The sphericity assumption was not violated in the earlier analysis. For this reason, it is reasonable to use the emotion × subjects (condition) MS_{error} term from the original ANOVA. This is 1,325, and so F = 40,296.7/1,325, and the contrast could be reported as: $F_{1,174} = 30.4$, $MS_e = 1{,}325$, $p < .01$.

As is often the case for a contrast, the explanatory power of the interaction contrast is of greater interest than the test. *SS* for the interaction is 44,231 and therefore the interaction contrast explains about 91% of the interaction effect ($r^2_{alerting} = .91$). This is equivalent to a correlation of .95 between contrast weights and the residualized cell means ($r_{alerting} = .95$). Although there is quite a bit going on in Figure 16.3, much of the variation is down to main effects. Of the variation that remains, the vast majority can be explained in terms of pride being harder to identify from facial expression than from body posture, and happiness being harder to identify from body posture than facial expression.

16.4.3 Effect size

Repeated measures and mixed designs are especially difficult to obtain appropriate standardized effect size metrics for. Many commonly calculated quantities (e.g., η^2_p or *g*) are not comparable to similar metrics obtained from independent measures designs. It a good idea to compare effects from different designs using unstandardized measures (e.g., simple mean differences) as a first – and possibly only – step. Standardized effect size metrics need to take into account factors that that may distort the standardizer (the variance or standard deviation used to scale the effect). An important contributor to the standardizer in an independent measures design is individual differences (which in a repeated measures design are estimated separately from other sources of variance).

The generalized effect size measures of Olejnik and Algina (2003) provide a starting point for 'design neutral' standardized effect size metrics. Their approach is to calculate generalized statistics that treat repeated measures equivalently to independent measures designs. One proviso is that for statistical power or sample size estimation, the appropriate metric is one that matches the design of the study being planned. In theory η_g^2 can be calculated with the formulas for other factorial designs by treating subjects as an additional measured factor. In practice, ANOVA software rarely provides the *SS* for such a calculation in a convenient format. Following Olejnik and Algina, I suggest calculating the required denominator for η_g^2 by subtraction. The goal is to exclude all manipulated factors or interactions with only manipulated factors (except the one under consideration). An indicator variable I is used to designate whether the effect under consideration is a manipulated factor (I = 1) or a measured factor (I = 0).

$$\eta_g^2 = \frac{SS_{effect}}{SS_{total} - \sum_{manip} SS_{manip} + I \times SS_{effect}} \qquad \text{Equation [16.11]}$$

This formula excludes manipulated factors from the denominator (and adding the focal effect back in only if the focal effect is a manipulated factor). Interactions with measured factors are considered measured factors. Repeated measures fixed factors tend to be manipulated factors, though it may be reasonable to treat them as measured factors in some situations.

Olejnik and Algina (2003) also extend ω_g^2 (generalized omega-squared) to designs with repeated measures factors. The correct formulas can become rather complex and the simplest solution is to refer to Tables 2, 3 and 4 of their paper.

Example 16.6. In the two-way mixed ANOVA for the pride data, the two factors are the emotion to be recognized and the experimental condition (whether the pictures showed face, torso or both face and torso). The experimental condition is a canonical example of a manipulated variable. Emotion is manipulated by the experimenter, and for comparisons with other experiments might be considered as such. For other purposes – for example to gauge impact on everyday performance – it may be considered a measured variable (because expressions of happiness, pride and surprise are a routine part of everyday experience).

Treating both variables as manipulated factors, η_g^2 for the interaction is:

$$\eta_g^2 = \frac{SS_{effect}}{SS_{total} + I \times SS_{effect} - \sum\limits_{manip} SS_{manip}} = \frac{44231}{521678.2 + 1 \times 44231 - (44231 + 5616 + 26060)} = .090$$

The interaction would account for around 9% of the total sample variance in an equivalent independent measures design. Replacing SS_{effect} with $SS_{contrast}$ allows versions of η^2 to be calculated for a contrast. Thus η_g^2 for the interaction contrast is:

$$\eta_g^2 = \frac{SS_{effect}}{SS_{total} + I \times SS_{effect} - \sum\limits_{manip} SS_{manip}} = \frac{40296.7}{521678.2 + 1 \times 40296.7 - (40296.7 + 5616 + 26060)} = .083$$

In practice it is usually better to evaluate contrasts relative to the effect they are decomposed from, rather than in terms of total variance. The contrast explains about 91% of the interaction effect and about 8% of sample variance in an equivalent independent measures design.

16.5 MANOVA

Multivariate analysis of variance (MANOVA) is technique of potential interest whenever correlated measurements are obtained. MANOVA is designed for applications in which several correlated outcome variables or *DV*s (dependent variables) are measured.

An important application for repeated measures designs is a form of MANOVA called profile analysis. This does not assume sphericity and will sometimes greater statistical power than an epsilon-corrected ANOVA. In this application, already considered in passing, the repeated measures are treated as correlated outcome variables (which, in a sense, they are). Knowing in advance whether the MANOVA tests are more powerful than the corrected ANOVA analysis is far from simple. It would be possible to estimate the statistical power of each technique if the population covariance matrix $\mathbf{\Sigma}$ were known. Obtaining a good enough estimate of $\mathbf{\Sigma}$ (e.g., in a pilot study) to determine the relative power of the two approaches is likely to be difficult (but see Miles, 2003). Estimates of variances and covariances from small samples tend to be very imprecise. The MANOVA approach will tend to have greater power when many repeated measures are taken, when the degree of sphericity violation is large and when samples sizes are large (but there are departures from this general trend).