A e	E
Major LR	B <sup>b</sup> 46, 62, 65, 70	g 55, 62, 67, 70	El <sup>b</sup> 51, 63, 67, 70	c 48, 63, 67, 72	A <sup>b</sup> 56, 63, 68, 72	f 53, 65, 68, 72	a <sup>‡</sup> 58, 65, 70, 73	C <sup>‡</sup> e <sup>b</sup>	F b	a <sup>‡</sup> F <sup>‡</sup>	C e	e C	b
Minor RL	f <sup>‡</sup> 54, 66, 69, 73	A 57, 64, 69, 73	c <sup>‡</sup> 49, 64, 68, 73	E 52, 64, 68, 71	g <sup>‡</sup> 56, 63, 68, 71	B 47, 63, 66, 71	F 53, 60, 65, 69	e <sup>b</sup> C	b E <sup>b</sup>	F <sup>‡</sup> g	e F	C b	F
Major LP	C 48, 67, 72, 76	e 52, 67, 71, 76	E 52, 68, 71, 76	g <sup>‡</sup> 56, 68, 71, 75	G <sup>‡</sup> 44, 68, 72, 75	c 48, 67, 72, 75	El <sup>b</sup> 51, 67, 70, 75	C c <sup>‡</sup>	E <sup>b</sup> a <sup>‡</sup>	g F <sup>‡</sup>	F g <sup>‡</sup>	b D	f <sup>‡</sup>
Minor LP	c <sup>‡</sup> 49, 64, 68, 73	A 45, 64, 69, 73	a 57, 64, 69, 72	F 53, 65, 69, 72	f 53, 65, 68, 72	C <sup>‡</sup> 49, 65, 68, 73	F <sup>‡</sup> 54, 66, 70, 73	c <sup>‡</sup> B	a <sup>‡</sup> F <sup>‡</sup>	F <sup>‡</sup> g <sup>‡</sup>	g <sup>‡</sup> B <sup>b</sup>	D e	G
Major PL	B 47, 71, 75, 78	b 59, 71, 74, 78	G 55, 71, 74, 79	g 55, 70, 74, 79	El <sup>b</sup> 51, 70, 75, 79	e <sup>b</sup> 63, 70, 75, 78	B <sup>b</sup> 58, 70, 74, 77	B g <sup>‡</sup>	F <sup>‡</sup> c <sup>‡</sup>	g <sup>‡</sup> B	B <sup>b</sup> a	e E <sup>b</sup>	a
Minor PL	g <sup>‡</sup> 56, 63, 68, 71	G <sup>‡</sup> 44, 63, 68, 72	c 48, 63, 67, 72	C 48, 64, 67, 72	e 52, 64, 67, 71	E 40, 64, 68, 71	B <sup>b</sup> 46, 62, 65, 70	g <sup>‡</sup> c <sup>‡</sup>	c <sup>‡</sup> F <sup>‡</sup>	B a	a E <sup>b</sup>	E <sup>b</sup> B <sup>b</sup>	B <sup>b</sup>

Harmonic sequences used in our study, expressed by MIDI values. Capital letters denote major chords, whereas small letters denote minor chords. The name of each sequence reflects the quality of its first triad and the parsimonious transformations involved in the chord progression. The 6 columns on the right side show the triads per sequence which were chosen for visual representatives of the 6 possible typologies (see Table I); each grey cell matches with the true response.

## Appendix B

	M Tr	SD Tr	Mdn Tr	M Tr <sub>0</sub>	SD Tr <sub>0</sub>	Mdn Tr <sub>0</sub>	M Tr <sub>1</sub>	SD Tr <sub>1</sub>	Mdn Tr <sub>1</sub>	M Tr <sub>2</sub>	SD Tr <sub>2</sub>	Mdn Tr <sub>2</sub>
MusPro	0.25	<b>0.32</b>	0.25	0.26	<b>0.30</b>	0.25	<b>0.26</b>	<b>0.30</b>	<b>0.26</b>	0.23	<b>0.38</b>	0.23
SciPro	0.24	0.23	0.23	0.23	0.25	0.24	0.24	0.23	0.24	0.25	0.18	0.24
MusStu	<b>0.27</b>	0.20	<b>0.27</b>	<b>0.27</b>	0.19	<b>0.26</b>	0.26	0.23	0.26	<b>0.28</b>	0.19	<b>0.27</b>
SciStu	0.24	0.25	0.25	0.24	0.26	0.25	0.24	0.24	0.24	0.24	0.25	0.25

	M C - A <sub>1</sub>	SD C - A <sub>1</sub>	Mdn C - A <sub>1</sub>	M C <sub>0</sub> - A <sub>1</sub>	SD C <sub>0</sub> - A <sub>1</sub>	Mdn C <sub>0</sub> - A <sub>1</sub>	M C <sub>1</sub> - A <sub>1</sub>	SD C <sub>1</sub> - A <sub>1</sub>	Mdn C <sub>1</sub> - A <sub>1</sub>	M C <sub>2</sub> - A <sub>1</sub>	SD C <sub>2</sub> - A <sub>1</sub>	Mdn C <sub>2</sub> - A <sub>1</sub>
MusPro	<b>0.44</b>	<b>0.33</b>	<b>0.81</b>	<b>0.49</b>	<b>0.48</b>	<b>0.42</b>	<b>0.43</b>	<b>0.40</b>	<b>0.66</b>	<b>0.45</b>	<b>0.35</b>	<b>0.48</b>
SciPro	0.21	0.25	0.00	0.13	0.10	<b>0.38</b>	0.14	0.14	0.10	0.25	<b>0.26</b>	<b>0.30</b>
MusStu	0.18	0.23	0.19	0.17	0.18	0.10	0.21	<b>0.28</b>	0.02	0.17	0.22	0.13
SciStu	0.16	0.18	0.00	0.21	0.24	0.10	0.21	0.18	0.22	0.12	0.17	0.09

	M C - A <sub>2</sub>	SD C - A <sub>2</sub>	Mdn C - A <sub>2</sub>	M C <sub>0</sub> - A <sub>2</sub>	SD C <sub>0</sub> - A <sub>2</sub>	Mdn C <sub>0</sub> - A <sub>2</sub>	M C <sub>1</sub> - A <sub>2</sub>	SD C <sub>1</sub> - A <sub>2</sub>	Mdn C <sub>1</sub> - A <sub>2</sub>	M C <sub>2</sub> - A <sub>2</sub>	SD C <sub>2</sub> - A <sub>2</sub>	Mdn C <sub>2</sub> - A <sub>2</sub>
MusPro	<b>0.42</b>	<b>0.28</b>	<b>0.97</b>	<b>0.50</b>	<b>0.48</b>	<b>0.48</b>	<b>0.48</b>	<b>0.38</b>	<b>0.67</b>	<b>0.31</b>	0.21	<b>0.43</b>
SciPro	0.22	<b>0.27</b>	0.00	0.06	0.10	0.00	0.19	0.20	0.14	<b>0.31</b>	<b>0.33</b>	<b>0.28</b>
MusStu	0.18	0.20	0.03	<b>0.38</b>	<b>0.39</b>	<b>0.48</b>	0.20	0.21	0.16	0.11	0.20	0.02
SciStu	0.18	0.24	0.00	0.06	0.07	0.03	0.13	0.21	0.04	0.26	0.25	<b>0.28</b>

	M Q - A <sub>1</sub>	SD Q - A <sub>1</sub>	Mdn Q - A <sub>1</sub>	M Q - A <sub>2</sub>	SD Q - A <sub>2</sub>	Mdn Q - A <sub>2</sub>	M D - A <sub>1</sub>	SD D - A <sub>1</sub>	Mdn D - A <sub>1</sub>	M D - A <sub>2</sub>	SD D - A <sub>2</sub>	Mdn D - A <sub>2</sub>
MusPro	<b>0.30</b>	0.24	<b>0.32</b>	<b>0.33</b>	0.19	<b>0.33</b>	0.23	<b>0.26</b>	0.23	0.16	0.20	0.13
SciPro	0.21	0.26	0.18	0.21	0.29	0.23	<b>0.28</b>	<b>0.28</b>	<b>0.26</b>	<b>0.29</b>	<b>0.28</b>	<b>0.32</b>
MusStu	0.23	0.25	0.20	<b>0.27</b>	0.23	<b>0.27</b>	0.23	0.20	<b>0.26</b>	0.23	0.23	0.25
SciStu	0.25	0.25	<b>0.29</b>	0.19	<b>0.29</b>	0.17	0.26	0.26	0.25	<b>0.31</b>	<b>0.30</b>	<b>0.30</b>



	<b>M B - A<sub>1</sub></b>	<b>SD B - A<sub>1</sub></b>	<b>Mdn B - A<sub>1</sub></b>	<b>M B - A<sub>2</sub></b>	<b>SD B - A<sub>2</sub></b>	<b>Mdn B - A<sub>2</sub></b>	<b>S<sub>1</sub> (0/1)</b>	<b>S<sub>2</sub> (0/1)</b>	<b>S<sub>3</sub> (0/1)</b>	<b>S<sub>4</sub> (0/1)</b>	<b>S<sub>5</sub> (0/1)</b>
<b>MusPro</b>	0.22	<b>0.30</b>	0.12	0.18	<b>0.29</b>	0.02	<b>0.58</b>	0.22	<b>0.47</b>	<b>0.40</b>	<b>0.27</b>
<b>SciPro</b>	<b>0.42</b>	0.24	<b>0.63</b>	<b>0.42</b>	0.19	<b>0.78</b>	0.08	0.03	0.06	0.20	0.16
<b>MusStu</b>	0.16	0.21	0.09	0.21	<b>0.26</b>	0.06	<b>0.29</b>	0.15	<b>0.47</b>	<b>0.40</b>	0.15
<b>SciStu</b>	0.20	0.26	0.16	0.20	<b>0.27</b>	0.14	0.05	<b>0.60</b>	0.00	0.00	<b>0.42</b>

Degree of belonging of different variables to each of the a priori-defined groups. Bold numbers indicate significant values—i.e. greater than 0.26, according to Fisher's table for (88-2) degrees of freedom. Data depends on previously analyzed variables (Tr: Time ratio; C: correct answer; Q: recognition of quality; D: distance; B: induced visual bias; S<sub>x</sub>: strategy number x); sub-indices for Tr and C stand for the number of shared pitch classes between the two last chords of the stimuli. Different attempts are equally specified when necessary (A<sub>1</sub>: attempt 1; A<sub>2</sub>: attempt 2). M stands for mean; SD for standard deviation; Mdn for median.