**Interpretation of similarity measures**

In this supplementary document, we want to provide the reader with a more detailed description and interpretation of the factor solution, specifically regarding the use of similarity measures. To compute the similarity of the time courses for each factor from the adult PCA, the correlation with each factor from the child PCA was computed. To measure topographic similarity, we first computed the average factor scores across participants at each electrode site for both stimulus types separately. Then, we computed the correlation between the average scores at the electrode sites separately for both conditions. This results in correlation matrices in which the rows represent factors from the adult PCA, and the columns represent factors from the child PCA. The closer the correlation coefficients are to 1, the more similar the factors are. For simplicities’ sake, we only report the similarities between the first six factors (i.e., the largest factors) from both PCAs but our scripts generate a full solution considering all factors. The time course and topographic similarities are reported in Table 1.

## Correlation of factor loadings

# The rows are factors from the adultPCA, the columns from the childPCA.

# Higher correlations indicate higher similarity.

Rloadings <- cor(rotFit\_adultPCA$loadings, rotFit\_childPCA$loadings)

## Compute average factor scores across participants

avr\_scr\_adPCA <- aggregate(. ~ cond + chan, data = scores\_adultPCA, mean)

avr\_scr\_chPCA <- aggregate(. ~ cond + chan, data = scores\_childPCA, mean)

## Correlation of factor scores

# Note: The excluded columns contain index variables for electrode sites etc.

# and need to be removed for the correlations to be meaningful.

Rtopo\_sta <- cor(avr\_scr\_adPCA[avr\_scr\_adPCA$cond == "sta",-c(1:4)],

avr\_scr\_chPCA[avr\_scr\_chPCA$cond == "sta",-c(1:4)])

Rtopo\_nov <- cor(avr\_scr\_adPCA[avr\_scr\_adPCA$cond == "nov",-c(1:4)],

avr\_scr\_chPCA[avr\_scr\_chPCA$cond == "nov",-c(1:4)])

# Please see script for some more convenience code which provides an automatic # ranking of the similarities.

Table 1

*Similarities between time courses and topographies from adult and child PCA*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | chFA1 | chFA2 | chFA3 | chFA4 | chFA5 | chFA6 |
|  | *Time course similarity* | | | | | |
| adFA1 | **0.95** | -0.20 | -0.19 | -0.11 | -0.20 | -0.22 |
| adFA2 | -0.17 | **0.99** | 0.20 | -0.14 | -0.09 | -0.10 |
| adFA3 | -0.18 | -0.09 | 0.44 | -0.11 | 0.41 | **0.96** |
| adFA4 | -0.17 | -0.18 | -0.14 | **0.85** | 0.10 | -0.13 |
| adFA5 | -0.14 | 0.36 | **0.91** | -0.14 | -0.08 | 0.13 |
| adFA6 | -0.17 | -0.12 | -0.07 | -0.07 | **0.90** | 0.16 |
|  | *Topographic similarity (standard sounds)* | | | | | |
| adFA1 | **0.97** | 0.32 | **0.97** | **0.98** | **0.97** | **0.94** |
| adFA2 | -0.93 | -0.18 | -0.92 | -0.93 | -0.94 | -0.95 |
| adFA3 | **0.95** | 0.25 | **0.94** | **0.95** | **0.96** | **0.96** |
| adFA4 | **0.95** | 0.24 | **0.95** | **0.96** | **0.97** | **0.98** |
| adFA5 | 0.44 | -0.33 | 0.52 | 0.50 | 0.47 | 0.56 |
| adFA6 | **0.98** | 0.34 | **0.98** | **0.99** | **0.98** | **0.97** |
|  | *Topographic similarity (novel sounds)* | | | | | |
| adFA1 | **0.92** | -0.60 | -0.45 | **0.87** | -0.76 | -0.74 |
| adFA2 | -0.93 | **0.94** | 0.40 | -0.60 | **0.90** | **0.94** |
| adFA3 | -0.85 | **0.92** | 0.30 | -0.49 | **0.87** | **0.92** |
| adFA4 | **0.90** | -0.50 | -0.24 | 0.90 | -0.68 | -0.71 |
| adFA5 | -0.69 | **0.85** | 0.26 | -0.28 | 0.77 | **0.87** |
| adFA6 | -0.07 | -0.05 | -0.15 | -0.20 | -0.04 | -0.12 |

*Note.* The values are Pearson correlations, higher numbers indicate higher similarity. The rows represent factors from the adult PCA, the columns represent factors from the child PCA. Similarities ≥ .80 are highlighted in bold. Note that high negative similarities indicate high topographic overlap but inverted polarities, for instance, a central negative and a central positive factor could correlate -1. This type of reverse overlap is not relevant for matching because we are looking for factors with nearly identical topography – including the polarity.

The pattern of high similarities (bold) in Table 1 clearly shows that the factors are more distinct in the temporal than in the spatial domain because there is typically only one factor with similar time course in both PCAs but there are several factors with similar topographies which is quite representative of typical ERP data (Dien, 2010a). We note that similarity measures must always be interpreted in the light of substantive expertise and should not be used in an automatic fashion. As our example will demonstrate, these measures can help to identify matching factors which are similar both in the temporal and spatial domain, but some factors differ considerably between age groups and would not be matched automatically. Therefore, we strongly recommend that researchers always report the criteria by which they identified or matched factors so that reviewers and readers can judge for themselves if they agree with the interpretations.

The easiest factors to match are those factors with similar time course *and* topography between groups, that is, factors without strong latency shifts. The Factors 1 and 2 from both groups seem to match very well with respect to time courses and topography. Considering the topographies and reconstructed ERPs, the mismatch of the topography for standard sounds (Factor 2) can be explained by the fact that Factor 2 seems to be almost absent for standard sounds in children or rather includes parts of the temporal N1c/T-complex (Bruneau et al., 1997). In both groups, Factor 1 occurs very late in the epoch with peak latencies around 0.730 s. The topography is characterized by a fronto-central negativity. Higher amplitudes were observed in children compared to adults and for novel sounds compared to standard sounds. Based on these characteristics, we labeled Factor 1 in both groups *late discriminative negativity (LDN).* In both groups, Factor 2 occurred with peak latencies between 0.170 and 0.180 s and was characterized by a central positivity and higher amplitudes for novel sounds compared to standard sounds. We labeled Factor 2 as *P2* in both groups.

For these two factors, which were of substantive interest in the original study, the matching was possible and relatively clear. However, we were also interested in differences between both groups in the P300 time range. Matching of factors in this time range is more challenging because it is known that the respective components occur later in children compared to adults (Riggins & Scott, 2020). Among Factors 3 to 6 in the adult PCA, the Factors 5 and 3 seem to represent early and late P3a, respectively, considering their latency between 0.2 and 0.3 s and centrally and fronto-centrally positive topography for novel sounds. According to the time course similarity, Factor 5 (early P3a, peak latency: 0.222 s) should be matched with child Factor 3 (0.240 s). This is a clear mismatch, considering that child Factor 3 is characterized by a large fronto-central negativity for standard sounds and almost no response to novels. The topographic similarities (standards: .52; novels: .26) underline this interpretation. Due to the latency shift, the time course similarity cannot be prioritized here. Instead, considering topographic similarity (especially for novels: .87), the child Factor 6 should be matched and labeled the child early P3a. Similarly, the adult Factor 3 (late P3a, 0.298 s) would be mismatched with the child early P3a (factor 6, 0.296 s) based on time courses alone. The next-best topographic match (novel similarity: .92) would be child Factors 2 and 6. Given that child Factor 2 (0.172 s) is in a clearly different time range and seems to be too early for a P3 component, and child Factor 6 was already matched as early P3a, we conclude that adult Factor 3 matches reasonably well with child Factor 5 (novel similarity: .87) and labeled it late P3a based on the fronto-central and positive topography for novels and a higher latency compared to the early P3a factors. The allocation of the factors to the early and late P3a is also in line with the typical central distribution of the early P3a and the often observed more anterior distribution of the late P3a (Escera et al., 2000).

To sum up, we identified and matched the following factors among the major factors: P2 (adult Factor 2, child Factor 2), the early P3a (adult Factor 5, child Factor 6), the late P3a (adult Factor 3, child Factor 5), and the LDN (adult Factor 1, child Factor 1).

Overall, our example dataset demonstrates that substantive reasoning is mandatory and cannot be replaced by automatic similarity matching. Nevertheless, such similarity measures can be very helpful – especially given the large number of factors in ERP research – because they offer an objective quantification as opposed to purely subjective judgements of similarity. We recommend complementing expert visual inspection with objective similarity measures to guide substantive interpretations.