Pre-registered report: Mapping Synesthetic Sequences in Space: Improving Consistency Test by Harnessing Cartography Tools.

Rémy Lachelin, Chhavi Sachdeva, and Nicolas Rothen

Psychology, UniDistance Suisse

Author Note

Rémy Lachelin https://orcid.org/0000-0002-8485-7153

Chhavi Sachdeva https://orcid.org/0000-0002-0074-4371

Nicolas Rothen https://orcid.org/0000-0002-8874-8341

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: *Rémy Lachelin*: conceptualization, methodology, formal analysis, data curation, and Writing - Original Draft. *Chhavi Sachdeva*: conceptualization, methodology, and Writing - Review & Editing. *Nicolas Rothen*: conceptualization, Writing - Review & Editing, supervision, project administration, and Founding Acquisition

Correspondence concerning this article should be addressed to Rémy Lachelin, Psychology, UniDistance Suisse, Schinerstrasse 18, Brig-Glis, Valais 3900, Email: remy.lachelin@fernuni.ch

Abstract

Sequence–space synaesthesia (SSS) is the experience reported by some people of spatial forms elicited by ordered symbolic sequences. SSS is identified using self-report measures (e.g., questionnaires) and behavioural consistency tasks. Consistency tests provide an objective measure to distinguishing SSS from non-synaesthes. Established consistency test’s diagnostic features focus on stimulus-specific measures of distance between repetitions, therefore overlooking ordinal and geometric properties of space-forms. In the *phase I* preregistered study, we first attempt at optimizing SSS diagnostics from consistency test by extracting novel geometric features on a space-form level. We evaluate classification performances of old and new features in a large, aggregated sample (N = 685) from three independent datasets. We harness a geographic toolbox to quantify topological space–sequence forms. Receiver operating characteristic (ROC) analyses are conducted to assess discriminative performances of each features. The results show that permuted topological validity of the space-forms performs as well as the perimeter between stimuli. In an upcoming *phase II*, we will examine predictive validity of the selected features for diagnosing SSS on an independent dataset that has yet to be collected. These findings highlight the relevance of topological principles in shaping space–sequence representations and will attempt at optimizing the consistency tests diagnostic of SSS.

Keywords: Space sequence synaesthesia, consistency test

Pre-registered report: Mapping Synesthetic Sequences in Space: Improving Consistency Test by Harnessing Cartography Tools.

Introduction

Sequence-Space Synaesthesia (SSS) or visuo-spatial forms is the phenomenon where people visualize ordered sequences in particular spatial positions. For example, numbers, weekdays or months (synesthetic *inducers*) are represented as arranged into specific spatial positions in space (synesthetic *concurrent*). The synaesthetic space forms are characteristic and form complex and variable patterns which might differ by categories of stimuli such as number-forms (Galton, 1880) weekdays and months being elliptical, zig-zags or curbed lines (Flournoy, 1893). While SSS is a specific "sub-type" of synaesthesia it also largely overlaps with other types. For example, colour-grapheme synaesthesia where a grapheme *inducer* triggers a *concurrent* colour (i.e. "A" seen as red), co-occurs at 71 % to 76 % with SSS (Sagiv et al., 2006). Across different synesthetic subtypes sub-types, self-reported SSS tend to cluster together (Ward & Simner, 2022), indicating a degree of internal homogeneity and justifying to study it as a distinct phenomenon.

Spatial-forms are idiosyncratic, which results in considerable heterogeneity in how this phenomenon is manifested across individuals. One source of heterogeneity is given by dimensionality: some SSS experiences involve three-dimensional (3D) and two-dimensional (2D) spatial arrangement (Eagleman, 2009; Price & Pearson, 2013). Another source is the reference frame, for example , the spatial forms take place in an external space around the body (*i.e.* projector) or in an egocentric internal space (*i.e.* associator) (Dixon et al., 2004; Smilek et al., 2007). Further variability can be explained by temporal-spatial properties, such as manipulation of their spatial forms such as "zooming" in and out, rotating or shifting perspectives (Gould et al., 2014; Jarick et al., 2009). Lastly, the shape, complexity and layout of the spatial forms are also heterogeneous such as forming for example ovals, lines or zig-zags or loops. With some recurring shapes being more frequent, such as ovals for months (Eagleman, 2009).

Despite the after mentioned heterogeneities, SSS is also phenomenological characterized (Seron et al., 1992). *Automaticity*: the *inducer* automatically triggers the *concurrent*. *Unidirectionality*: while the *inducer* triggers the concurrent, the concurrent does not trigger the inducer. *Developmentally early*: the experience was already present during childhood. *Consciousness*: The concurrent is consciously perceived. *Consistency*: the inducer-concurrent pair remains stable in time within subject. These distinctions can be quantified with self-reported questionnaire (i.e. for development and consciousness), or more objectively using behavioural tests such as consistency tests (Baron-Cohen et al., 1993).

Consistency tests

The rationale behind consistency tests is to measure the variability in *inducer-concurrent* associations over time. These test are used as an objective validation, or genuineness of self-reported synaesthetes, and therefore mainly useful in experimental settings to compare synaesthetes with control and characterize synaesthetic experiences and their origins. Consistency tests have proven effective for colour-grapheme synaesthesia. Measures of individual consistency can be derived using colour pickers while repeatedly presenting the same inducer (i.e. "A"). The quantification of the distance between concurrents from the same inducers (i.e. similar reds) is therefore crucial for discriminating consistent synaesthetes from controls. For colour-grapheme, the euclidean distance in CIE L*u*v colour space, which is designed for perceptual uniformity, yields optimal classification accuracy when using a cut-off estimated as deduced from a larger representative sample (Rothen et al., 2013). A similar rationale to that used for colour-grapheme synaesthesia has been applied to characterize SSS.

In SSS consistency test's task, instead of using a colour picker, participants are asked to repeatedly position each inducer on the computer screen according to the spatial location of their concurrent experience. Brang et al.(2010), evaluated consistency as the distance between repeated responses to the same inducer or stimuli (e.g. January and January) relative to adjacent stimuli (e.g. February). A response was defined as consistent if it fell within 1.96 *z* scores. However, this criterion was noted to be potentially too conservatory, as it identified synaesthesia in 4 of 81 self-reported synaesthetes. Geometrical features such as area and perimeter between the coordinates of repeated inducers have been used as a measure of the distance between same inducers, hence as a metric for consistency. Less consistent responses leading to smaller triangle perimeters and area (Rothen et al., 2016a). A cut-off corresponding to less than 0.203 % average responses area in proportion to the total screen area to classify as SSS has been suggested (see also Ward et al., 2018).

However, several general caveats of consistency tasks using response level metrics have been identified. Some geometrical features may favour certain space-forms (Ward et al., 2018) in particular those with linear or highly regular spatial layouts, i.e. elliptical patterns for months (Brang et al., 2010; Eagleman, 2009; Flournoy, 1893). Another concern is that participants that respond always with the same position, for example when not knowing the response, will have artificially high consistencies (Rothen et al., 2016a). This has led to alternative approaches such as to additionally include standard deviation of responses and questionnaire cut-off (Ward et al., 2018) and permutation-based comparison of individual responses to chance levels for colour-grapheme (Root et al., 2021) and SSS (Ward, 2022).

Present study

The goal of this registered report is to compare different consistency features on their classification performances of individuals with SSS and controls using Receiver Operator Characteristics (ROC) analyses. In the present *Phase I*, we merge previously available datasets to replicate established consistency test methods from the literature and to evaluate additional, novel features. These new features are designed to take advantage of two properties of synesthetic responses.

First, synesthetic spatial forms follow ordinal rules (e.g. Monday → Tuesday → Wednesday). Some studies have systematically investigated ordinality by investigating metrics of adjacent inducer-concurrent pairs such as distances (Brang et al., 2010) or angles (Eagleman, 2009). However, ordinality of the concurrent space-form might be particularly relevant considering all the stimuli of a certain category (i.e. weekdays or months). Second, the space forms of the concurrents often take the form of structured spatial configuration space-forms such as lines or polygons that may follow distinctive geometrical rules. For example circular layouts have been already described in early accounts of number forms (Flournoy, 1893; Galton, 1880).

Attempting to capture both sequential and geometrical properties of synesthetic forms, we harnessed a geospatial analysis package (Pebesma, 2018) to extract geometrical features from participant's 2D (x,y) coordinate responses. This package allows, for example, to build and analyse strings or polygons and then extract multiple geometrical descriptors or features. These individual geometrical features are then compared using ROC analyses to compare their classification performances of syntheses and controls.

In the present *phase I*, we compare ROC on three merged datasets using the same task on both SSS and control groups Ward (2022). Several analytical approaches were evaluated. First, we aimed to reproduce the diagnostic criteria based on stimulus-level consistency metrics, such as area and perimeter.

Second, we explored a new approach base on comparing geometrical features across repetitions of the same categories (weekday, month and numbers).

Third, we applied higher-order analytical approaches such including permutation-based methods. On the one hand, we reproduce Root et al. (2021) permutation test for colour-grapheme synaesthesia and adapted it to SSS as in Ward (2022)). On the other hand, we implemented permutations across repetitions to derive a permuted measure of consistency. Specifically, instead of using repetitions in chronological order, we randomized the order of stimulus presentation and extract the geometrical feature from x,y coordinates of same stimulus categories in non-chronological order.

In a future *phase II*, we will assess whether the features identified *Phase I* generalize and are validated on an independent dataset that is yet to be acquired (registered report on the Open Science Framework: https://osf.io/9efjb/**).**

Phase I. **Methods**

Phase I. **Datasets**

We merged four datasets. Two datasets were collected in laboratory settings [Rothen et al. (2016a) data in https://osf.io/6hq94/files/osfstorage and Van Petersen et al. (2020b) data in https://data.ru.nl/collections/di/dcc/DSC_2018.00019_653]. The two others come from on-line testing [Ward (2022); data in https://osf.io/nu5v4/overview] and additional dataset gently provided by private communications with Pr. Ward. To match the other datasets, stimuli form (Van Petersen et al., 2020b) are translated from Dutch to English and for the stimuli, only numbers from 0 to 9 are included here (excluding 50 and 100). The demographics for each datasets can be found in Table 1.

***Phase I.* Participants**

We kept 685 from the total 689 participants. First, we excluded 0.02 % empty trials (i.e. skipped responses) including trials flagged for having the same x or y coordinates across conditions and repetitions, causing the depletion of 2 participants(as in Rothen et al., 2016b; Ward et al., 2018). An additional 2 participants were excluded for having less than 4 coordinate points since polygons need at least 4 coordinates, see Section 2.5.3. Note therefore, that 31 participants of the final sample do not have responses in all three conditions X repetition cases. x and y coordinates were then separately normalized (*z*-score) per participant.

From the final sample of N = 685, 396 were synaesthetes and 289 controls. Table 2 breaks down the number of synaesthetes and control contributed that are included in the following analyses.

Regarding the synaestheses profiles, we are limited in precise descriptions from the data by Pr. Ward (i.e. 573 participants) since it also included questionnaire responses. We describe the self-reported profiles for the stimulus categories used in the consistency test (i.e. number, weekdays and month), see Figure 1. From this venn-diagram we see that 240 of the SSS report having space-forms for Numbers, Days and Month; 42 only Days and Month; 21 only Number and Day. Also, 127 controls report space-forms for either Days, Numbers, Months or combinations.

***Phase I.* Procedure**

For the consistency test, each stimuli is presented randomly and sequentially centrally on the screen. Participant are instructed to click on the screen position where they visualize them. In Van Petersen et al., (2020b) and Rothen (2016a) the participants were allowed to skip responses.

Stimulus included here are 7 weekdays (Monday to Sunday), 12 months (January to December) and 9 numbers (0 to 9). For Ward's data the stimuli were presented in randomized order with the constraint that no stimulus was repeated until all unique stimuli (N = 29) had been presented once. The median display resolution was 1440 X 768, with a maximum of 2560 X 2025 and a minimum of 308 X 149 .

***Phase I.* Analyses**

First, we reproduced features extracted from consistency tests found in the literature (Root et al., 2021; Rothen et al., 2016a; Van Petersen et al., 2020a; Ward, 2022). These methods compute consistency metrics at the stimulus level, they asses the consistency for each stimulus *within* the repetitions.

Second, we extract features at the space-form level. Taking the stimuli as an ordered sequence we can consider them as a geometrical segments (i.e. open geometrical form) and polygons (i.e. closed geometrical form), similarly as originally described in Galton (1880). First, we consider the space-forms as segments, from here we compute the number of self-intersections. We harnessed a geography package to extract geometry-based features. Informed by the ordinality of the stimulus, we defined segments and polygon by conditions and repetitions. The rationale here is to determine whether, when considering the stimuli as ordered coordinates (i.e. as segments or polygon) they remain consistent *between* repetitions for each individuals.

Since these methods are also rely on repetition order we also compute the best AUC features by permuting them. We predicted that permuted averaged features would yield superior classification performance (higher AUC values) compared to chronologically ordered features. The rationale is that genuine synaesthetes should exhibit stable spatial forms across repetitions regardless of presentation order, whereas controls would show random variation

Each feature's performance in classifying SSS from controls was compared with Receiver Operating Characteristics (ROC) analyses. Area Under the Curve (AUC) was used to determine which feature is best at classifying SSS from Controls. The optimal cutoffs were calculated using Youden's index (Youden, 1950). Discriminant Power (DP) was used as an additional estimate of discrimination performance according to the optimal cutoff (Equation 1). DP around 1 being interpreted as inefficient discrimination.

$$DP = \frac{sqrt(3)}{\pi}\left(log(X) + log(Y)\right) \qquad (1)$$

where: $X = sensitivity/(1 - sensitivity)$ and $Y = specificity/(1 - specificity)$

Additional analyses in the Appendix attempt to address several concerns raised in the literature on consistency tests: that some criteria might be more beneficial toward a certain type of SSS than others, and circularity. Regarding the concern that some criteria might bias toward types of synaesthesia, we compared the groups by subtype of SSS (i.e., weekdays, months, and numbers) with the classifications using the cutoffs for permuted validity and perimeter. In an attempt to circumvent the circularity of perfecting consistency tests based on self-reported groups, we attempted two additional approaches: first, testing subsamples based on percentiles of the questionnaire score sample distribution (see Section 9.4). The rationale is that a good consistency test should have similar discriminative performance for strong SSS and weaker SSS.

To do this, we resampled the SSS and controls based on their questionnaire distribution by percentiles (i.e., comparing 10% highest questionnaire scores vs. 10% lowest, then 20% vs. 20%, etc.). Second we correlated the questionnaire scores with the results from different features extracted from the consistency test. Finally, since we will attempt to predict the best diagnostic features of a to be collected dataset, we also carried analyses across the different datasets separately, to test whether each consistency test's feature remain constant, see Section 9.3 .

Phase I. **Features extraction**

Stimulus level: area and perimeter between repetitions

The stimulus repetition distance's features are calculated as the **area**, see Equation 2, and the **perimeter**, see Equation 3, formed by the x,y coordinate between three repetitions (i.e. $(x1, y1), (x2, y2), (x3, y3)$).

$$Area = (x1y2 + x2y3 + x3y1 - x1y3 - x2y1 - x3y2)/2 \qquad (2)$$

$$Perimeter = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} + \sqrt{(x3 - x2)^2 + (y3 - y2)^2} + \sqrt{(x1 - x3)^2 + (y1 - y3)^2} \qquad (3)$$

Each stimuli's area is then averaged by participants. The area metric is in % of the screen area to be able to compare with the consistency across studies using different screen sizes. In addition, since also individual spread of responses can vary, we calculated area and perimeters on individually z-score transformed x,y coordinates.

Permuted consistency

Next, we reproduce (Root et al., 2021) permuted consistency method. For each individual, the x and y coordinates are randomly permuted and the areas are calculated. After 1000 permutations per individual, a z-score is calculated with the observed means compared to the permuted distribution, see Equation 4. The distribution of the permuted results form an individual chance level distribution, hence the z-score reflects where the observed area lies on the chance level distribution.

$$Zscore = [(Observed) - (MeanPermuted)]/(SDPermuted) \qquad (4)$$

Form level: additional features

First, we extract the number of *self-intersections* of each segments. Conceptually, SSS should have less chance to produce that self-intersect than control. The number of self-intersections are added separately for each repetitions and conditions and averaged per participants.

The next geometrical features are extracted using the simple feature sf package (Pebesma, 2018) to generate ordered segments and polygons based on the individually z-score transformed x and y coordinates. sf has originally been developed for geography.

Topological validity. We assessed topological validity using geometric validation test. Topological validity tests if a polygons is well-formed and valid according to the Open Geospatial Consortium (OGC) Simple Features Specification (Herring, 2010). A polygon is considered topologically valid if it satisfies the following criteria: (1) if polygon rings are simple (i.e., they do not touch or self-intersect), (2) boundary rings to not cross (3) boundary rings may only touch points tangentially (4) rings that define holes are contained within the exterior ring (5) the polygon rings must not splits the polygon (*PostGIS 3.6.2dev Manual*, n.d.).

For each participant we tested for topological validity across individual 3 categories (weekdays, month and numbers) and 3 repetitions, hence 9 forms. The binary outcome (1 = is valid, 0 = invalid) is then averaged across all 9 forms. Hence, it can span from 1 (all 9 are valid) to 0 (none of the 9 forms are valid). Thus, the topological validity score ranges from 0 (none of the nine forms are valid) to 1 (all nine forms are valid). For example, if a participant's forms for weekdays were valid across all three repetitions, months were valid in two of three repetitions, and numbers were invalid in all repetitions, the topological validity score would be 5/9 ≈ 0.56.

Topological simplicity. We assessed topological simplicity, which evaluates whether a geometric object has a simple structure without self-intersections or self-tangencies. According to the OGC Simple Features Specification (Herring, 2010), a polygon boundary is considered simple if it does not pass through the same point more than once (*PostGIS 3.6.2dev Manual*, n.d.). For polygons, simplicity requires that each ring does not self-intersect or self-touch. Note that simplicity is a condition for validity: a polygon can be simple but still invalid (e.g., if an interior ring extends outside the exterior ring).

Finally, we also attempted a correlational approach. Conceptually, consistent coordinates should correlate across repetitions and between vertical and horizontal axes. We present a naive approach was to correlate x and y coordinates across participants and conditions. For example, correlate all x coordinates for weekdays (3*7 = 21) to the y coordinates. Then the absolute correlations are averaged across participants.

Permutation-based feature extraction

The form-based features computed before were relying on the chronologically ordered repetitions. For example, when a stimulus such as Monday was repeated three times, the coordinates for the first presentation of Monday were always paired with the coordinates for the first presentation of Tuesday to construct segments or polygons. However, if synesthetic forms are truly consistent, they should remain stable independently of the chronological order of stimulus presentation. Hence, we permute the repetitions within each condition. Specifically, for each participant and each category, we randomized the response repetition order. For example segments for weekdays were constructed with the x,y coordinates where randomly permuted across repetition order, for example $Monday1^{st}$, $Tuesday^{3rd}$, $Wednesday^{2nd}$, ect.

***Phase I.* Results**

Descriptively, the area between repetition for SSS was surprisingly larger than for other datasets: 0.26%, compared to 0.14% in Rothen (2016a) and 0.15 % in Ward (2018), see Table 4. This difference seems to be mainly driven by the dataset with more participants from (Ward, 2022) with 0.26% area. Note that despite being the largest sample it is also the more variable with regards to areas: *SD = 0.50* % for the Ward's (2022) data compared to *SD = 0.09 %* in Rothen et al. (2016a)], see Table A2. This difference between datasets is diminished when using individually standardized coordinates (M = 0.05 (*SD = 0.06*) (Rothen et al., 2016a), M = 0.08 (*SD = 0.15*) (Ward, 2022), M = 0.09 (*SD = 0.15*), Ward2). Hence, for the perimeter, we only consider standardized coordinates.

ROC analysis

The results of the ROC analyses are summarized in Table 3 and Figure 2. We compared all the ROC's to permuted validity since it has the highest AUC (80.09), using bootstrapping (1000 times), all p-values false discovery rate (FDR) adjusted. We did not find any significant difference in total AUC between permuted validity and perimeter (D = 0.98, $p = 0.35$). But the permuted validity led to significantly higher AUC than the non permuted one (D = 3.05, $p < .05$. Hence the two best AUC performances are from the averaged permuted validity score (AUC = 80.09; DP = 2.06, cut-off = 0.17 corresponding to 1.53 valid of 9 forms) and the other for standardized perimeter between the repetitions of each stimuli (AUC = 78.5; DP = 2.04, cut-off = 1.92 z-scores). However, at the proposed cut-off, the permuted validity leads to higher specificity (77.5 vs. 74.29) hence best to reject controls (i.e. less false negatives) but smaller sensitivity (70.4 vs. 73.6), hence perimeter is best at detecting SSS (i.e. leads to less false positives). See Figure A1 in Section 9.1 for the densities of each standardized features suggesting bi-modal distribution (see also difference in median, see Table A1.

In sum, the permuted validity and perimeter both provide similarly best AUC compared to the other features. While the permuted validity is best to have less false negatives (i.e. controls) while perimeter is best for less false positives (i.e. SSS).

Higher-order analyses

Some features might be more suited when having well defined samples such as "strong" SSS and controls with no space-forms at all than when tested on more general population. To test the stability of AUC of all features, we computed them across several percentiles based on the questionnaire scores. Again, only the data from Pr. Ward is included there. We computed AUC, sensitivity and sensibility on sub samples based on the percentiles of the questionnaire scores(see Ward et al., 2018), taking the 10-0% highest questionnaire scores against the 10 % worst (i.e. 90-100 %) . We further sub-sampled by 10 % questionnaire scores and recalculated AUC, sensitivity and sensibility, see Figure A2 for AUC and Figure A3 for sensitivity and specificity. Using this method we can see that while the total AUC remains stable across percentiles.

We proceeded with the same method to sub-sample the data based on the source of the data and found different features best fit different datasets, see Figure A4 for AUC and DP. Interestingly the best features by datasets differ between the data sources. While the standardized area indeed lead to the best AUC for the data from Rothen et al., (2016a), the perimeter leads to better AUC in Ward et al.,(n.d.). The permuted validity leads to AUC > 80 across all datasets.

Another important point addressed in the literature about consistency test is circularity. The circularity is given in that if consistency tests are used to classify SSS and controls, then those groups will by definition differ in consistency (Root et al., 2025). Hence we made a correlation matrix with the questionnaire scores and all the features, see Section 9.4. The two highest correlations are indeed between the questionnaire score and perimeter (r = .58) and permuted validity (r = .50), see Figure A6.

Phase II Methods

Additional data will be collected in the future using the same consistency test, with the procedural exception that there will be four repetitions per stimuli instead of three. The same feature will be extracted from this dataset. Hence for the stimulus levels feature we will compare the area and perimeter of a rectangle of four coordinates pairs instead of a triangle. Materials are more details on this study are pre-registered on OSF (https://osf.io/pjb6e/?view_only=d467ebf4c1f94076ae4ac61298255065). The population (i.e. sample size and recruitment method) are also pre-registered (https://osf.io/6h8dx**).**

Phase II Analyses

We will compute the same ROC analyses as in this pre-registration on all the features. Since we obtain different best AUC per features depending on the dataset, see Section 9.3, it is likely that the best features here will not match with the best feature in the to be collected dataset.

Phase I. Discussion

Our investigation of four datasets found two main features leading to the optimal classification of SSS from control: standardized perimeter between repetitions and permuted topological validity. While perimeter is calculated on the stimulus level, i.e. distance between repeated stimuli, topological validity is calculated on the whole space-form level, e.g. x,y coordinates formed by Monday to Sunday. Both criteria display bimodal distributions across the groups (see Figure A1 Table A1). We also obtain variable best features by datasets, in particular from the stimulus based metrics (area and perimeter), while topological validity lead to > 80 AUC across datasets. Considering only the data by Pr. Ward that contains questionnaire data, we obtain similar AUC and DP for more extreme Synaesthete and Control groups when sub-sampled by questionnaire's score distribution percentiles (see Figure A2). However, perimeter lead to slightly higher correlation with the questionnaire (r = .58) than topological validity (r = .50).

Limitations

Although an optimal test to classify SSS might be particularly relevant for experimental purposes, several important limitations need to be considered. First, consistency tests contain only a limited set of sequential stimuli (i.e. months, weeks and the first ten natural numbers). Other ordinal categories such as temperature, clock time, musical keys might be more relevant for some individuals with SSS. For numbers specifically, better consistencies might be obtained using a larger set size (i.e. including decades and hundreds), as descriptively interesting form changes occur at different decimals in base-10 number systems (Galton, 1880).

Second, the use of diagnostic cut-offs assumes categorical distinctions between groups, but SSS may exist on a continuum and scores might be more suitable (Price & Pearson, 2013). This limitation is related to the circularity mentioned previously: diagnostic cut-offs as calculated here depend on how SSS and controls are classified in the first place (Simner, 2012). Measures of other characteristics of synaesthesia such as automaticity or visual Gabor detection might be necessary to establish external validity (Ward et al., 2018)). Indeed, determining the prevalence of SSS in the general population requires definitional choices that lead to more conservative or lenient criteria(Brang et al., 2010; Jonas & Price, 2014; Sagiv et al., 2006; Ward et al., 2018).

Indeed, current prevalence estimates spans between 4.4 % (Brang et al., 2013), 8.1 % (Ward et al., 2018) and 14 % (Seron et al., 1992).

Third, as shown in Section 9.3, we find variable optimal criteria varies across datasets. he differences across datasets might be due to different sampling biases, for example (Van Petersen et al., 2020b) recruited SSS from over one hundred candidates to maximise participant reporting synesthetic experiences. Another possible explanation may relate to the tasks option to skip responses, leaving empty cases for some participants. Indeed, having fewer responses (i.e., fewer coordinates) increases the likelihood of producing a valid spatial form. Conversely, perimeter and area are differentially affected by missing responses.

Topological validity

Surprisingly, topological validity of the space-forms led to similar results than the perimeter (or the distance) between the responses. This result might be informative about how SSS map ordinal stimuli in space. Indeed, it seems the patterns of space-forms follow topological rules analogous to geographical space structures. One of the advantages of using topological validity is that it would classify responses with the same coordinates as controls compared to area and perimeter.

Analogies between maps and neuroscience have a long history (i.e. retinotopy, sonotopy or somatotopy) (Eagleman & Goodale, 2009). Therefore space-forms might originate from the mapping between ordinal or sequential stimuli on idiosyncratic visuo-spatial abilities following topological rules.

Interestingly, ordinality is very important an important semantic property of numbers (REF). Moreover, that numbers are acquired sequentially (i.e. 1 is learned before 2) (REF).

Hence the importance of ordinality in SSS is coherent with developmental accounts of Synaesthesia (Price & Pearson, 2013).

Intermediary Conclusion

Phase I compared traditional stimulus-level consistency measures with novel form-level geometric features for detecting SSS. We found optimal classification performance from permuted topological validity and standardized perimeter between repetitions. The success of topological validity as a diagnostic feature is theoretically meaningful: genuine synesthetes should produce spatially coherent, well-formed structures that satisfy geometric constraints compared to controls. Similarly, the perimeter feature captures the stability of the overall spatial configuration across repetitions, which should remain consistent for true SSS individuals regardless of presentation order.

In *phase II,* we will attempt at validating the criteria on a yet-to be acquired dataset.

References

Baron-Cohen, S., Harrison, J., Goldstein, L. H., & Wyke, M. (1993). Coloured Speech Perception: Is Synaesthesia what Happens when Modularity Breaks Down? *Perception*, *22*(4), 419–426. https://doi.org/10.1068/p220419

Brang, D., Miller, L. E., McQuire, M., Ramachandran, V. S., & Coulson, S. (2013). Enhanced mental rotation ability in time-space synesthesia. *Cognitive Processing*, *14*(4), 429–434. https://doi.org/10.1007/s10339-013-0561-5

Brang, D., Teuscher, U., Ramachandran, V. S., & Coulson, S. (2010). Temporal sequences, synesthetic mappings, and cultural biases: The geography of time. *Consciousness and Cognition*, *19*(1), 311–320. https://doi.org/10.1016/j.concog.2010.01.003

Dixon, M. J., Smilek, D., & Merikle, P. M. (2004). Not all synaesthetes are created equal: Projector versus associator synaesthetes. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(3), 335–343. https://doi.org/10.3758/CABN.4.3.335

Eagleman, D. M. (2009). The objectification of overlearned sequences: A new view of spatial sequence synesthesia. *Cortex*, *45*(10), 1266–1277. https://doi.org/10.1016/j.cortex.2009.06.012

Eagleman, D. M., & Goodale, M. A. (2009). Why color synesthesia involves more than color. *Trends in Cognitive Sciences*, *13*(7), 288–292. https://doi.org/10.1016/j.tics.2009.03.009

Flournoy, T. (1893). *Des phénomènes de synopsie (audition colorée): Photismes, schèmes visuels, personnifications*. Alcan. https://books.google.bj/books?id=JISQxpcyGUMC

Galton, F. (1880). Visualised Numerals. *Nature*, *21*(533), 252–256. https://doi.org/10.1038/021252a0

Gould, C., Froese, T., Barrett, A. B., Ward, J., & Seth, A. K. (2014). An extended case study on the phenomenology of sequence-space synesthesia. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00433

Herring, J. R. (2010). *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*.

Jarick, M., Dixon, M. J., Stewart, M. T., Maxwell, E. C., & Smilek, D. (2009). A different outlook on time: Visual and auditory month names elicit different mental vantage points for a time-space synaesthete. *Cortex*, *45*(10), 1217–1228. https://doi.org/10.1016/j.cortex.2009.05.014

Jonas, C. N., & Price, M. C. (2014). Not all synesthetes are alike: Spatial vs. Visual dimensions of sequence-space synesthesia. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01171

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

PostGIS 3.6.2dev manual. (n.d.).

Price, M., & Pearson, D. (2013). Toward a visuospatial developmental account of sequence-space synesthesia. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00689

Root, N., Asano, M., Melero, H., Kim, C.-Y., Sidoroff-Dorso, A. V., Vatakis, A., Yokosawa, K., Ramachandran, V., & Rouw, R. (2021). Do the colors of your letters depend on your language? Language-dependent and universal influences on grapheme-color synesthesia in seven languages. *Consciousness and Cognition*, *95*, 103192. https://doi.org/10.1016/j.concog.2021.103192

Root, N., Chkhaidze, A., Melero, H., Sidoroff-Dorso, A., Volberg, G., Zhang, Y., & Rouw, R. (2025). How "diagnostic" criteria interact to shape synesthetic behavior: The role of self-report and testretest consistency in synesthesia research. *Consciousness and Cognition*, *129*, 103819. https://doi.org/10.1016/j.concog.2025.103819

Rothen, N., Jünemann, K., Mealor, A. D., Burckhardt, V., & Ward, J. (2016a). The sensitivity and specificity of a diagnostic test of sequence-space synesthesia. *Behavior Research Methods*, *48*(4), 1476–1481. https://doi.org/10.3758/s13428-015-0656-2

Rothen, N., Jünemann, K., Mealor, A. D., Burckhardt, V., & Ward, J. (2016b). The sensitivity and specificity of a diagnostic test of sequence-space synesthesia. *Behavior Research Methods*, *48*(4), 1476–1481. https://doi.org/10.3758/s13428-015-0656-2

Rothen, N., Seth, A. K., Witzel, C., & Ward, J. (2013). Diagnosing synaesthesia with online colour pickers: Maximising sensitivity and specificity. *Journal of Neuroscience Methods*, *215*(1), 156–160. https://doi.org/10.1016/j.jneumeth.2013.02.009

Sagiv, N., Simner, J., Collins, J., Butterworth, B., & Ward, J. (2006). What is the relationship between synaesthesia and visuo-spatial number forms? *Cognition*, *101*(1), 114–128. https://doi.org/10.1016/j.cognition.2005.09.004

Seron, X., Pesenti, M., Noël, M.-P., Deloche, G., & Cornet, J.-A. (1992). Images of numbers, or "when 98 is upper left and 6 sky blue". *Cognition*, *44*(1), 159–196. https://doi.org/10.1016/0010-0277(92)90053-K

Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology*, *103*(1), 1–15. https://doi.org/10.1348/000712610X528305

Smilek, D., Callejas, A., Dixon, M. J., & Merikle, P. M. (2007). Ovals of time: Time-space associations in synaesthesia. *Consciousness and Cognition*, *16*(2), 507–519. https://doi.org/10.1016/j.concog.2006.06.013

Van Petersen, E., Altgassen, M., Van Lier, R., & Van Leeuwen, T. M. (2020b). Enhanced spatial navigation skills in sequence-space synesthetes. *Cortex*, *130*, 49–63. https://doi.org/10.1016/j.cortex.2020.04.034

Van Petersen, E., Altgassen, M., Van Lier, R., & Van Leeuwen, T. M. (2020a). Enhanced spatial navigation skills in sequence-space synesthetes. *Cortex*, *130*, 49–63. https://doi.org/10.1016/j.cortex.2020.04.034

Ward, J. (n.d.). *Optimizing a Measure of Consistency for Sequence-Space Synaesthesia*. https://doi.org/10.31234/osf.io/5cnr7

Ward, J. (2022). *Optimizing a measure of consistency for sequence-space synaesthesia*. https://osf.io/5cnr7_v1

Ward, J., Ipser, A., Phanvanova, E., Brown, P., Bunte, I., & Simner, J. (2018). The prevalence and cognitive profile of sequence-space synaesthesia. *Consciousness and Cognition*, *61*, 79–93. https://doi.org/10.1016/j.concog.2018.03.012

Ward, J., & Simner, J. (2022). How do Different Types of Synesthesia Cluster Together? Implications for Causal Mechanisms. *Perception*, *51*(2), 91–113. https://doi.org/10.1177/03010066211070761

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3

Table 1

My Caption

	Dataset	(Rothen et al., 2016a)	(Van Petersen et al., 2020b)	(Ward, 2022)	Ward 2 (personal communication)
Descriptive	Group				
N	Synaesthete	33	23	252	88
N	Control	37	21	215	17
n women	Synaesthete	24	20	202	-
n women	Control	37	19	178	-
Age	Synaesthete	23.1	23.2	37.2	-
Age	Control	28.2	21.6	19.9	-

Note. Demographics were not provided in the dataset Ward 2

Table 2

Summary of data sources

Data Source	Controls	Sequence-Space Synesthetes
PeterCor	21	22
Rothen	37	32
Ward	213	252
Ward2	18	90

Note. Number of participants for each data source by initial classification

Table 3

Summary of ROC analysis

(#tab:tbl-mytb3)

Feature	AUC	DP	threshold	sensitivity	specificity
Permuted valid structure [mean binary/9]	80.09	2.06	0.17	70.40	77.50
Perimeter [z-score]	78.50	2.04	1.92	73.60	74.29
Valid structure [mean binary/9]	77.07	1.81	0.17	71.73	71.43
Segment self-intersections [n/9]	71.71	1.68	1.17	79.47	58.93
Area of the polygon [z-score]	70.96	1.30	1.29	64.53	67.50
Area [z-score]	70.94	1.72	0.08	75.47	65.36
Simple topology [mean binary/9]	69.91	1.26	0.28	60.53	70.36
Permuted Chance level [z-score]	65.80	2.18	-3.67	93.87	37.86
Perimeter of the polygon [z-score]	60.65	1.32	10.46	83.47	43.21
Perimeter [screen size %]	56.31	0.69	0.03	75.20	40.00
Correlation between coordinates [r]	55.85	1.54	0.67	95.20	19.64
Area [screen size %]	51.93	0.82	0.21	76.27	41.79

Note. *AUC* = Aurea Under the Curve. *DP* = Discrimination Power

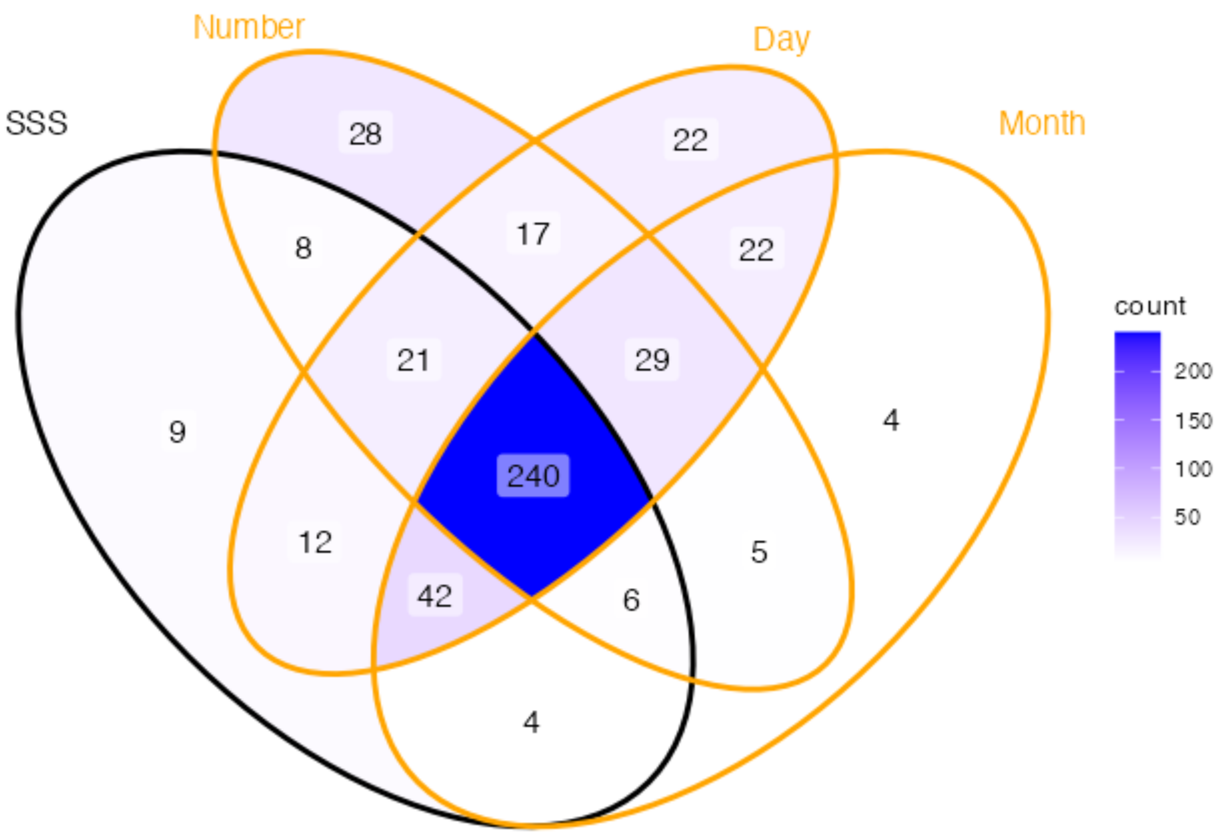
Table 4

Descriptives of each features

Feature	Controls	Sequence-Space Synaesthetes
Permuted valid structure [mean binary/9]	0.12 (0.16)	0.35 (0.23)
Perimeter [z-score]	3.33 (1.72)	1.61 (1.17)
Valid structure [mean binary/9]	0.13 (0.19)	0.37 (0.26)
Segment self-intersections [n/9]	8.83 (10.45)	1.71 (4.69)
Area of the polygon [z-score]	1.03 (0.99)	1.86 (1.24)
Area [z-score]	0.27 (0.33)	0.08 (0.14)
Simple topology [mean binary/9]	0.21 (0.22)	0.38 (0.26)
Permuted Chance level [z-score]	99.79 (790.15)	14.27 (217.45)
Perimeter of the polygon [z-score]	9.61 (3.41)	8.39 (2.49)
Perimeter [screen size %]	0.03 (0.03)	0.03 (0.09)
Correlation between coordinates [r]	0.40 (0.27)	0.32 (0.19)
Area [screen size %]	0.47 (0.88)	0.26 (0.54)

Figure 1

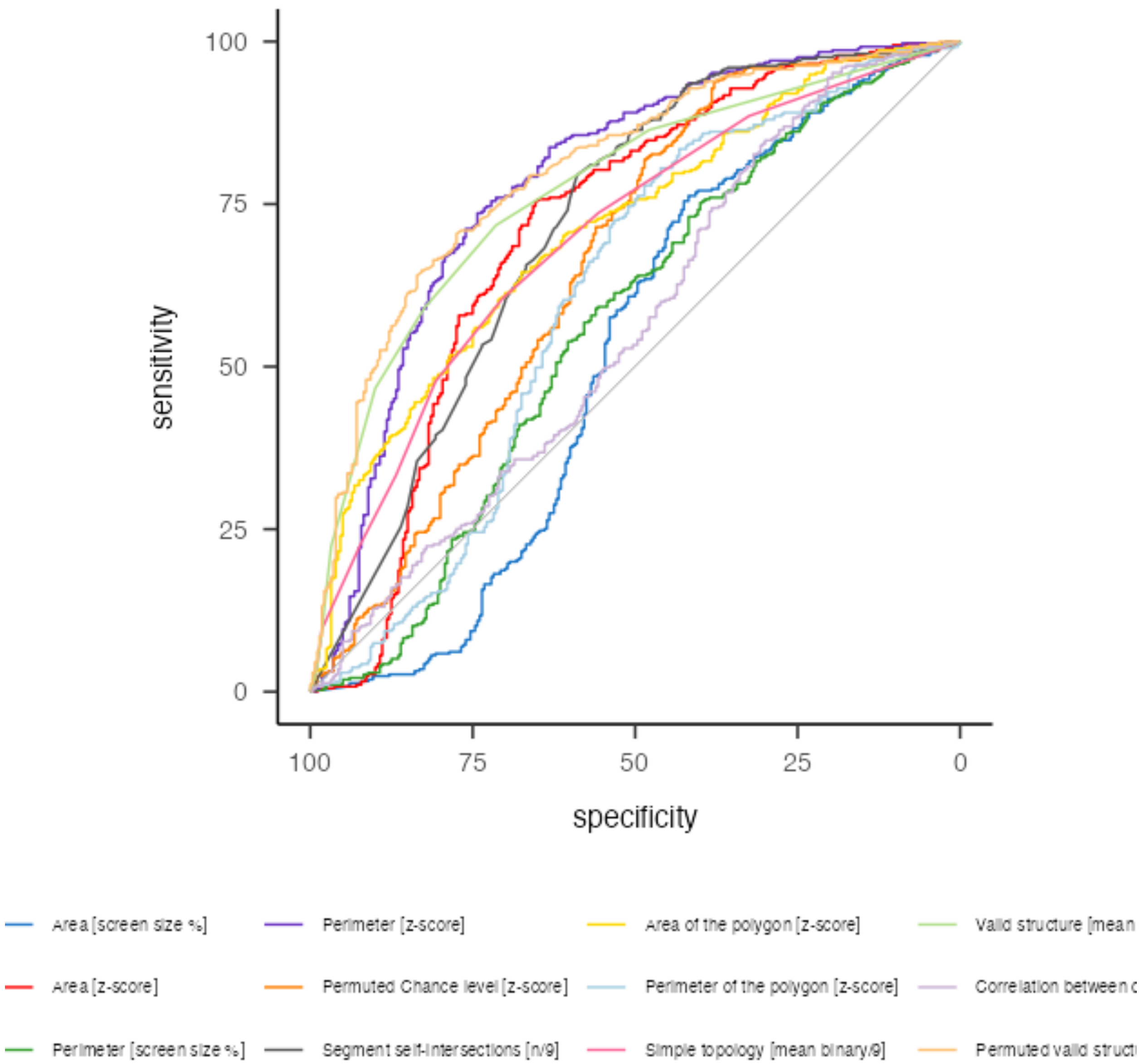
Venn diagram of the types of self-reported SSS



Note. Only data from Pr. Ward is represented here. SSS = Sequence-space synaesthesia

Figure 2

Receiver Operating Characteristic (ROC) curves of all features



Note. Grey line indicates chance level

Appendix

Appendix phase I

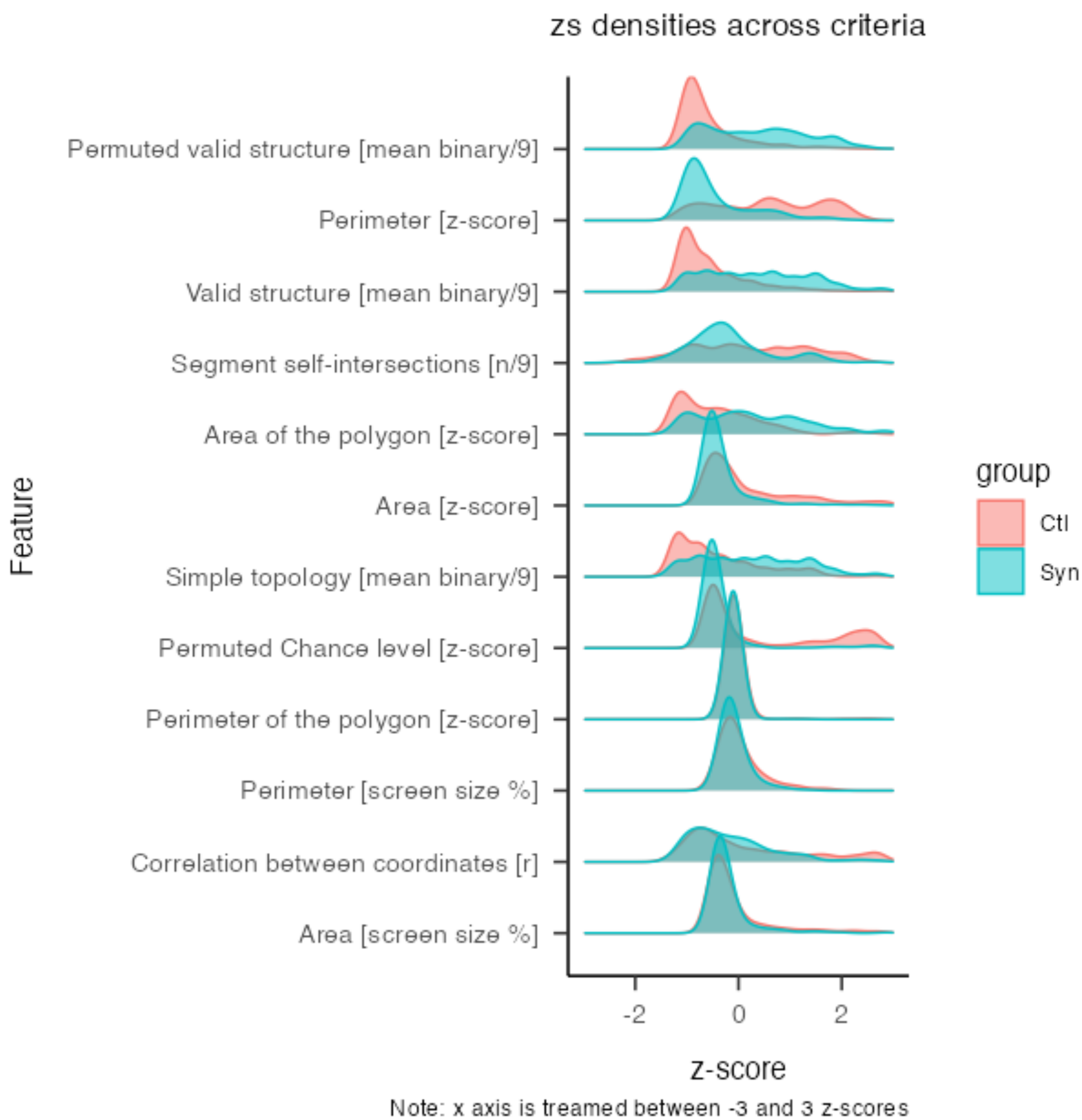
To additionally test the validity of the criteria, we computed the ROC again by sub-sampling the groups based on the questionnaire scores so to have more extreme groups. This was done only on the data from Ward, since the other did not include a questionnaire in the data.

Appendix 1 Feature distributions

Regarding the distributions of each features across the SSS vs. controls, Figure A1 we compared the density plots of each standardized criteria (in order to make them comparable) which visually indicates a bimodal distribution. Median z-scores per group are also presented in Table A1.

Figure A1

Density plots of all the features comparing SSS and controls



Note. all feature's score have been z-score transformed in order to be compared

Table A1

Median and SD of each standardized features

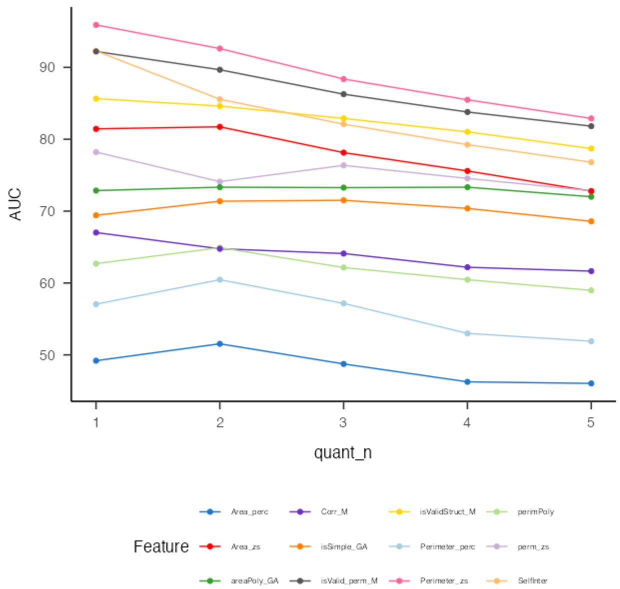
(#tab:tbl-mytb05)

Feature	Ctl	Syn
Permuted valid structure [mean binary/9]	Med = -0.82 (0.68)	Med = 0.35 (1.01)
Perimeter [z-score]	Med = 0.62 (1.03)	Med = -0.73 (0.70)
Valid structure [mean binary/9]	Med = -0.60 (0.72)	Med = 0.24 (1.01)
Segment self-intersections [n/9]	Med = 0.16 (1.15)	Med = -0.30 (0.84)
Area of the polygon [z-score]	Med = -0.53 (0.82)	Med = 0.16 (1.02)
Area [z-score]	Med = -0.16 (1.28)	Med = -0.51 (0.55)
Simple topology [mean binary/9]	Med = -0.75 (0.85)	Med = 0.11 (1.01)
Permuted Chance level [z-score]	Med = -0.29 (1.24)	Med = -0.52 (0.55)
Perimeter of the polygon [z-score]	Med = -0.10 (1.45)	Med = -0.10 (0.40)
Perimeter [screen size %]	Med = -0.11 (0.43)	Med = -0.17 (1.27)
Correlation between coordinates [r]	Med = -0.22 (1.17)	Med = -0.30 (0.83)
Area [screen size %]	Med = -0.30 (1.24)	Med = -0.34 (0.76)

Appendix 2 Sub-sampled data by questionnaire quantiles (20% steps)

We compared the data sampled by the questionnaire score. Based on the distribution of the questionnaire score, we sampled the 10 % with the lowest and 10 % with the highest scores. Those are then compared with the 20 and 20 % and so on until 40 and 40 %. The rationale of this procedure is that AUC, sensitivity and specificity should remain stable across percentiles for a feature to be valid, see Figure A2. In other words the ROC should remain unchanged if we take extreme groups compared to less extreme ones.

(A) Area Under the Curve (AUC)



(B) Discrimination Power (DP)

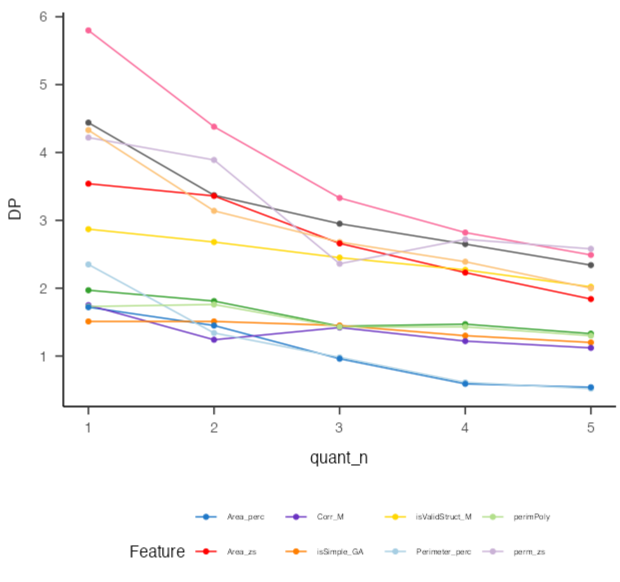
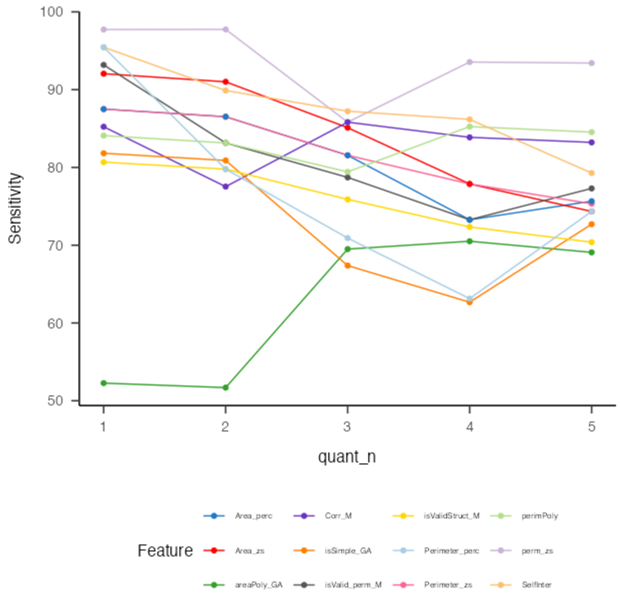


Figure A2

Lineplots of AUC and DP by percentile

(A) Sensitivity



(B) Specificity

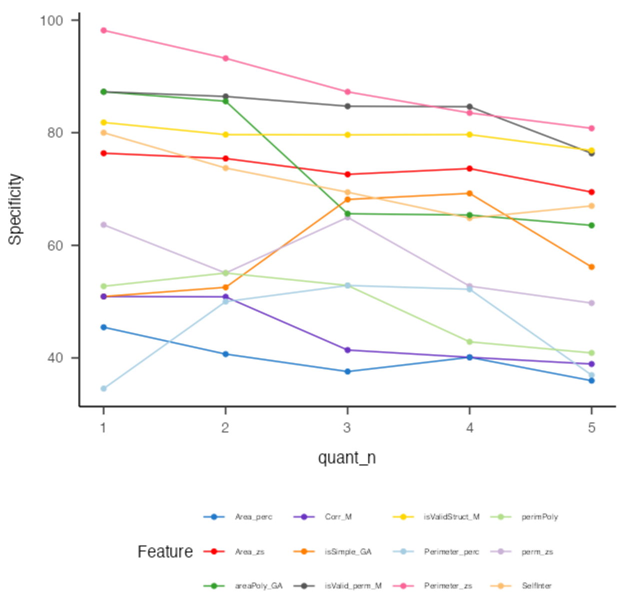


Figure A3

Lineplots of Sensitivity and Specificity by percentiles

Appendix 3 By dataset

Here we compare the ROC for each data sample. Note that the dataset labelled as Ward 2 has more synaesthetes than controls (5:1), see Table 2. Hence, we only present the descriptives per dataset for the two main features and other datasets.

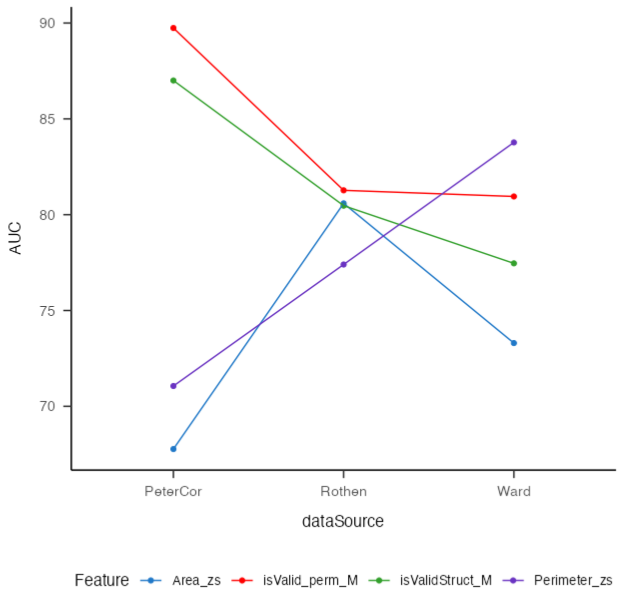
Table A2

Average feature for each group and dataset

(#tab:tbl-mytb4)

dataSource	Feature	Ctl	Syn
PeterCor	Perimeter_zs	Med = 1.24 (0.79)	Med = 0.73 (0.24)
Rothen	Perimeter_zs	Med = 3.07 (1.37)	Med = 1.18 (0.96)
Ward	Perimeter_zs	Med = 3.76 (1.65)	Med = 1.14 (1.19)
PeterCor	isValid_perm_M	Med = 0.07 (0.10)	Med = 0.56 (0.25)
Rothen	isValid_perm_M	Med = 0.09 (0.16)	Med = 0.27 (0.22)
Ward	isValid_perm_M	Med = 0.05 (0.15)	Med = 0.31 (0.24)

(A) Area Under the Curve (AUC)



(B) Discrimination Power (DP)

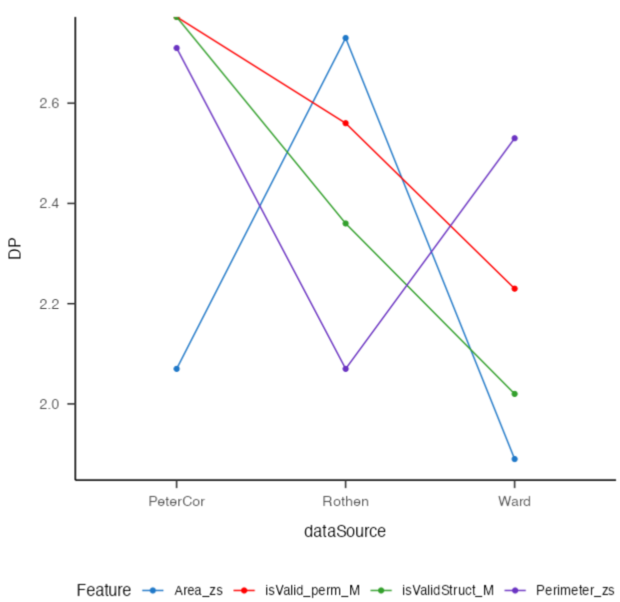
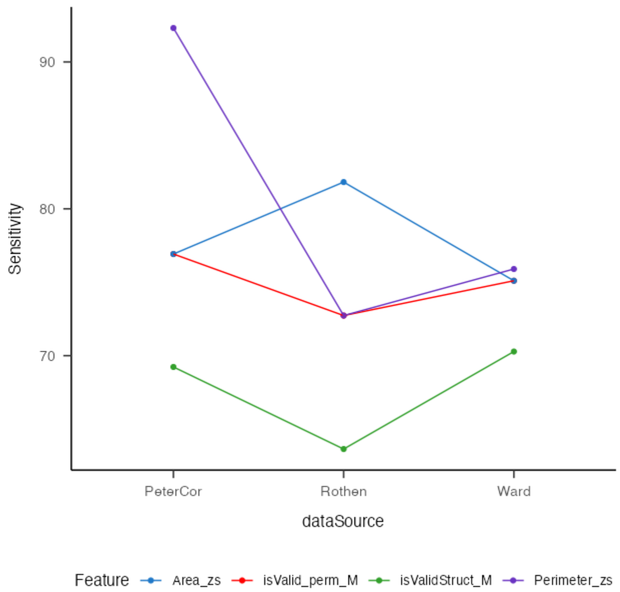


Figure A4

Lineplots of AUC and DP by data source

(A) Sensitivity



(B) Specificity

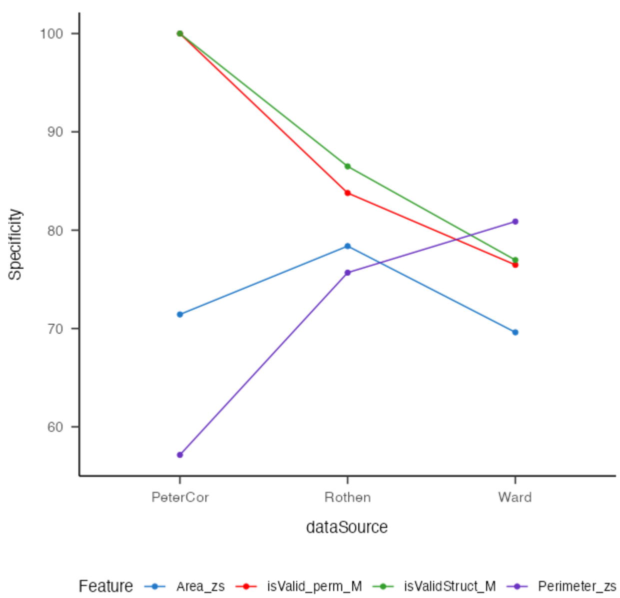


Figure A5

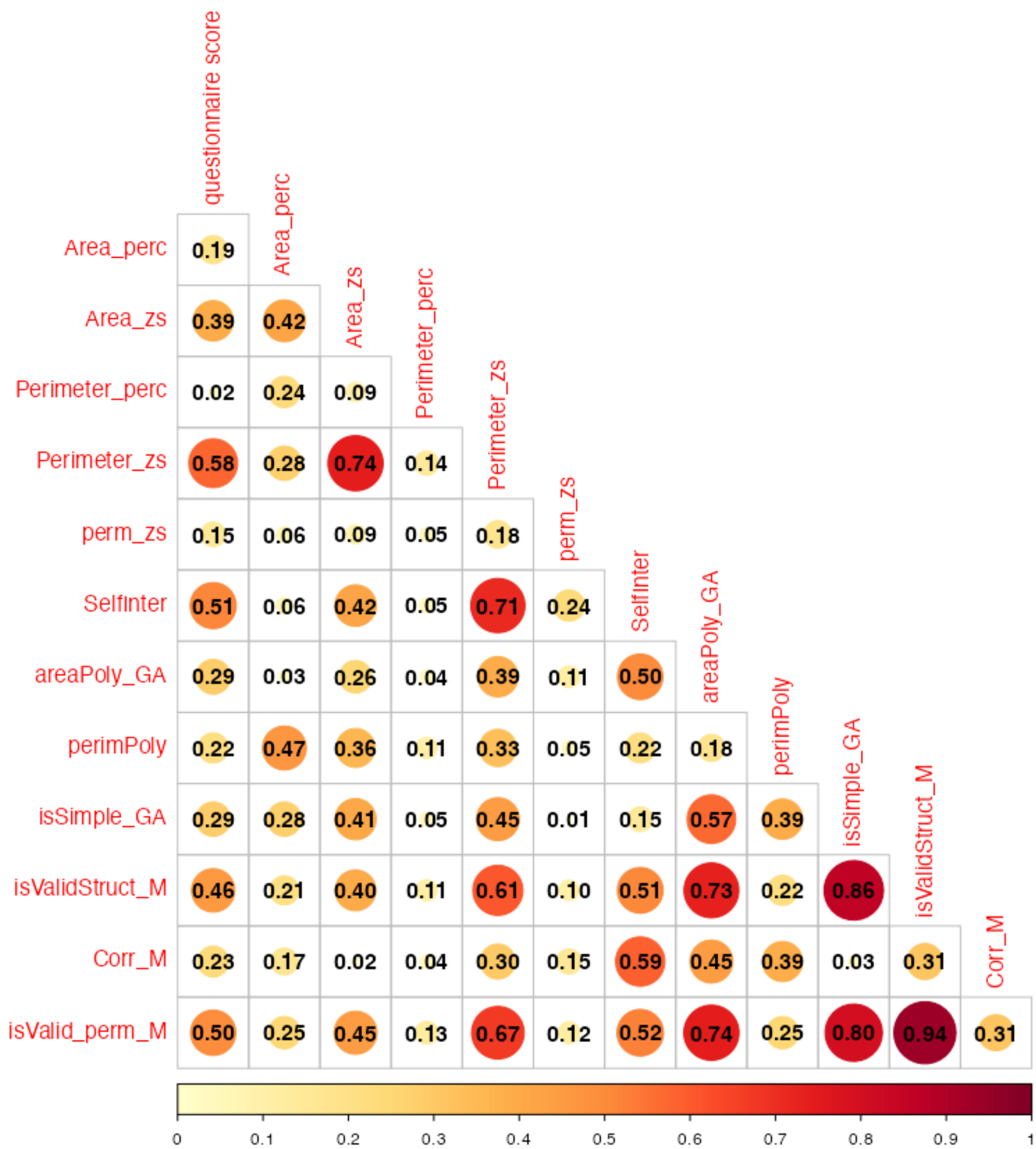
Lineplots of Sensitivity and Specificity by data source

Appendix 4 Correlation with self-report

The best criterion should also best correlate with SSS self-reported questionnaire score. Works only with Ward’s aggregated data, see Figure A3.

Figure A6

Correlation with self-reported questionnaire



Note. Only data from Ward is included here

Appendix 5 Code to visualize all space-forms

This exports many pdf's. It plots each ID and condition z-score x and y coordinates. Since each coordinate is repeated 3 times, these are represented by triangles. The line paths connect average coordinates to visualize forms (stimulus are ordered, i.e. 1 to 9, Monday to Sunday, January to December). Finally in the top right corner, each dots indicates if the ID would pass / fails depending on the criteria.